CHALMERS
UNIVERSITY OF TECHNOLOGY

Increasing Throughput Potential in Functional-Oriented Machining Plants

Master of Science Thesis
in the Quality and Operations Management master's Programme

ADAM LEWESTAM
HENRIK MÄKI

Department of Technology Management and Economics
Division of Logistics and Transportation
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2015
Report No E2015:079

MASTER'S THESIS E2015:079

Increasing Throughput Potential in Functional-Oriented Machining Plants

ADAM J. LEWESTAM

HENRIK S. U. MÄKI

Department of Technology Management and Economics
Division of Logistics and Transportation
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2015

Increasing Throughput Potential in Functional-Oriented Machining Plants
ADAM J. LEWESTAM
HENRIK S. U. MÄKI

Master's Thesis E2015:079

Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone + 46 (0) 31- 772 1000

Chalmers Reproservice
Göteborg, Sweden 2015

Abstract

Productivity and throughput are central aspects in a production system´s ability to satisfy market demand. Throughput is thus an indicator of system performance and is dependent on how the production resources are utilized. Machining plants, for which a functional oriented production layout is given, are exposed to a number of factors limiting their maneuverability when striving to increase throughput and better satisfy their customers. Based on this, the purpose of this thesis is to investigate capacity losses, analyze the potential for increasing throughput and propose an approach for how to fulfill this potential.

To fulfill this purpose, the thesis is based on a case study at a Swedish company in the Aerospace industry that finds itself in this situation. The research strategy is a combination of both quantitative and qualitative methods in terms of production data analysis, interviews and observations.

The analysis revealed that the case company's production system loses 28 % of its available capacity with its current set-up and planning and control procedures. This capacity loss is mainly caused by two factors. The first one, high variability, negatively impacts the lead time at any given utilization level and creates a disrupted product flow. The second one, high levels of Work-In-Process, leads to unnecessary bottlenecks, long lead times and an inertly production system. High variability and Work-In-Process are in turn effects of a number of identified physical and managerial constraints that must be elevated in order to increase throughput.

The thesis results in a recommended framework that aims at increasing throughput for functional-oriented machining plants, without major investments or physical rearrangements. The framework combines new layout and routing principles with the use of simpler planning and control procedures, which primarily will reduce variations, complexity, unnecessary Work-In-Process and lead times. This creates a pulling product flow that is more predictable and transparent. The new principles are in turn supported by an aligned business governance and performance measurement strategy, matching the current manufacturing situation.

Keywords: Functional-oriented layout, Variability, Work-In-Process (*WIP*), Throughput, Theory of Constraints, Virtual Groups, Simplified Drum-Buffer-Rope (*S-DBR*), Capacity utilization, Pull production.

Preface and Acknowledgements

This thesis was conducted as a part of the examination from the Master's programme Quality and Operations Management at Chalmers University of Technology. The thesis was carried out during the spring of 2015. The research was initiated by GKN Aerospace Sweden AB and conducted on their behalf.

We would like to thank everyone involved in guiding us through this process and provided us with the information and tools needed. A special thanks to Fredrik Plyhm, Håkan Colliander and everyone else at Logistics Development for their expertise, advice and willingness to help with data support.

Finally, we would like to thank our supervisors Lars Medbo at Chalmers and Torgny Almgren at GKN. Your advice and support has been valuable to us throughout this process.

We hope that you will enjoy the reading and that our results can be useful in facing the challenges to come.

The authors

Adam Lewestam   Henrik Mäki

Table of Contents

List of Figures

List of Tables

This page intentionally left blank.

1 Introduction

This chapter introduces the reader to the problems the case company currently experiences and to related theory, which leads to the purpose of this study. Furthermore, the reader is presented with a problem analysis, resulting in three research questions that the study will answer in order to fulfill the purpose.

1.1 Problem Background

The ultimate goal of any organization is to satisfy their customers (Bergman & Klefsjö, 2010). In operations management theory it is believed that higher levels of customer's satisfaction are achieved by increasing the productivity, e.g. by shortening the delivery lead time. A company's productivity is a measurement of how efficiently it can transform its inputs, materials and resources, to outputs, products (Anderson et al., 1997).

GKN Aerospace Sweden AB's (GKN) production unit in Trollhättan, Sweden, manufactures components for aerospace engine producers. Their capital-intensive production is characterized by long and varying machining operations in a functional layout, which generate a complex and disrupted production flow. As a consequence of these aspects, GKN currently experiences problems with delayed customer deliveries due to limited capacity caused by constraints in their production.

The production resources and the company's managerial principles and rules thus define the upper capacity limit for the production system. However, when resources are not used for productive purposes in relation to their full capacity, capacity losses emerge (Slack et al., 2010). This means that the production system has processes operating below their maximum capacity, while one or several processes might operate close to their capacity limit. The processes pushed to their capacity limit, caused by physical limits or managerial planning and decisions, therefore become capacity constraints for the whole production system. Both physical and managerial constraints, determine the pace for the whole production system and thereby the maximum output rate. Elimination and better management of constraints is thus necessary to increase the productivity rate and to achieve a smooth and swift flow of products through the processes (Slack et al., 2010; Schmenner & Swink, 1998).

However, the characteristics of GKN´s situation reduce the maneuverability of managing the limiting constraints. The activities in the production are mainly machining operations and each work center perform in general only one type of operation. The machines are to a large extent also impossible to move due to physical constraints in the production facilities. This implies that the current functional layout is rather fixed. Furthermore, the available funds for investment in more production resources are limited.

The question that remains is thus how a company with a fixed functional layout without investment resources, such as GKN, can address the problem with capacity losses and overcome existing physical and managerial constraints to increase their throughput and thus satisfy their customers.

1.2 Purpose

The purpose of this study is to investigate capacity losses and analyze the potential for increasing throughput in a machining factory where a functional-oriented layout is given. Furthermore, it aims to propose an approach for how to fulfill this throughput potential.

1.3 Problem Analysis

GKN is currently experiencing problems with fulfilling customer orders and is constantly behind their delivery schedule. The reason for the seemingly stable delay is a continuous revision of the delivery plan[1]. This implies that GKN constantly overestimates their production capacity and puts a too high workload on the system. To be able to better meet their customer demand, GKN therefore needs to increase their throughput. In order to improve the throughput there is initially a need to understand why they are falling behind and thus determine what the actual available capacity of the system is. This leads on to the below research question.

- *What is the actual capacity of the production system with the current set-up?*

A capacity constraint can arise from natural causes, e.g. when the physical limit of a certain resource type is reached, or is caused by aspects in the process logic such as production steering and product routing (Slack et al., 2010). Since investment capital is limited, the potential to increase the throughput of the system is dependent on if these non-natural capacity constraints can be elevated. The next step is thus to investigate whether there are resources in the system that are not used optimally, that when optimized, can elevate the constraint. However, since the goal is to increase the throughput of the system, there is no intrinsic value in increasing the resource utilization if it does not improve the product flow (Goldratt & Cox, 1986). The improvement potential consists therefore of the slack resources, i.e. capacity losses, that when utilized increase the product flow. This leads on to the second research question of this study.

- *How extensive are the capacity losses in the system and what are the main causes?*

The potential for throughput improvement is constrained by physical characteristics mentioned above, such as big machines that cannot be moved and limited available investment capital. The solution space that remains consists thus of actions on the system level that optimizes the production logic, such as

[1] Logistics Developer 1 [Logistics Development]. Interviewed by the authors 2015-01-27.

optimizing production steering and product routing, and actions on a work center level that increase the efficiency of the work center. Again, increasing the efficiency on the work center level is only justifiable if the capacity of the work center constrains the capacity of the entire system (Slack et al., 2010). The last part of the study´s purpose is to estimate the throughput potential and give recommendations on how GKN can use these actions to fulfill this potential and increase the productivity of the production system. The last research question is thus formulated as below.

- *Given GKN's situation, what is the potential for increasing the throughput and how can it be achieved?*

1.4 Delimitations

The study focuses on the production system in the factory. Issues and topics concerning the processes of material inflow to and outflow from the factory are thus not discussed since it is out of the scope of the study. Furthermore, the recommendations to the company are only on a conceptual level. The entire implementation phase with needed tools and software is left outside the scope of this study.

2 Theoretical Framework

This chapter covers the theory needed for understanding and analyzing the situation at the company and can be considered to be the theoretical framework for the research process of this study. The first sections up to 2.3.2 define key concepts in performance measurement, capacity management and production steering and control. They also provide the tools necessary for understanding the current situation and its issues and give the reader an understanding of the subject. The remaining sections are to create a framework for how to improve the situation and resolve the problems.

2.1 Performance Measurement

Performance measurement is the base for all improvement efforts and is thus key for creating a sustainable competitive advantage in an ever faster changing environment with increasingly high customer demands (Slack et al., 2010). Traditional performance measurement based on accounting has proven insufficient and counter-productive as basis for manufacturing improvement actions (Abdel-Maksoud et al., 2005). Manufacturing performance has however great impact on both top line growth and bottom line efficiency, why performance indicators focused on manufacturing is key for the overall company performance (Ghalayini et al., 1997).

This section covers measurements of manufacturing performance. First, definitions and explanations of relevant system level performance indicators and how they are related and connected to each other, are presented. Thereafter the concept of variability is presented and the impact variability has on the production system performance is covered in depth.

2.1.1 Performance indicators

The performance indicators that are defined and explained below are measures of manufacturing performance in terms of quality and time, and will be needed throughout this study. The indicator´s impact on the overall company performance is covered.

Little's Law

Little´s law states the relationship between three essential performance indicators, throughput time, Work-In-Process and cycle time, as below (Slack et al., 2010).

$$Throughput\ Time = WIP \cdot Cycle\ Time \ (1)$$

The throughput time of a production system is the time a product that is about to enter the shop floor will stay within the production system. Throughput time has a great impact on the company´s ability to meet customer order due dates and how quickly it can respond to environmental changes (Slack et al., 2010). Work-In-Process, *WIP,* is the average number of products in the system at any given time. Besides having an impact on the throughput time, *WIP* also implies cost for invested capital (Schönsleben, 2007). The cycle time is the intermediate time between two products leaving the production system. The cycle time is directly connected to the available production resources and their efficiency (Hopp & Spearman, 2011).

Value Added Ratio

This indicator is a measurement of to what extent the throughput time is used for value adding activities and is defined as below.

$$R_{VA} = \frac{Value\ Added\ Time}{Throughput\ Time} \ (2)$$

Non-value adding activities is waste and should be reduced as much as possible since it implies unnecessary lead time and costs (Schönsleben, 2007).

Capacity Utilization

Capacity implies investment in production resources and a high utilization of said resources implies a high return on investment (Hopp & Spearman, 2011). However, this is only true under the condition that a higher utilization will increase the throughput of the system, i.e. the system will produce more sellable products that will add to the result of the firm (Goldratt & Cox, 1986). Capacity utilization is defined as below (Schönsleben, 2007).

$$Utilization = \frac{Load}{Capacity} \ (3)$$

Hopp and Spearman (2011) and Slack et al. (2010) present an alternative way to calculate capacity utilization. The utilization can be calculated as

$$u = \frac{r_a}{r_e} \ (4)$$

$$where$$

$$r_a: Rate\ of arrivals\ in\ jobs\ per\ time\ unit$$
$$r_e: Effective\ processing\ rate$$

2.1.2 Variability

There are several possible sources of variability affecting processes and these can be divided in two major categories, process time variability and flow variability.

Process Time Variability

Process time variability (Slack et al., 2010; Hopp & Spearman, 2011), is caused by variability in the nature of the jobs to be processed, in process steps themselves or the items processed (Schmenner & Swink, 1998). The most prevalent sources of variability affecting processes in manufacturing are according to Hopp and Spearman (2011):

Natural Variability - occur since the jobs' natural process time varies, the nature of the operations is different, varying operator performance and that different operators perform tasks in slightly different ways. Automated processes have less natural variability than manual ones. Even though these are tightly controlled, there will always be some natural variability.

Unplanned Breakdowns - are in many production systems the greatest factor causing variability.

$$t_e = \frac{t}{A} \quad (5)$$

where

$$A = \frac{MTTF}{MTTF+MTTR} \quad (6)$$

Non-preemptive Outages - represent downtimes that to some extent can be controlled. Examples of this are changeovers, set-ups, preventive maintenance, breaks, operator meetings and shift changes. These outages occur in general more often between jobs than during them. Planning these downtimes carefully and continuous work with reducing them is important to reduce the variability.

Rework - is another major source of variability in manufacturing systems caused by quality problems. Rework has an effect of stealing capacity, requiring divergent set-ups, disrupts planning and flow, and thereby generates great variability of the effective process times. Thus, more rework causes more variability, which causes longer queues, higher WIP and longer throughput time.

Flow Variability

The second type, flow variability, stems from the demand of the process, timing and the transfer of jobs between processes, where variability at one process affects other process in the same production flow (Slack et al, 2010; Hopp & Spearman, 2011; Schmenner & Swink, 1998).

The variability in the flow is defined by the arrival of jobs to processes. Clearly, if an upstream process has high variability, this will also be the case further down in the stream among the processes. Hopp and Spearmans' (2011, p. 318) *law of variability placement* states the relationship as:

"In a line where releases are independent of completion, variability early in a routing increases cycle time more than equivalent variability later in the routing."

The variability in the flow can be measured by the coefficient of variation of the interarrival times, c_a

$$c_a = \frac{\sigma_a}{t_a} \quad (7)$$

Effects of Variability

The influence of variability causes undesirable effects on processes and the production system, and results in reduced effective capacity (Slack et al., 2010). Greater random variability in and between processes make the processes less productive and thereby decrease the throughput for the system (Schmenner & Swink, 1998). Hopp and Spearman (2011, p. 309) defines this relationship as the *Law of Variability*:

"*Increasing variability always degrades the performance of a production system"*.

Greater variability in processes will according to Slack et al. (2010) make the processes endure from reduced utilization and longer throughput, which build up queues and thereby higher *WIP*. Processes with high variability must therefore provide extra capacity to compensate from the reduced utilization (Slack et al., 2010), or accept longer throughput times and higher inventory levels. Hopp and Spearman (2011, p. 309) state these relationships as the *Law of Variability Buffering* where:

"*Variability in a production system will be buffered by some combination of:*

1. *Inventory*
2. *Capacity*
3. [Throughput] *Time"*

This non-linear relationship has been described by numerous authors, e.g. Hopp & Spearman, 2011; Slack et al., 2010), and clearly shows that as a process approaches 100 % capacity utilization, the longer the average queue time will be. The only way to guarantee short waiting times is to keep low process utilization. The relationship is visualized in Figure 1.

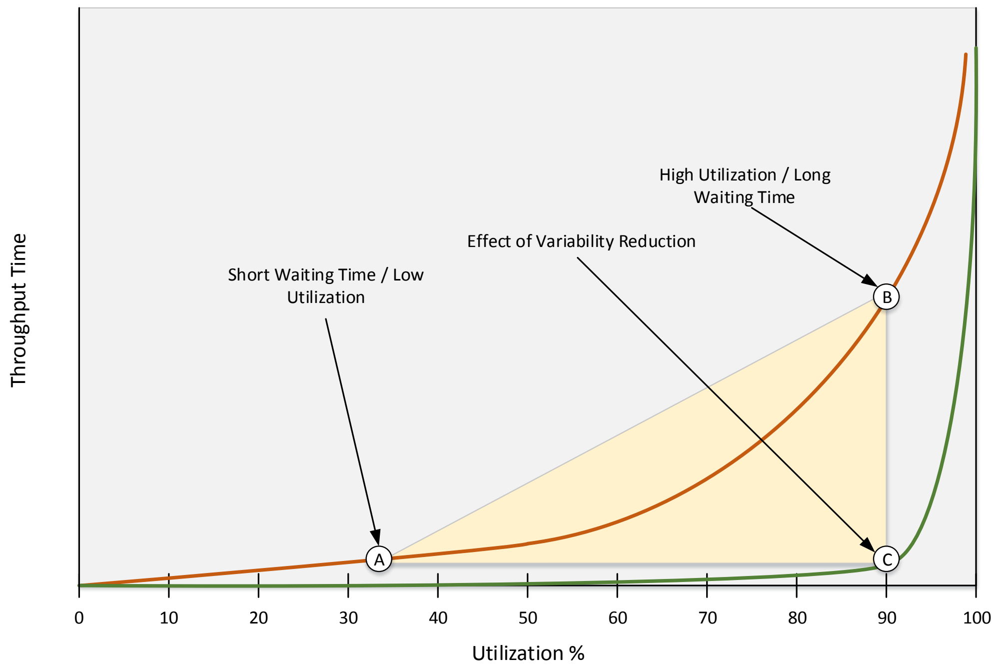

Figure 1: Relationship between utilization, variability and throughput time.

The figure shows that there are three different scenarios a company can reach when designing their processes (Slack et al., 2010):

A. Achieve high utilization but accept long waiting times.
B. Accept low utilization and thereby short waiting times.
C. Reduce the variability in flow and process time and attain high utilization and short waiting times.

Hopp and Spearmans' (2011, p. 317) *Law of Utilization* synthesizes this as:

"If a station increases utilization without making any other changes, average WIP and cycle time will increase in a highly non-linear fashion."

Measures of Variability

Variability is measured by using standard measures as variance, standard deviation and mean from statistics. The level of variability can be divided into three classes (Hopp & Spearman, 2011). To put variability into a context, the measure must be relative instead of absolute. Hopp and Spearman (2011) describe the *coefficient of variation,* c_v, where *t* is the mean and σ the standard deviation

$$c_v = \frac{\sigma}{t} \quad (8)$$

Table 1 shows variability classes for c_v where low variability has c_v less than 0.75, moderate variability when c_v is between 0.75 and 1.33 and high variability when c_v is greater than 1.33.

Table 1: Classes of variability (Hopp & Spearman, 2011).

Variability Level	Coefficient of Variation
Low	$CV < 0.75$
Moderate	$0.75 \leq CV \leq 1.33$
High	$CV > 1.33$

Variability Interplay - Queueing Systems

The interplay between process and flow variability can be described by a queuing system. If variability would not exist, queues would not exist at all since one would be able to adapt the capacity of the work station to exactly match its demand (Slack et al., 2010). However, queues and performance of a work station can be characterized by a queuing system, given parameters such as process time and variability in arrival and process time (Hopp & Spearman, 2011).

There is an endless variety of different queuing systems for an endless variety of situations. However, the complexity of a queuing system quickly increases as the situation of a work station becomes more special. Modeling the complexity of each work station in a factory would therefore quite quickly get out of hand. The queuing system described will hence be general and serve more as an approximation than an exact description of the system.

Queuing systems that are primarily considered are of type G/G/m since they, according to Hopp and Spearman (2011) and Slack et al. (2010), can be directly useful in practice when modeling systems of work stations. The authors mean that is not completely accurate but is good for practical purposes. G/G/m describes systems with general, e.g. normal or uniform but non-exponential arrival rates and process times distributions (Slack et al., 2010; Hopp & Spearman, 2011).

The expected time in the queue is

$$W_q = \left(\frac{c_a{}^2 + c_e{}^2}{2}\right)\left(\frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)}\right) t_e \ (9)$$

$$where$$

$$m: number\ of\ paralell\ servers\ at\ work\ station$$

$$c_e: process\ time\ variation\ coefficient$$

$$c_a: arrival\ rate\ variation\ coefficient$$

$$t_e: average\ processing\ time$$

$$W_q: expected\ queue\ time$$

$$u: utilization\ rate$$

Managing Variability

Due to the disrupting effects that process time and flow variability cause, it is desirable to reduce the levels in order to achieve shorter throughput times. The production flow will be more even when the variability is narrowed down (Schmenner & Swink, 1998). Shifting towards a pull-system, more even scheduling and reduction of batch releases are examples of approaches for reducing variability (Hopp & Spearman, 2011). Pull systems will be explained further in section 2.3.2. Hopp and Spearman (2011) suggest that reducing variability should start at bottlenecks or other high-utilization work stations since such results will give the most effect in the production system, see section 2.2.2. Furthermore, reducing process time variability at work stations upstream in the production flow will result in more even arrival rates and flow in the whole production system (Hopp & Spearman, 2011).

2.2 Capacity Management

This chapter covers the term, and management of, capacity at an individual resource level, in section 2.2.1, as well as on a system level, in 2.2.2.

2.2.1 Capacity Efficiency

The capacity of a resource is measured in the hours of work, or number of products if it only handles one type of product, it can produce. The capacity of a work center is thus a combination of machine hours and labor hours produced by the person manning the machine. This maximal capacity, the *theoretical capacity,* is simply the number of resources multiplied by the number of hours the resources are active (Slack et al., 2010). However, the *theoretical capacity* is rarely fully available for utilization due to several types of losses that first has to be deducted. These losses can affect the performance from two sides. They can limit the time available for loading a resource but also limit how much of the loading time that can be used for value adding operations (Slack et al., 2010). The latter is covered in the section *Overall Equipment Effectiveness* further down.

System losses occur due to how the production system is designed. It contains aspects that affect the available loading time such as unplanned lack of labor in terms of sick leave and vacations, and allowances in terms of small breaks and bathroom visits. These losses are hard to accurately calculate and are therefore often estimated and treated as a standard deduction (Ellegård, 1992). Furthermore, machine stoppage caused by planned or unplanned maintenance has to be deducted from the available capacity. The magnitude of these losses can greatly vary depending on the type of machine and the environment in which it operates. Planning decisions regarding deductions caused by maintenance must thus be based on past performance and experience (Schönsleben, 2007).

The capacity that remains is called *rated capacity* and should be available for loading. However, *demonstrated capacity* based actual past data and experience is often distinct from rated capacity and is usually seen as the load limit. The difference is used as *protective capacity* and should be able to protect the resource from variations, complexity effects and extreme events that would otherwise result in exponentially increasing lead times as discussed in section 2.1.2 (Schönsleben, 2007; Slack et al., 2010).

Resources that constrain the performance of the system can be loaded up to its *demonstrated capacity* while loading a non-constraining resource beyond its *productive capacity* will not increase the throughput of the system but only result in excess inventory, i.e. waste (Cox & Schleier, 2010). The difference between the *productive capacity* and the *demonstrated capacity* for a non-constraint is called balancing loss, i.e. capacity that should not be used due to that the system is constrained by another resource (Slack et al., 2010). This is covered in more depth in the following sections. Definitions above are depicted in Figure 2.

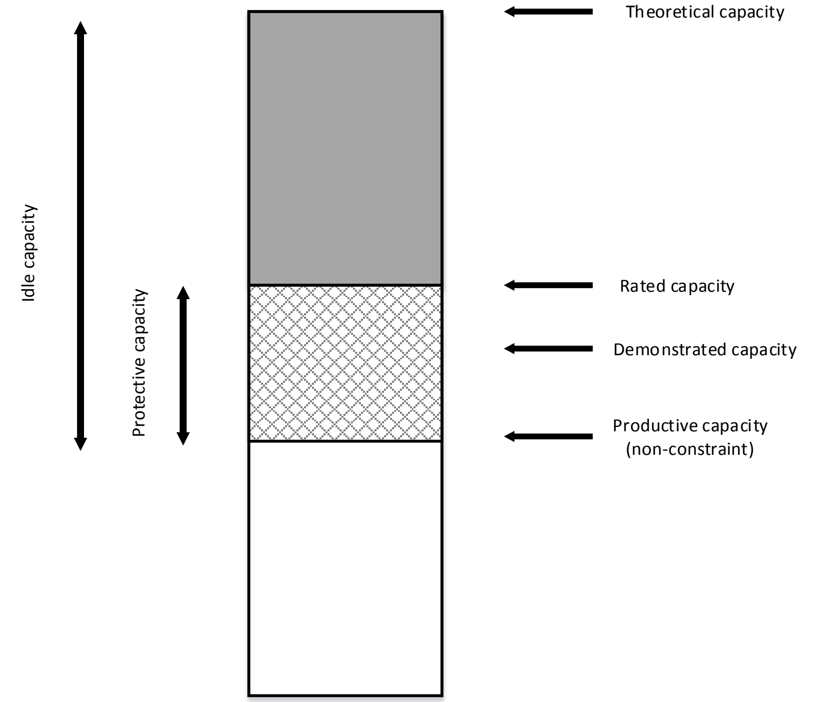

Figure 2: Resource capacity measurements.

Capacity efficiency is a measure of how much of the theoretical capacity is available for utilization and is defined as below, where *effective capacity* is either *productive capacity* or *demonstrated capacity* depending on whether the resource is a constraint or not (Slack et al., 2010).

$$Efficiency = \frac{Actual\ output}{Effective\ capacity} \quad (10)$$

Overall Equipment Effectiveness

A further measurement in the capacity analysis is the *Overall Equipment Effectiveness, OEE,* which determines how much of the loading time is used for value adding operation, and is defined as follows (Slack et al., 2010).

$$OEE = Availability\ rate \cdot Productivity\ rate \cdot Quality\ rate \quad (11)$$

$$where$$

$$Availability\ rate = \frac{Total\ operating\ time}{Loading\ time} \quad (12),$$

$$Productivity\ rate = \frac{Net\ operating\ time}{Total\ operating\ time} \quad (13)\ and$$

$$Quality\ rate = \frac{Valuable\ operating\ time}{Net\ operating\ time} \quad (14)$$

The availability rate measure deducts availability losses such as set-up time, handling time and unplanned work stoppage, distinct from maintenance stoppage mentioned above. Performance rate covers speed losses such as equipment running below its stated speed. Beside poor performing machinery, variations in the skill level of the operators will result in operating times that varies around a stated average. Planning has to be based on the slowest operator to avoid problems, and thus performance is lost (Ellegård, 1992). Time invested in a unit that later has to be scrapped or reworked due to quality issues is lost and these losses are covered by the quality rate (Slack et al., 2010). This can be seen in Figure 3.

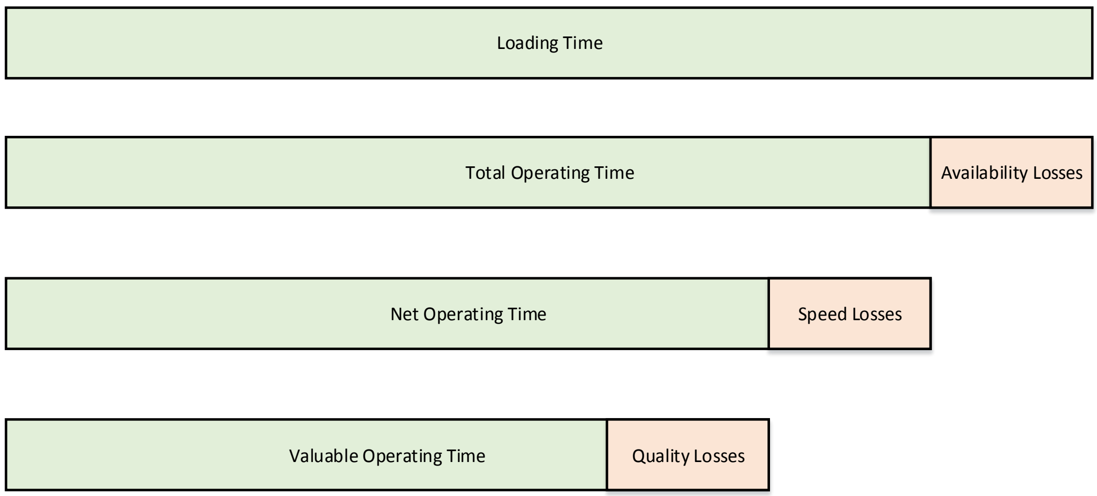

Figure 3: Visualization of OEE measurement.

OEE improvement of a resource that constrains the system performance will increase the throughput of the system through a reduction of the cycle time in accordance with *Little´s law*, stated above (Slack et al., 2010).

2.2.2 Theory of Constraints

The concept *Theory of Constraints* was developed by Eliyahu Goldratt in the 1980s as a competing approach to *Materials Requirements Planning, MRP & MRPII,* and *Just-In-Time, JIT*. All three methods were meant to challenge old assumptions and achieve a sustainable competitive advantage in a changing environment where time and product quality became key competitive factors (Rahman, 1998). *TOC* evolved from Goldratt´s own logistical system *Optimized Production Timetables*, *OPT*, with the objective to reduce throughput time and increase inventory turnover, and thus meet the challenges of the new business environment (Goldratt, 1988). The trend with an increasingly rapid changing business environment has continued to current date (Chopra & Meindl, 2007), which implies that the theory behind *TOC* still is valid.

Philosophy

TOC is based on the realization that a system must have at least one constraint that limits its production capacity (Goldratt, 1988). If this was not the case a commercial organization would be able to achieve unlimited profit, which is infeasible (Rahman, 1998). A constraint is, according to Goldratt (1988, p.453), "*Anything that limits a system from achieving higher performance versus its goal"*. This definition makes it clear that a constraint can be made up of not only production resources but also by anything that influences the production system (Cox & Schleier, 2010). Beside physical constraints, such as material availability, people and machines, the system performance can be limited by managerial constraints, such as non-optimal procedures, policies and methods, and the market demand (Schönsleben, 2007).

A company that wants to increase its throughput must direct its improvement efforts towards the constraint that limits the performance. This since increasing the capacity of a non-constraint will not increase the performance of the system. Above reasoning leads on to a central aspect in *TOC,* which is to focus all efforts on the constraint in order to improve the system performance (Cox & Schleier, 2010). If the effort is successful the constraint is elevated and one of the previous non-constraints pose as the new system limit, making the working principle behind *TOC* an iterative process.

Principle

The working principle of *TOC* is a cyclical process built on continuous improvements similar to the P-D-C-A cycle of *Lean Production* and the D-M-A-I-C of the *Six Sigma* methodology (Rahman, 1998; Bergman & Klefsjö, 2010). The *TOC* cycle consists of five steps that will be described below and can be seen in Figure 4.

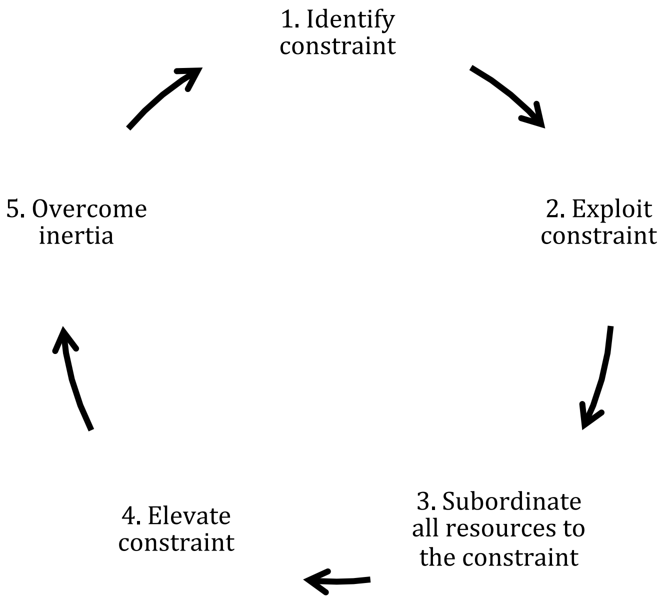

Figure 4: The five steps in the TOC cycle.

[1] The first step is to identify the constraints of the system in order to be able to prioritize and direct improvement efforts (Goldratt, 1990). Constraining production resources can be identified using different methods depending on the requested accuracy. Based on past data, the resource that had the longest work queue over a certain period of time can be identified as the system bottleneck. Identification of the

system bottleneck is, beside the first step towards improving system performance, also the base for production control. However, for performing production control tasks at a detailed level, information regarding past constraints is not always sufficient (Schuh et al., 2012). This issue will be covered further in section 2.2.2.

[2] When the system constraints have been identified it has to be decided how to exploit them. A physical constraint can be exploited by either adding resources of that type or making the use of the existing resources more effective and efficient (Rahman, 1998). Improving resource efficiency implies increasing its *value-added ratio,* and thus its *OEE*, e.g. by reducing set-up time, cross-training workers and making process improvements to improve quality (Hopp & Spearman, 2011). However, all constraints are always subordinate to the market constraint, meaning that nothing that is not in demand should be produced. Scheduling work to a capacity constrained resource, *CCR,* just to increase its utilization only builds inventory and might waste capacity that could have been used for other products for which there is an actual demand (Schragenheim et al., 2009). Managerial constraints should not be exploited but eliminated in favor for policies and methods that support throughput increase (Rahman, 1998).

[3] All non-constraining parts of the system should be utilized in a way that support an optimal utilization of the constraint. Utilization of a non-constraint beyond its productive capacity will not increase the throughput of the system but only build excessive inventory and should therefore not take place (Rahman, 1998; Lockamy & Cox, 1991).

[4] Efforts can now be made to elevate the constraint and increase its performance. A performance increase of a *CCR* will increase the performance of the system since more of the *non-CCR* capacity can be used for productive purposes (Goldratt, 1990).

[5] The last step is to ensure a cyclical process. When a constraint is elevated a previous *non-CCR* will become the new system bottleneck and the organization must start over at step one of the cycle. Environmental changes, such as changing product mix and market demand levels, implies that constraints might move around the system making past decisions invalid which also creates a need for the process to be frequently run through (Schuh et al., 2012).

2.2.3 Bottleneck dynamics

TOC assumes static constraints, meaning that a *CCR* is assumed to remain a constraint until above cycle is run through and the constraint has been elevated. As mentioned above, the constraints can move around the system due to environmental changes and internal issues (Shen & Chen, 2010). Frequent runs through the cycle can to some extent capture these changes, but it is still a discrete process. For a system exposed to large complexity and variations, in terms of e.g. product mix, production volumes and product flow, production control decisions need to be made at the same time as the *CCR* is identified, which requires a continuous real-time method (Schuh et al., 2012).

The issue with shifting bottlenecks can thus be approached from two directions, either by reducing the system complexity or using a real-time method for identifying constraints. There is no comprehensive method for real-time identification of bottlenecks to current date (Shen & Chen 2010; Schuh et al., 2012), which implies that the second approach requires a relatively complicated combination of several tools, such as mathematic modeling and simulation. Measures that enable use of simple rules and methods, i.e. the first approach, are therefore recommended (Schuh et al., 2012).

2.3 Production Planning & Control

The main aspect of *TOC* is as mentioned to focus on the capacity constrained resources and to subordinate all other decisions to how the constraints should be elevated. There has to be a system in place with planning and control techniques and methods that support this approach. This chapter will cover the *Drum-Buffer-Rope* logistical system of *TOC* and how different types of manufacturing situations and customer demand impacts the set-up of the production system and the planning and control procedures.

2.3.1 Manufacturing Situation

A production system consists of a set of interdependent and serial connected production resources with intermediate buffers. The time it takes for a certain product to pass through this system from start to end is makes out the total lead time (Slack et al., 2010). The ultimate goal of every production system is to deliver perfect customer service, in terms of the right quality at the right time to the right price, and doing this by using as little resources as possible, thus maximizing the profit (Bergman & Klefsjö, 2010; Chopra & Meindl, 2007; Slack et al., 2010). This implies that producing a volume that deviates, in both directions, from the true market demand can be regarded as waste and should be avoided. A production volume higher than demanded is a waste of resources while a volume less than the market demand is a waste of sales opportunities (Schönsleben, 2007). Thus, the perfect situation is to be able to start the production, or even procurement, only when there is an actual and firm customer order (Chopra & Meindl, 2007). For this to be possible, the customers must tolerate a waiting time equal to, or longer than, the total lead time of the production system. This is rarely the case so the point that represents the maximum tolerable waiting time will lie somewhere within the total lead time. This point in time is called *Customer Order Decoupling Point, CODP,* and from that point on a production order is directly tied to a specific customer order (Olhager, 2010). This is described in Figure 5.

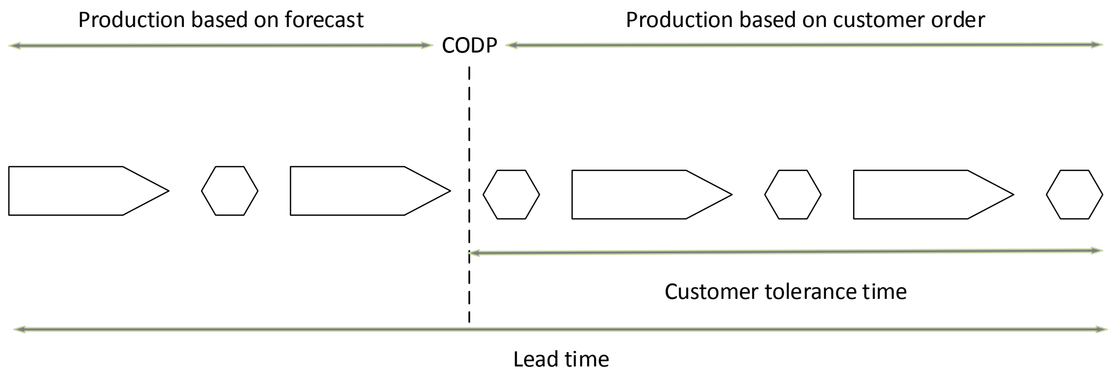

Figure 5: Customer order decoupling point in lead time.

Thus, dependent on characteristics of the customer base and the lead time of the production system, firms can find themselves in either of four general manufacturing situations, *Make-to-Stock, Assemble-to-Order, Make-to-Order, Engineer-to-Order* or the special case *Make-to-Availability,* which is a combination of *MTS* and *MTO*.

Make to Stock

In this situation the entire planning and control procedure has to be made based on forecast of future demand since the customer expects to be able to get the product delivered at the same time as their demand arises. The production system must therefore be set up in a way that ensures that stock levels are sufficient for serving demand (Schönsleben, 2007).

Assemble to Order

This situation is a special case of the *MTO* situation for production systems where assembly is a significant part of the lead time. In this situation assembly lies within the customer tolerance time and will not be started before a firm customer order is received (Schönsleben, 2007). Fabrication or procurement of ingoing components and sub-assemblies has to be made based on forecasts and be available for assembly. The last stock-keeping point before delivery lies thus just before the assembly (Schragenheim et al., 2009).

Make to Order

The last stock-keeping point for this situation is the raw-material warehouse, meaning that all operations in the production system can be made within the customer tolerance time. This implies that no production has to be made based on forecasts, which is the optimal situation in accordance with the reasoning above (Schragenheim et al., 2009).

Engineer to Order

This is an extreme case of the *MTO* situation where even the development of the product lies within the tolerance time. This situation is mostly found in high-tech industries such as military and aerospace.

These situations are presented graphically in Figure 6.

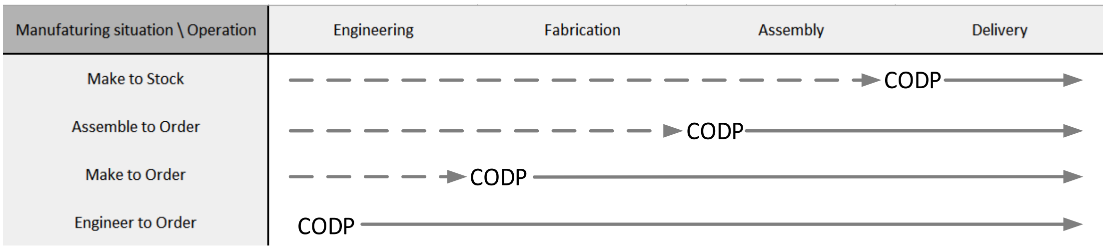

Figure 6: Visualization of different manufacturing situations.

Make to Availability

Both *MTO* and *MTS* has the objective of generating as high return on investment in production resources as possible, but there is a clear conflict between the approaches in terms of the assumptions and logic they are based upon (Cox & Schleier, 2010). The logic behind *MTS* is that high resource utilization will generate a high return on investment, i.e. if the resource is never idle the firm captures the entire value potential of that investment. However, for this to be true there are a number of assumptions that must hold (Schragenheim et al., 2009).

1. If no production with the purpose of filling stock took place there would be a lot of idle time in the production system.
2. Everything that is produced can always be sold.
3. Everything that is produced can always be sold at full price.
4. The only way to achieve high resource utilization is making to stock.

Assumption [2] and [3] can never at the same time be completely true in a competitive market since supply and demand, for a certain price-point, in such a market is always in equilibrium, i.e. the production volume is equal to the demanded volume at all times. In order to sell more, the price has to be lowered which is a contradiction of assumption [3] (Mankiw, 2014).

MTO is based on the opposite assumptions. Customer demand is uncertain and excess finished inventory will remain unsold or be significantly marked down (Schragenheim et al., 2009), which is in line with the macroeconomic reasoning above. To avoid waste in terms of missed sales opportunities and excess inventory, the optimal solution is, as mentioned above, to only produce to fill an actual customer order, i.e. *MTO*. The issue with *MTO* is that it is not always possible since the delivery time is limited by the customer tolerance time.

Make to Availability is based on the same assumptions as *MTO* but adjusted for the case where part of the production has to be made without a firm customer order. As little as possible should be produced based on forecasts and the unavoidable forecast-based production should be tied to ensure availability of a certain product to a certain customer (Cox & Schleier, 2010). The main difference between *MTS* and *MTA* is thus the base for the forecasts. While the forecast for *MTS* is based on experience, past data and trends, *MTA* forecasts requires a closer relationship with the customers and are based on a customer blanket order (Schragenheim et al., 2009). A blanket order is an indication of what the customer expects to order in the future. It contains a maximum and minimum number of units that the customer undertakes to order. This potential quantity gap narrows as production start date approaches and ends up in a firm customer order (Schönsleben, 2007). The producing company undertakes to deliver the quantity stated in the blanket order and uses it to plan the production outside of the customer tolerance time (Chopra & Meindl, 2007).

According to Schragenheim et al. (2009) *TOC* has historically been implemented in mainly *MTO* environments, since those situations fully appreciate that the market constraint is the superordinate system constraint, but can also be implemented with similar benefits in *MTA* situations. This since the logic behind *MTA* supports the central rationale in *TOC* to avoid overproduction when possible. *MTA* can thus be seen as a superior approach to *MTS* in situations where the customer is known and a relationship can be built, which is the case for most *B2B* transactions (Cox & Schleier, 2010).

With these premises it is possible to understand how the *TOC* approach and thus a throughput focus can be supported by planning and control procedures and the set-up of the production system. This is covered in the next section.

2.3.2 Planning & Control System

As established above, the ultimate situation is to be able to produce only when there is a firm customer order available. What hinders companies from reaching this situation is that their lead time is longer than the delivery time the customers can accept. This realization is the core of transforming the thinking behind *TOC* into an industrial application through a planning and control system that appropriately manages the system bottlenecks. The focus of this system is to increase throughput through a reduction of inventory, *WIP*, and thus a reduction of the lead time (Schragenheim et al., 2009). This relationship is self-inducing since the shorter the lead time, the bigger part of the production that can be made to fill a firm customer order. This leads to smaller deviations in production volumes from the true demand and therefore less waste in terms of over- or under- production (Schönsleben, 2007). In turn, this leads to less need for buffers and safety stock, i.e. less *WIP*, and at length shorter lead times in accordance with *Little´s law* mentioned in section 2.1 (Slack et al., 2010). Moreover, the shorter the lead time, the faster the production system can respond to changes in demand. The faster the system can respond, the shorter the forecast horizon for the part of the production that lies beyond the *CODP* will become and the more reliable the forecast will be (Cox & Schleier, 2010). More stability and predictability in the production system will imply that simpler planning and control methods can be used and the more transparent the system will be (Schuh et al., 2012). This is especially important in production systems with complex operations and product flows (Schönsleben, 2007; Chopra & Meindl, 2007).

Appropriate management of bottlenecks through application of *TOC* to production planning and control will not only increase throughput but also reduce working capital and throughput time variation and improve adherence to delivery dates (Schuh et al., 2012). Thus, it will improve the company result by both supporting top line growth as well as improving the bottom line (Baumol & Blinder, 2011).

The industrial application of *TOC* is called *Drum-Buffer-Rope*, *DBR*, and was introduced by Goldratt (1984). It was later simplified and renamed to *S-DBR*, making it even more suitable for complex production systems (Schuh et al., 2012). This method is covered in depth in the following section while the differences and similarities to *DBR* will be pointed out briefly throughout. The later parts of this section will cover a comparison between *S-DBR* and competing methods such as traditional *MRP* and the Japanese *JIT*.

S-DBR Principle

S-DBR, like most modern production philosophies, e.g. *JIT*, facilitates the transition from a traditional pushing production to a pulling product flow. This implies that a work station should only be active when there is a need for material further down the production chain, i.e. customer demand initiates all production (Schragenheim et al., 1994). This is also applicable for other situations than *MTO*, but then it is consumption from the finished goods stock that initiates production (Cox & Schleier, 2010). The production chain consists of a *Drum* that sets the production pace, a *Rope* that connects all parts of the production chain to each other and the market, and a *Buffer* that protects the system from disruptions and ensures a timely delivery. The *Drum* is the system constraint, either an active *CCR* if a production resource has less capacity than demanded by the market, or otherwise always the customer demand. The *Rope* is the production planning that pulls material through the production chain in the pace that is set by the *Drum*. The *Buffer* is the production control that ensures material is available to be connected to the *Rope,* see section *S-DBR Control* (Cox & Schleier, 2010; Schragenheim et al., 1994; Schragenheim et al., 2009). The S-DBR set up is visualized as in Figure 7.

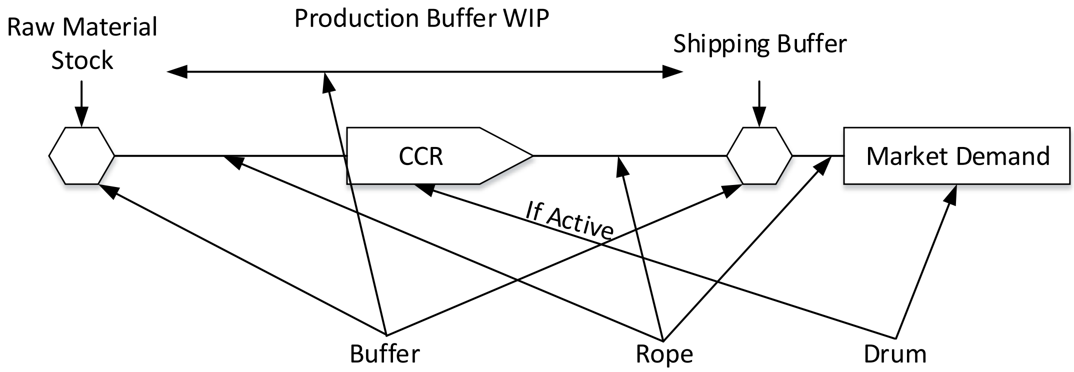

Figure 7: Conceptual visualization of the S-DBR production system.

In accordance with the *TOC* principle all resources in the production system are subordinated to the system constraint. A key difference between *S-DBR* and traditional *DBR* is that the market constraint is considered as the only real constraint in *S-DBR*, which implies that there is no point in making sure that the *CCR* never is starved unless there is an actual demand (Cox & Schleier, 2010). This means that there is no need for a protective buffer before the *CCR* as there is in *DBR* since the shipping buffer together with the production buffer will make sure that the market demand is never starved, which is of sole importance (Lee et al., 2010).

When it comes to the production planning, the *CCR* is however of great importance, so in the following section the planning procedure is described as if the *CCR* is active. Furthermore, the set-up of an *S-DBR* system is highly dependent on the manufacturing situation. As it has been established earlier that *MTA* is superior to *MTS* in most *B2B* transactions and that *MTO*, although optimal, is not possible for all companies, *MTA* will be covered in depth while comparisons with *MTO* are made throughout.

S-DBR Planning

Assuming that everything the production system can produce will be consumed, i.e. the *CCR* is active and thus the *Drum*. This, following the principle of *TOC,* means that all other parts of the production system should be subordinated to the pace of the *CCR* and at length that the *CCR* is the only planning point in the system. The objective of *TOC* is to have as little *WIP* in the system as possible, i.e. all orders released onto the shop floor will be finished within one standard lead time promised to the customers. A reasonable planning horizon for the *CCR* is thus just one standard lead time (Cox & Schleier, 2010). In contrast to traditional *DBR* where the *CCR* is planned in detail with a fixed production schedule, the *S-DBR* planning of the *CCR* does not take the order of the units into consideration. Each production order is represented only by the time it will consume at the *CCR*, the time will be added to the currently existing planned production orders at the *CCR*, i.e. the planned load, until the planned load reaches the standard lead time (Lee et al., 2010). The *CCR* is however never loaded to its maximum capacity in accordance with the reasoning in section 2.2.1. The maximum planned load at all times is thus a percentage, e.g. 80%, of the standard lead time. This can be visualized as in Figure 8.

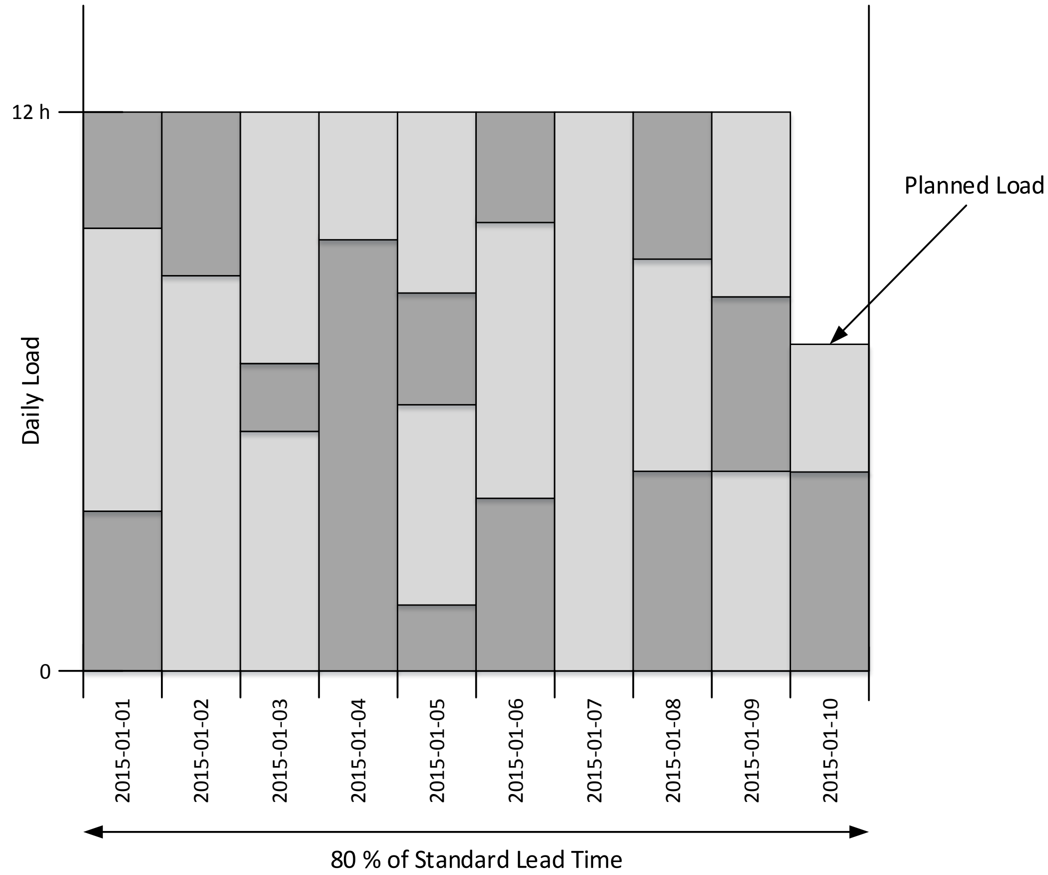

Figure 8: Visualization of how individual jobs are loaded onto the CCR.

In the example above the loadable part of the standard lead time, i.e. 80 %, is ten days and the CCR is in operation 12 hours per day. Each bar represents an order that has been planned and released onto the shop floor. This is a finite loading technique, meaning that the capacity of the work station is considered fixed and cannot be adjusted in the short term. This is the case when expensive and complicated machinery is needed or when labor is specialized and requires special training. If this is not the case and the capacity is flexible, work is loaded onto the work station regardless if it exceeds the current capacity, the capacity will be adjusted accordingly. The tradeoff when deciding which approach to take is whether the capacity or the delivery due dates is the most flexible and the decision is thus context-dependent (Schönsleben, 2007).

It is rather clear that there is no need for a buffer before the CCR since it is reasonable to assume that the time it takes for a production order to get from the start of the production chain to the CCR is less than 80 % of the standard lead time, regardless of where in the production chain the CCR is located. This since the CCR is the weakest link in the production system and all other resources will have excess capacity and thus no queue (Schragenheim et al., 2009). The first rule for releasing orders onto the shop floor is thus to check whether there is room for the order at the CCR so that the planned load does not exceed the load limit

(Schragenheim, et al., 1994). Breaking this rule will only increase *WIP* in the system and thus create waste (Goldratt & Cox, 1984).

S-DBR Control

Since the production schedule for the *CCR* in *S-DBR* is neither planned nor fixed, there has to be a method for prioritizing among the production orders that arrive at the work station. Clearly there are some orders that are more urgent than others. In traditional *DBR* this is determined beforehand as the production order sequence is planned onto the *CCR* (Goldratt & Cox, 1986). The issue with this approach is that between the points in time when the orders are assigned a time slot at the *CCR* and when they arrive at the *CCR*, something might happen that changes the demand or somehow creates a need for reprioritizing orders. Since the goal with the system is to be as responsive as possible, this cannot be the optimal solution (Schragenheim et al., 2009). In *S-DBR*, the production control is instead done through buffer management. Buffer management is made up of three steps, which are presented in Figure 9, and each step is then moreover described.

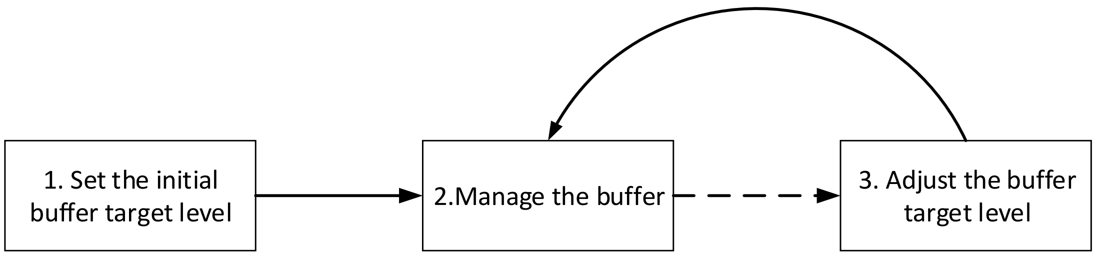

Figure 9: The three steps of buffer management in S-DBR.

[1] The buffer in the system is referred to as the *production buffer* and consists of two parts; finished orders in the shipping buffer and orders that are somewhere on the shop floor, i.e. *WIP*. The state of the *production buffer* controls both production initiation as well as the priority for individual production orders, which implies that the predetermined size of the buffer is paramount (Schragenheim et al., 2009).

For an *MTO* situation, where nothing is produced unless to a firm customer order, there will be no physical stock. The *production buffer* in such a situation is made up of time, i.e. the lead time of the system with an additional safety buffer. The safety buffer implies that the quoted lead time is somewhat longer than the actual lead time of the system and protects the system against failing to meet delivery due dates caused by disruptions and variations in the production chain. The initial challenge is to set a reasonable lead time (Cox & Schleier, 2010). The net operating time is just a fraction of the total lead time due to the unnecessary high *WIP* levels when using traditional planning and control methods. Several researchers show that cutting the current lead time in half is a reasonable initial production buffer level, when transitioning from traditional methods to *S-DBR* (Schragenheim et al., 2009; Schragenheim et al., 1994; Cox & Schleier, 2010). As mentioned earlier, the relationship between *WIP* and lead time is self-inducing, i.e. lead time reduction will imply that orders are released later onto the shop floor which reduces the *WIP*. The new *WIP* level supports the shorter lead time, all according to *Little´s law*.

However, the *MTA* situation does imply that the buffer consists of physical product units and setting the appropriate level for the *production buffer* is more challenging and not as straight forward as for *MTO*. Since customers do not accept to wait for delivery as long as the lead time of the production system, a first approach is that the demand during the replenishment time, i.e. lead time has to be satisfied from the buffer (Cox & Schleier, 2010). The challenge is thus to determine a demand during the replenishment time that ensures no starvation of the market and still a reasonable *WIP* level. Since *MTA* implies a commitment of availability to a specific customer the forecast is based on customer blanket orders over time. The quantity of those orders varies but the system has to be able to handle all eventualities and the *production buffer* should be sized accordingly. The replenishment time is not fixed either, but fluctuates due to variations and disruptions. The appropriate *production buffer* level can thus be calculated as below (Schragenheim et al., 2009; Cox & Schleier, 2010).

$$Target\ level = Max\ sales\ within\ \emptyset\ replenishment\ time \cdot f_u \quad (15)$$

Where f_u, the uncertainty factor for the replenishment time, is a measurement of variation and can be based on past performance of the system.

[2] For controlling the production, the state of the production buffer is a comparison between the current level and its target level and will guide all decisions on the shop floor. The buffer state is calculated as per below (Schragenheim et al., 1994).

$$Production\ buffer\ status = \frac{Target\ level - Finished\ goods\ stock - WIP}{Target\ level} \quad (16)$$

The buffer status is divided into three state categories, green, yellow and red, and is described by Figure 10.

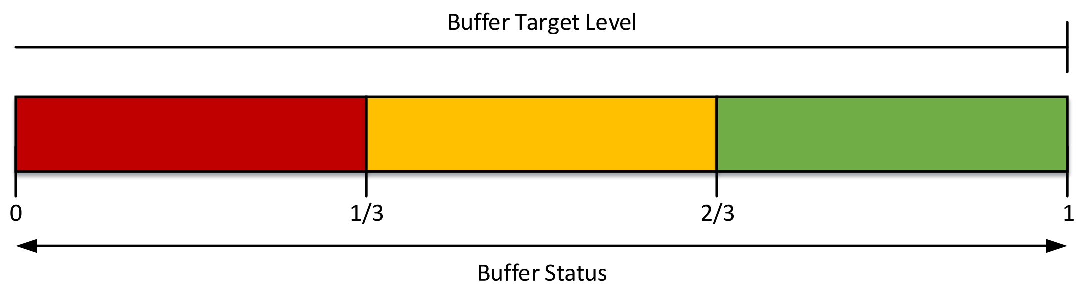

Figure 10: Buffer status levels.

If the *production buffer* status for a product is in the green area there is no risk for a stock out and starvation of the market and replenishment need is not urgent. The yellow area is considered normal. These orders should be monitored but does not require any immediate actions. Production orders that lay within the red zone face an imminent risk of failing to meet the market demand and require immediate management attention and actions. For *MTO* situations, the penetration is measured in time instead of numbers of units as for *MTA* but the principle is the same (Schragenheim et al., 2009; Cox & Schleier, 2010).

For prioritizing among orders at the *CCR* the rule is simple, yellow orders are handled before green ones and red orders are always highest priority. When two or more orders have the same buffer status, the operator at the work station has to make the decision. This decision should take into consideration if handling one of the orders before the others could imply higher efficiency in terms of lowering the total amount of set-up by batching, and should be done so if that is the case (Schragenheim et al., 1994). Each order has a calculated measure of its *production buffer* penetration so even if the orders lie within the same category there can still be a significant gap between their penetrations, which should be taken into consideration. These types of decisions are however always subordinated to the above rule (Schragenheim et al., 2009). Here is another clear advantage of using *S-DBR* with only one buffer over traditional *DBR* that has a separated buffer in front of the *CCR*. The issue of having multiple separated buffers is that every order gets more than one buffer status, one for each buffer. There is thus a possibility of having a red production order at the *CCR* while the same order is green at the shipping buffer. There is no immediate threat of starving the market but this order will be prioritized at the *CCR* and maybe over an order that faces an immediate threat, i.e. red at the shipping buffer, but that is green or yellow at the *CCR* (Schragenheim et al., 2009). The *production buffer* status is updated in real-time and together with the prioritization rule at the *CCR,* the *S-DBR* system is also more flexible and responsive to changes than *DBR* as mentioned above (Lee et al., 2010).

The level of the *production buffer* should be aimed at the target level and when consumption of the shipping buffer lowers the buffer status to below 1, a new production order should thus be released. This is the second order release rule but is always subordinate to the first rule mentioned in the planning paragraph, to only release an order if the *CCR* planned load is less than the maximum load (Schuh et al., 2012).

Companies often face a situation where some products are made to order and some that has to be made to availability (Cox & Schleier, 2010). With the *S-DBR* system, this is however not an issue. Regardless if the buffer is made up of time or actual products, the penetration and prioritization at the CCR are treated the same, a red order has always the highest priority (Schragenheim et al., 2009).

[3] Regarding the *production buffer* status, stagnation is considered a sign for caution since a buffer should be buffering, i.e. fluctuating. If this is not the case the target *production buffer* level should be adjusted. The ideal situation is a *production buffer* that fluctuates between green and yellow penetration since it implies a low risk for starvation as well as a reasonable *WIP* level (Cox & Schleier, 2010). A stable green penetration indicates that the company carries too much inventory and the target levels should be adjusted downward. A product that resides in the red zone over a longer period of time indicates that the *production buffer* level for this product type should be increased (Lee et al., 2010). This should however be done with caution since increasing the *production buffer* implies increased *WIP* levels. If this is done when the planned load of the *CCR* approaches or exceeds the limit, this only builds inventory, which will increase the replenishment time causing more orders to enter the red zone, worsening the situation (Schragenheim et al., 2009).

S-DBR and MRP

The production in both systems is triggered by some demand at the end of the production chain. In an *MTO* situation, this demand is represented by a due date of the customer order and for *MTA* it is a need for replenishment. This demand is then inherited up-stream in the production system (Schönsleben, 2007). The difference in this regard between the systems is that in *MRP* the generated production order is split up in a number of intermediate orders, one for each ingoing component (Steele et al., 2005). These different orders are then assigned individual due dates for when the component is needed in the production. A component can be shared between several products and production orders for these components will be combined into a single order for the common component, which implies a significant loss of transparency in the system, a situation that is worsened by complicated batching policies (Steele et al., 2005). The multiple due dates also create confusion after being released onto the shop floor since there is no way of knowing if an order is on time or not. In *S-DBR* there is only the final due date, as soon as the buffer status is below 1 a single production order for ingoing components is generated and released, and the buffer management makes sure that this date or replenishment need is met (Cox & Schleier, 2010).

Order release in *MRP* is done by only establishing a latest date for release, which opens up for releasing orders early if there is room at the gate resource, i.e. the first resource in a product flow. This implies unnecessary high levels of *WIP* and a risk for a situation where a critical order is delayed at a resource because the resource is occupied by a non-critical order (Steele et al., 2005; Schragenheim et al., 2009). With *S-DBR* each released order has a time slot at the *CCR* and are moved through the production chain as quickly as possible. *MRP* also fails to recognize that capacity is finite which implies a risk for overloading the system (Schragenheim et al., 2009).

It is possible to ensure that a *MRP* system will not overload the shop floor by attaching a *CONWIP,* Constant-Work-In-Process, system to it. The logic behind a *CONWIP* system is that a production order for a certain product will not be released until the market has consumed a unit of that specific product, thus keeping *WIP* constant in the system. The drawback with such an approach is that it adds even more complexity to the already complicated *MRP* system (Hopp & Spearman, 2011). This implies a loss of transparency, which can hamper improvement efforts and make the system difficult to overview. Furthermore, buffer management in *S-DBR* gives a clear upfront indication when the *WIP* levels need to be adjusted while with the *CONWIP* system, this will be noticed when the effects of wrong *WIP* levels already has occurred (Cox & Schleier, 2010).

S-DBR and JIT

The *DBR* concepts and *JIT* are similar in many regards. They both focus on the product flow and aims to reduce inventory and lead time to make the production as flexible and responsive as possible (Schragenheim et al., 2009). The key difference is that *JIT* states that all inventory is harmful and eliminating it is the ultimate goal, while *S-DBR* recognizes that some inventory, although very limited, is necessary in order to protect the system against variations. *JIT* assume that process and flow variations can be minimized by applying *Six-Sigma* techniques (Huang, 2002). However, it is rarely possible to eliminate all special-cause variations, regardless of the *Six-Sigma* efforts the company undertakes. This is especially prominent in situations where specification tolerances are very tight. There is also a point where the cost for defect prevention exceeds the cost for quality defects (Schragenheim et al., 2009).

The planning and control system of *JIT* is called *Kanban* and is considered to facilitate pure pull production. Production initiation throughout the production system is controlled by e.g. cards, that the downstream work station detach from the bin of the components they need and send to the upstream station (Hopp & Spearman, 2011). This card initiates the production in the upstream station to fill the buffer between the two stations. The size of such buffers and thus the *WIP* level of the entire system is determined by the number of cards in circulation, which is constant until a decision is made to change this number (Schönsleben, 2007). However, the lowest amount of cards possible for the system to work is one card between each work station. For production of slow moving products made in many steps this method implies unreasonable high *WIP* levels relative to the production volume (Slack et al., 2010). *Kanban* and *JIT* are thus suitable for production of few production variants in large volumes and few steps, but not for low volume production.

The *DBR* concept takes a more holistic view aiming to optimize the overall profitability of the firm and also recognizes that the market demand can be affected by internal efforts (Schragenheim et al., 2009). When the internal constraints are elevated and the market constraint is the active bottleneck limiting the system performance, the next step in the *TOC* cycle is to try to elevate this constraint, increasing the throughput of the system. In this situation, the logic behind *JIT* is that the higher efficiency of the internal production system enables a reduction of the resources employed (Huang, 2002). The end result is that the *S-DBR* system will generate higher throughput than a *JIT* system in the same situation (Cox & Schleier, 2010).

Demand Driven MRP

There is however one aspect that speaks in favor for an *MRP* approach compared to both *S-DBR* and *JIT* and that is the completeness of the system. The two more pull-based approaches lack tools for materials planning, i.e. breakdown of products into components and ingoing raw material and generation of procurement orders, which *MRP* does with ease (Mula et al., 2012). A solution to this dilemma is to simply combine the production planning and control logic of the pull-based approaches with the material requirements planning of a traditional *MRP*. This combined approach has recently been developed under the name *MRP IV* or *Demand-Driven-MRP*, *DDMRP*, which utilizes the pull-based production logic of *S-DBR* or *JIT* and keep the superior handling of bills of materials and raw material supply of the traditional *MRP* (Plossl & Orlicky, 2011). This was the initial purpose of *MRP,* which was later extended through *MRPII* to a system that covers the entire operation of an enterprise, *ERP*. This is a very complex and complicated system with low visibility that created the original need for development of simpler and more transparent approaches such as *DBR* and *JIT* (Smith & Smith, 2013).

2.4 Layout and Product flow

The full potential of an optimal planning and control approach can only be realized if it is supported by an appropriate production layout. This chapter covers the physical layout of the production as well as the allocation of products to machines or work stations. The layout decision is dependent on the product type, manufacturing situation and the environment in which the company operates. The different contexts and corresponding layouts are therefore briefly presented first. It is then described how the scope of appropriate layouts for a certain context can be widened by applying virtual techniques.

2.4.1 Production Layout

Which production layout that is appropriate for a production system is determined by the production volume and product variety, resulting in product flows ranging from intermittent to continuous. The scales are continuous which implies vague context boundaries and an overlap of suitable layouts, as in Figure 11 (Slack et al., 2010).

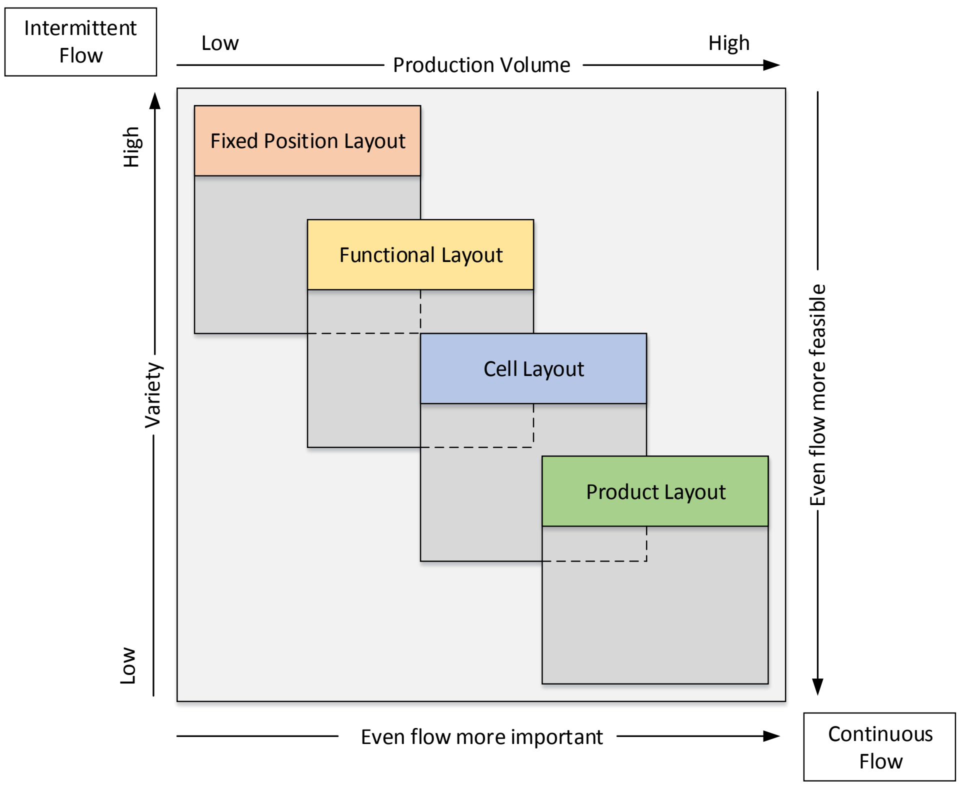

Figure 11: Production layout matrix (Slack et al., 2010).

Fixed Position Layout

This layout is suitable for products that are impossible to move. The product stays in one location and the production resources are brought to that location in the order they are needed. This layout is flexible and typically suitable for *ETO* situations where each product is unique and only made in one copy (Schönsleben, 2007).

Functional Layout

Machines and other production resources are in this layout organized and grouped together based on their functions, e.g. all lathes are located in one area and the milling machines in another. The products are moved between the groups in the order it needs the different functions (Schönsleben, 2007). This layout is suitable for moveable products, when a large range of product variants is needed, the volume for each variant is fairly low and in typically *MTO* situations (Slack et al., 2010). This layout enables high equipment utilization and great flexibility but has a series of drawbacks. Since products are moved back and forth between the different groups, handling and transport costs are high. The complex product flows also require a complicated planning and control system, since every machine is its own planning point, and high *WIP* levels to ensure availability. All this leads to long lead times and high working capital investments (Prince & Kay, 2003).

Cell Layout

This layout implies that machines and other production resources are grouped together in cells to serve a specific product flow. Products and components are grouped into families by what operations they require using *Production Flow Analysis*, which is explained in the following sections. Machines and production resources are then assigned to such a product family and will serve only that family (Schönsleben, 2007). By doing so, the product flow complexity will be drastically reduced since the number of intersections is minimized, which also lowers handling and transport costs. Furthermore, now each cell is a planning point instead of every machine, which implies that simpler and more transparent planning and control systems can be used. Moreover, it implies less *WIP* is needed to ensure availability, which shortens the lead time, according to *Little´s law*, and reduces the investment in working capital (Slack et al., 2010). Thus, a cell layout should be chosen over the functional ditto. However, there are some issues when transitioning to a cell layout. Firstly, machines might be large and difficult to move and the cost of obstructing the production to perform the transition might be significant. Furthermore, a cell layout might imply a need for investing in more production resources, e.g. in a situation when the number of product families that require the same type of resource exceed the number of such resources currently available. This also require that enough space in the facility is available for the new duplicate resources. This situation is especially prominent when the product variety is large. Lastly, an unstable product mix might require that routing flexibility is maintained which will be an issue since moving machines frequently is time- and resource demanding (Prince & Kay, 2003).

Product Layout

This layout is suitable when the production volume is high enough and the range of product variants is small enough to justify the set-up of a single, and often driven, line dedicated to mass production of that individual or standard product. This implies a single planning point for the entire production system and thus is a continuous production flow easily achieved (Schönsleben, 2007).

2.4.2 Virtual Manufacturing

This section explains the concept of Virtual Cell Manufacturing and its further development, Virtual Groups.

Virtual Cell Manufacturing

Virtual Cell Manufacturing, *VCM,* is an alternative to traditional cellular manufacturing where virtual cells are formed temporarily instead of physical cells as in Group Technology, *GT*, (Kannan & Gosh, 1996). The concept was originally introduced by McLean et al. (1983) as an extension of the concept of *GT* by presenting the illusion of a permanent set of assigned resources. McLean et al. (1983) states that *VCM* is a way of improving the performance of cell-based manufacturing systems in unstable production flows and environments by taking control of processes and virtually arrange them into machine pools and virtual cells. Rheault et al. (1996) have a similar reasoning but mean that firms in these turbulent environments traditionally are forced to apply a functional layout, but suggest that *VCM* can be an alternative. Rheault et al. (1996) state that firms experiencing such environments often are small *MTO* firms or subcontractors producing a high variety of parts in variable volumes, which requires a highly flexible and competitive manufacturing system.

Prince and Kay (2003) suggest that the main advantage of *VCM* is that the virtual cells are temporary. When cells have finished processing these jobs, the machines in that cell can all be released and delegated to a new cell. These authors mean that *VCM* therefore can give a similar level of machine utilization as in a job shop layout and eliminate the surplus capacity that frequently exisits in physical manufacturing cells. Kannan and Gosh's (1996) research shows that *VCM* is more responsive to changes in demand and routings than cell- and functional layouts. Furthermore their results suggest that *VCM* has shorter throughput time, improved due date adherence and higher robustness against demand variability. Finally, Kannan and Gosh (1996, p. 519) suggest one of the main advantages of *VCM* is that it "*combines the set-up efficiency typically obtained by GT cellular manufacturing systems with the routing flexibility of a job shop*".

VCM does however seem to have some drawbacks. McLean et al. (1983) indicate that the variable routing needs flexible material handling capabilities and that longer transport might be a result. Prince and Kay (2003) agree with this reasoning and mean this is due to the fact that machines are not physically rearranged. The firm cannot therefore benefit from lower throughput times and reduced *WIP* that comes with an effective material handling.

According to Prince and Kay (2003), Drolet et al., (1991) and Hyer and Wemmerlov, (1982), there is a clear connection between *VCM* and functional layouts. Furthermore, Kannan and Gosh's (1996) research clearly demonstrates that if *VCM* is implemented on a functional layout, it outperforms an application of cellular manufacturing.

Virtual Groups

Virtual Groups, *VGs*, is a further development and a broader view of the VCM concept introduced by Prince and Kay in 2003 (Nomden et al., 2006). Prince and Kay (2003) present the concept of *VGs*, where virtual cells can be applied and implemented onto a functional layout. *VGs* improve the performance of a production system with a functional layout, where a reorganization of machines is not an option due to e.g. high moving costs, investment constraints, large and bulky machines or where routing flexibility must be maintained (Prince & Kay, 2003). *VGs* have gained increased attention during the last years and have been implemented in firms where a physical re-arrangement of machines is not possible.

VGs consist of product and part families and machine groups where (Prince & Kay, 2003) lean and agile production concepts can be applied in order to improve production performance (Krishnamurthy & Yauch, 2007). Meanwhile, it is possible to create a decoupling point between the *VGs* to uphold a required level of buffer inventory (Krishnamurthy & Yauch, 2007). On the other hand, since the machines are unable to move, they are kept in a functional layout, but with *VGs* they are managed as they would in a production system with a product or cell layout (Prince & Kay, 2003).

The main difference between *VGs* and virtual cells in *VCM*, is that *VGs* are not only managed by an ERP- or computer system, *VGs* are instead from a managerial point of view regarded as physical (Prince & Kay, 2003). *VGs* therefore offer more simplified scheduling and planning of production since they exist longer than virtual cells, which are immediately released after finishing its current jobs (Prince & Kay, 2003). The only common denominator of *VGs* and *VCM* is that machines are not physically situated next to each other in the factory (Prince & Kay, 2003). Moreover, *VGs* but also VCM, differs from cellular manufacturing, where a cell traditionally is exclusively assigned to a product family (Slack et al., 2010), since *VGs* allow machine sharing. This provides the possibility to preserve the flexibility advantages offered by a functional layout (Prince & Kay, 2003).

There are both benefits and drawbacks with *VGs*. Since *VGs* allow machine sharing among families, the *VGs* can therefore be both process- and product-oriented, which thereby allows for flexibility (Nomden et al., 2006). However, even if machine sharing can provide higher machine utilization and lower throughput times, it is often not ideal and can result in complex production scheduling (Prince & Kay, 2003).

While *VCM* focuses on managing the process, *VGs* primary focus is the management of products (Balakrishnan & Cheng, 2005). Therefore, Prince and Kay (2003) mean that process managers can be substituted by group managers who will switch focus from managing processes to production of products instead. This give group managers influence of the whole production performance of products and better control of resources.

3 Method

This chapter explains the research approach and analysis methods specific for this project. The purpose of the chapter is to set up guidelines for how the study should be performed, which increase the trustworthiness of the study and makes it possible to recreate.

3.1 Research Approach

There are in general two main research approaches; the deductive and inductive approach (Patel & Davidson, 2003). Furthermore is the third one, the abductive approach, a combination of the two (Wallén, 1996). The deductive approach means that the study takes a starting point in existing theory and uses it to create a hypothesis, analyze and interpret the data gathered in the study (Patel & Davidson, 2003). The inductive approach aims at formulating theory from the collected empirical findings, which is a common approach when there is a lack of theory in the field of the study (Patel & Davidson, 2003).

Wallén (1996) explains that the choice of research approach depends on the researchers' perception of the available amount of theory in relation to the empirical data. Wallén (1996) suggest that the deductive approach is favorable when relatively much theory is available, while the inductive method is preferred when the amount of theory is limited and focus needs to be placed on data collection (Wallén, 1996). The combination of the two, the abductive approach, allows the researchers understanding to grow during the process since the researchers are moving between theory and empirical findings (Wallén, 1996). Shuttleworth (2008) states that the starting point in a research process often is inductive since it gives room for rational reasoning, which then allows for deductive reasoning to design the research process (Wallén, 1996).

The field of this study is fairly well theorized and after the purpose was developed the authors could use the existing theory to interpret the empirical findings, further refine the research questions and at length put the study results into perspective in order to achieve an accurate analysis of the situation. The approach of this study can thus be identified as mainly deductive. However, the study also had some inductive elements at the very beginning of the project before the purpose had been fully established. As Shuttleworth (2008) points out, rational reasoning without limitations from existing theory led to an initial hypothesis and theory that could determine the purpose of the study, which then opened up for designing the rest of the research process in an deductive manner.

3.2 Research Strategy

Bryman and Bell (2011) state that there are two main research methods when performing a study: qualitative and quantitative. The difference between the methods lies in how empirical data is collected and then processed.

A quantitative method implicates that collected data is entailed to numbers and statistical analysis in order to answer the research questions (Holme & Solvang, 1997). The main mechanisms focus on testing the theories with reality (Bryman & Bell, 2011), which is why quantitative research is strongly correlated with a deductive research approach. In qualitative research it is instead the researchers' perceptions that lay the foundation for the study (Holme & Solvang, 1997) and its main objective is to study and understand the social interactions and then build a theory based on the observed interactions. However, Holme and Solvang (1997) mean that there is not often a clear distinction between qualitative and quantitative research methods since both methods are often needed in a study and can therefore be used to complement or verify each other (Holme & Solvang, 1997; Patel & Davidson, 2003).

The research strategy of this study is based on a mix of quantitative and qualitative methods. Emphasis lies on quantitative data since it lays the foundation for the analysis of the production system and concrete recommendations in the end. The quantitative data must be put into context in order to understand it and verify it. Complementing with qualitative data has therefore been equally important in order to understand the situation in the production system and the quality of the quantitative data. Qualitative data complement, support and strengthen arguments and relationships seen from a quantitative point of view and gives new insights of areas in the quantitative data to further explore and understand.

3.3 Research Design

The research design of a study is a framework that determines a given set of parameters of how a study will be carried out in order to answer its research question (Creswell, 2013). The framework dictates how the data collection and analysis of data will be done (Bryman & Bell, 2011).

The design chosen for this research is a case study. Esienhardt (1989) states that a case study aims at understanding the relationships and dynamics that are present in a given setting. A case study also tries to create a deeper and more detailed understanding of subject in a single case of interest (Bryman & Bell, 2011). Creswell (2013) means that case studies are frequently used, and are often incorporating both qualitative and quantitative measures.

As the aim of the study is to investigate multiple performance parameters given the setting of a machining factory with several physical constraints, a case study is the most suitable approach. Only one study (Prince & Kay, 2003) has been identified of previously investigating related questions to this study. This eliminates the room for a research design based on a meta-analytic. The case study will serve as an example of how discussed questions in factories with similar settings can be approached.

3.4 Data Collection

Patel and Davidson (2003) describe several data collection methods such as interviews, surveys, forms and observations in order to collect relevant empirical data. Thomas (2010) and Creswell (2013) covers similar methods but also mention opinions, focus groups, measurements and tests, official statistics and other numerical based data. These methods can then be categorized in primary and secondary data (Creswell, 2013).

All data collection methods are not relevant for a study as this one. The choice of data collection methods depends on what is considered to give the best information in order to be able to answer the purpose in relation to the available time and resources (Patel & Davidson, 1991). In order to answer the stated research questions and thus fulfill the purpose, this study require collection and analysis of both qualitative and quantitative data from both primary and secondary data sources. Qualitative collection methods that were used in this study were mainly observations and interviews, and how these methods have been used is described below. The quantitative part of this study relied on secondary data that could be found in the company´s data bases. Data from different databases were retrieved and combined in order to develop the information needed for the purpose of this study.

3.4.1 Observations

Observations were initially performed to build knowledge about the production system and the current situation. The use of observations in the initial exploratory phase of a study is common (Patel & Davidson, 2003). The obtained knowledge and insights lays the foundation for further data collection with other methods (Patel & Davidson, 2003). Observations were also used later on in the study used to confirm or complement previous collected data or to get a deeper understanding of a specific subject. Observations are according to Patel and Davidsson (2003) a good way to complement information and a way of validating previously collected data (Bryman & Bell, 2011). The observations were conducted through two different methods, shadowing and visits. The researchers shadowed certain key personnel of interest, identified based on the purpose and research questions of this study, during a typical workday, which generated many new insights as well as complemented and confirmed previous knowledge and theories. Visits by the researches were used more frequently and have been conducted in two different ways, namely by a guide which at the same time have been interviewed, or independently in certain areas of interest.

3.4.2 Interviews

An interview is a data collection method, which roots in forming questions (Patel & Davidson, 2003). Interviews work best when these are not too formal and in those cases where the interviewer does not represent an objective expert in relation to the interviewee (Wallén, 1996). An interviewer must consider two aspects: the level of standardization and the level of structure (Patel & Davidson, 2003). The level of standardization means how strict the questions are formed and in which order they are to be asked while the level of structure dictates how freely the interviewee can answer. Based on these two aspects, interviews can be structured, semi-structured or unstructured (Thomas, 2010).

Unstructured interviews were used initially in order to build the authors knowledge and create an understanding of the system and problem situation. Through letting the interviewee decide the agenda and using a tone similar to a normal conversation the unstructured interviews created, in accordance with Czarniawska (2004), a narrative environment in which the interviewee could share their knowledge and viewpoints instead of just answering questions that in the initial stage of the study would have been of rather poor quality. This enabled the authors to ask more relevant questions and use more structured interviews later on in the study. These are, in accordance with Blumberg et al. (2011) and Bryman & Bell (2011), more precise and efficient which allowed the authors to focus more of their time on analysing the data instead of collecting it. These later interviews were held in a semi-structured manner with an agenda in form of topics instead of a fixed list of questions allowed the interviewee to speak more freely and were not coloured by the questions. This led, in accordance with Thomas (2010), to answers that are more representative for the thoughts and knowledge of the interviewee and to that the interviewers could get access to important information that they did not know they needed. The semi-structured approach also allowed the authors to fill in with follow up questions that were formed as the interview progressed.

3.4.3 Sample

In order to obtain representative data to build the empirical data for the study, choosing the right persons that will contribute with information is important (Denscombe, 2009). Denscombe (2009) mean that there mainly are two fundamental sample methods; probability and non-probability sample. Probability sample is most suitable in cases where the purpose is to obtain information that represents a population, while non-probability sample is used where all persons are not considered to have enough knowledge about all areas in order to contribute with knowledge and information (Denscombe, 2009). The latter is the situation for this study, which is why non-probability sample was used in combination with three selection methods; convenience sampling, subjective sampling and snowball sampling.

Convenience sampling means persons considered suitable for including in the sample are chosen because it is convenient for the researchers (Denscombe, 2009). Subjective sampling means that the researchers know which persons are appropriate and therefore handpick these as interview objects (Denscombe, 2009). Lastly, snowball sampling is conducted by asking a respondent to recommend other persons that have the experience, information or the knowledge needed (Bryman & Bell, 2011; Denscombe, 2009).

A combination of these three methods was used to generate a strong and representative sample in a convenient and time efficient way. The sample of the interviewees was initially built up by using convenience sample. This sample consisted generally of people in the near surrounding that the researchers knew had certain knowledge or information of a topic. Continually the sample was built up by using the snowball methodology, where the interviewees in the clear majority of the cases recommended further persons with knowledge relevant for the study. These persons were chosen based on a subjective notion of who would be most representative by the researchers.

3.5 Literature Review

The literature search was formed around the purpose and initial research questions of the study. The found literature within the subject were used as a framework that guided the process of collecting and analysing data. It is important during this review to be open for all literature since it allows the researchers to gain a wide perspective of available information (Holme & Solvang, 1997). The literature is also used as a reference point for assessing the recommendations and conclusions of the study. A broad and iterative literature review was therefore continuously carried out during the majority of the study where mainly printed sources, of type journal articles and textbooks from the author´s prior education, were reviewed. The process of finding relevant literature followed the six steps of the systematic search method presented in Forsberg and Wengström (2003):

1. State research questions
2. Specify the scope of the project
3. Plan the search
4. Formulate key search words
5. Do the search
6. Interpret and evaluate the result

The literature was mainly found through using the online database directory at *Chalmers Library,* using formulated key search words. The researchers prior knowledge of the field and certain books and articles covering the theory gave a starting point for further search.

The initial search formed the fundamental knowledge base required to be able to design the rest of the study. This consisted of theory originated in performance measurement, capacity management, production planning and control, and production layout and flow. This knowledge was needed in order to be able to understand how the production system was designed and worked, but also in order to somewhat understand what the root causes of the problems are that GKN currently experiences.

With the found sources as a starting point, the search was deepened by a review of the relevant references found in these primary sources in order to find more relevant material. This is also a way to validate the information found in the primary sources (Nyberg, 2000). In the later stages of the study, the search focused more on possible solutions applicable to GKN’s context.

3.5 Data Analysis

The model used for data structuring and presentation as well as layout for this report can be viewed in Figure 12. This model was considered to be the most suitable model for structuring the empirical data and for future recommendations.

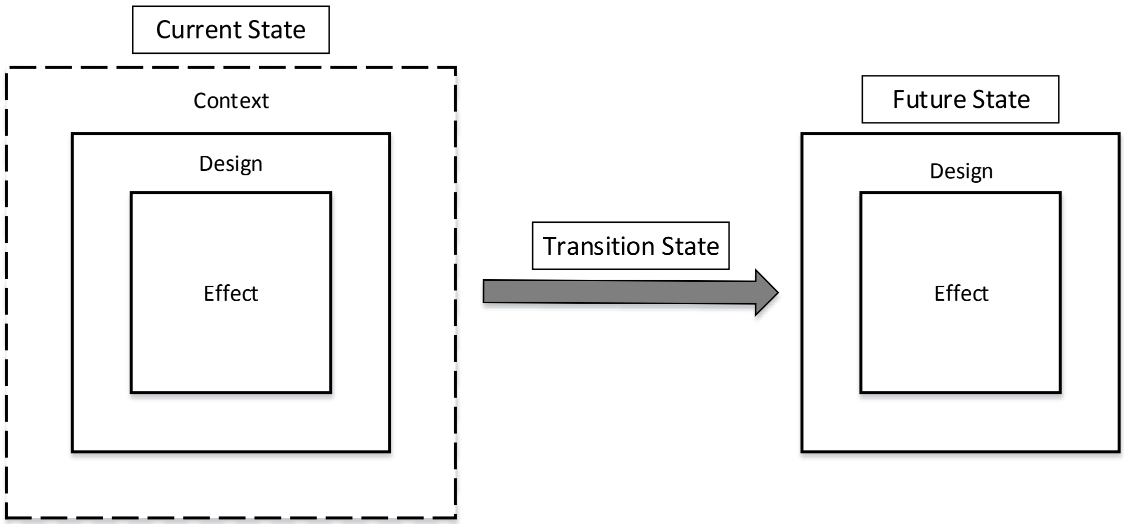

Figure 12: Analysis model.

The left part of the model describes the current state of the company's production system. The model shows that the *effect*, or outcome, is dependent on both the *design* of the production system, which the company can control, and the environment, or *context*, which the company cannot control. The company can change certain parameters, e.g. the supply chain, that affect the context of the production system, but such changes and transformations are on a much longer term, which is why the environment the company operates in must be seen as given.

The distinction of which aspects that are seen as within the company control and which aspects lies beyond the company´s control is determined by the scope of this study. The *context* for this study includes aspects such as customer relations, product range and company characteristics. *Design* contains such aspects as business model, production layout, production planning and control, and product routing. The *effect* is the result and contains a series of performance measures based on, and calculated from the collected data.

The future state in the model describes how the company's production system should be designed based on the given recommendations and solutions. After the transition period where these recommendations and solutions are managed and implemented, the new *design* of the production system will give rise to an improved *effect* in terms of the earlier discussed performance measurements.

3.6 Research Quality

The research quality of the study depends on a set of criteria that all needs to be met in order to achieve a high quality of the research. There are three different criteria that need to be considered: reliability, replication and validity (Bryman & Bell, 2011). Since this study is a mix of quantitative and qualitative methods the more general research criteria of validity and reliability were chosen to be discussed instead of the concept of trustworthiness that primarily is chosen for qualitative studies (Patel & Davidson, 2003).

Validity concerns whether the researchers have been consequent in how the results are interpreted and the integrity of the conclusions generated from the research. Bryman and Bell (2011) mean that it shows if a certain level of credibility of the research has been achieved.

Reliability of a study is a criterion that concerns if the result of the study would be the same if the study were repeated or not. Measuring reliability is according to Bryman and Bell (2011) often mostly considered when performing a quantitative study. Furthermore is the concept of replication linked to reliability since it concerns if it is possible to replicate the study (Bryman & Bell, 2011). This main difference between the concepts is that replication has more focus on if the study itself is replicable instead of the results of the study, which reliability concerns (Bryman & Bell, 2011).

The study mixes both primary and secondary data from several sources. This is a way to triangulate the information and secure the validity of the data in the study (Bryman & Bell, 2011). The triangulation has secured that not only one method has been used to collect certain data where possible. This ensures a higher trustworthiness of the study (Denzin, 1978; Patton, 1999). Furthermore, triangulation of sources has been used throughout the research. This has given the advantage of multiple perspectives on certain subjects and opinions (Denzin, 1978; Patton, 1999).

The quantitative analysis is primarily based on secondary data concerning the production system and products, such as operating times for the work centers and operation lists for the products. There are several advantages of using data from secondary sources, of which the most prominent is that it is time-efficient and allows for spending more resources on the analysis (Bryman & Bell, 2011). This data has been available and accessible and can be used directly from the company's databases. However, the researchers have identified that the quality of the secondary data has been fluctuating. Bryman and Bell (2011) points out that one drawback of using secondary data is that quality and validity of the data cannot always be guaranteed. The researchers has therefore triangulated the data from different databases in order to ensure it has been correctly recorded, but also eliminated data that is highly unlikely since it would give a skewed image of the current state. The researches have through interviews and observations been able to understand how the data is recorded by the system or workers, in order to further secure the validity. The qualitative input has been vital in order to be able to interpret and manage the data. Several of the presented key performance indicators have not been available in the databases, and therefore been created of data mined by the researchers.

To achieve a high reliability, both researchers have been present for most of the interviews. Patel and Davidson (2003) mean that this reduces the risk for interpretation errors. The researchers were well prepared before the interviews, which assumingly lead to a reduced *interview effect* and thereby less biased results (Patel & Davidson, 2003). After the interviews, the researchers have held internal discussions to assure that they agree on the findings and that their reflections have been noted. This ensures that the researchers have interpreted the interviewee in the same way. If there were any uncertainties, the researchers contacted the interviewee to get a clarification. Some persons were interviewed or/and observed more than once to further gain that person's knowledge or information. Some of the interviewed persons have had similar positions at the same or different departments. This has allowed the interviewers to confirm what has been said in previous interviews or gain new aspects of the same topic.

4 Empirical Findings

This chapter covers the empirical findings of this study and is structured according to the analytical model presented in chapter 3. Firstly, there is a presentation of the case study company together with a description of the context in which it operates. Thereafter, production procedures and policies are presented as part of the design, and lastly the result of the quantitative data analysis is covered as the effect of context and design. Paragraphs in this chapter without a direct reference to an interview or database can be considered as results of observations made by the authors throughout this study.

4.1 Context

This section addresses parameters that influence company performance but lies outside the company span of control or the scope of this study, such as market characteristics and customer relations.

4.1.1 Company Presentation

In this section the case object GKN Aerospace Engine Systems in Trollhättan is introduced along with its products, history and position in the corporate structure.

GKN

GKN is a global engineering company founded in 1759 in United Kingdom that in 2014 employed approximately 49,000 people in 30 countries. GKN Group is divided into four divisions, as per Figure 15, and generated total sales of £7.6 billion in 2013.

Figure 13: Divisions in GKN Group (GKN, 2014).

The company´s four divisions all hold a leading market share in their segments, out of which the Aerospace division is one of the biggest with a turnover of £2.2 billion. The Aerospace product line ranges from aero structure components, such as wing structures and fuselages, engine components and subsystems, such as structures, shafts and turbine fans, to special products such as fuel tanks, cabin windows and anti-icing systems. The products are sold both on the commercial market to major airplane manufacturers and on the military market with component presence in the major active, and under development, military aircrafts such as F35, F18, JAS 39 Gripen and the Sikorsky Blackhawk. The split in revenue with regard to product range and market can be seen in Figure 14.

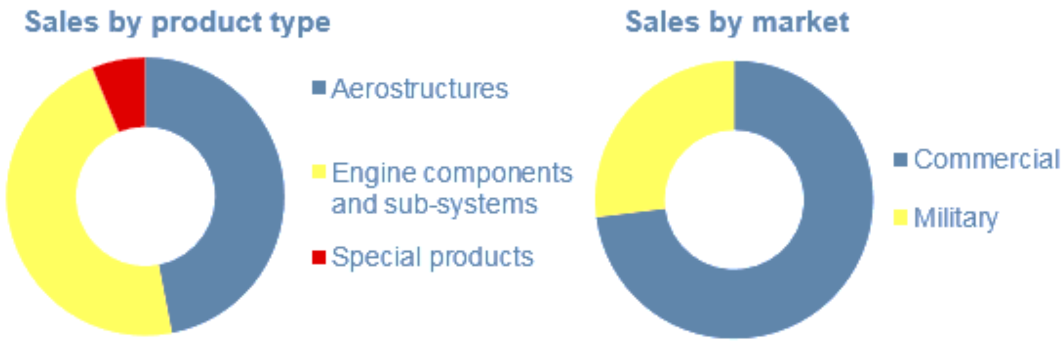

Figure 14: Shares of sales for each division (GKN, 2014).

In 2012 GKN Engineering Group acquired Volvo Aero AB, with headquarters in Trollhättan, from AB Volvo, which took the name GKN Aerospace Sweden AB and came to be a big part of the Aerospace division in the Engine Products segment.

GKN Aerospace Sweden AB

The production of aircraft engines in Trollhätan started in 1930, on request of the Swedish Aviation Authorities for the Swedish Air Force, under the company name Nohab Flygmotorfabriker. Until the 1970´s production consisted mainly of military engines and components. Although projects for entering the civil aircraft engine market started in the early 1940, they did not result in any commercial products until 1977 after the technical transition to jet propulsion had taken place. In 1969 AB Volvo were the sole owners of the company that changed name to Volvo Flygmotor AB and later to Volvo Aero Corporation, which was the name until it was acquired by GKN.

GKN Aerospace Sweden AB, which hereafter will be referred to as just GKN, is a second tier supplier on the aircraft market with all the major engine producers, such as General Electric, Pratt & Whitney and Rolls-Royce, as customers. The company has component presence in almost all larger commercial aircrafts in operation, including the major players Airbus, Boeing and Bombardier. The products produced in Trollhättan and their position in a jet engine can be visualized as in Figure 15.

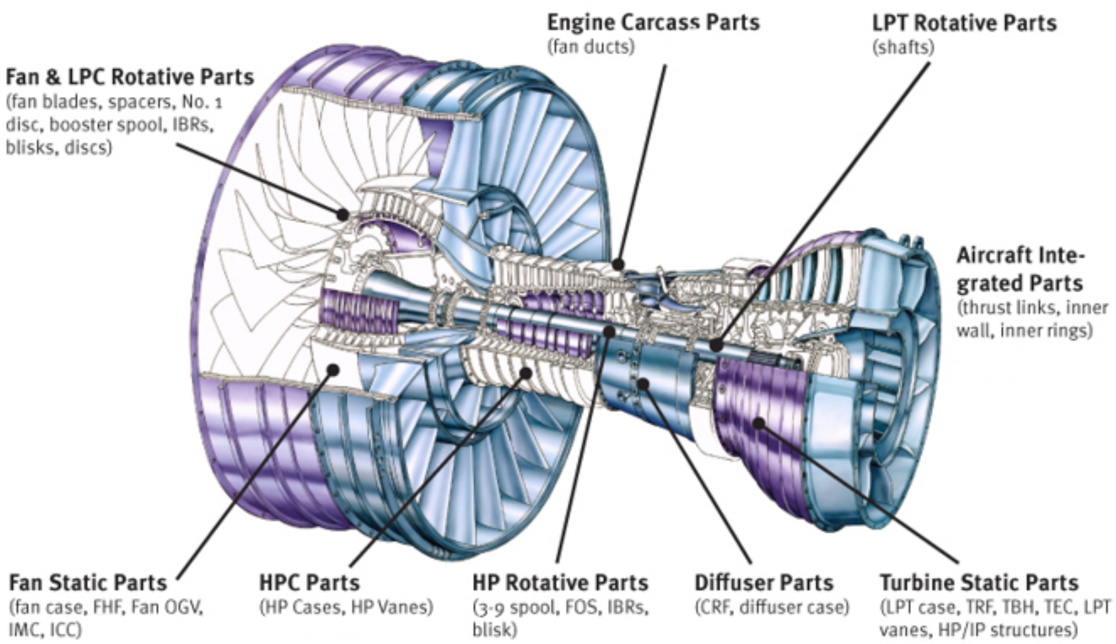

Figure 15: Product portfolio visualization (GKN, 2014).

As seen in the Figure 15, the variety in the product portfolio is large with parts of different functions in different parts of the engine. Rotation and the big difference in temperature between different parts of the engine puts high demands on the components that goes into the engine, and the production system. Each product is made according to unique customer specifications and produced in small volumes for only that specific customer. The product refinement throughout the production system mainly consists of machining operations down to very tight tolerances. The incoming material is mainly castings that, due to the high tolerance demands, are of very high value.

To be able to develop materials and production methods capable to meet these demands and the high tolerances, the Trollhättan site also employ a large number of materials- and production engineers. The Research & Development department is at the very front of technology development with projects in, among others, additive layer technology, robotic laser welding and composite structures.

4.1.2 Business and Customer Relations

The market in which a company operates has a great impact on how the company can design and optimize its operations. Customer characteristics and relations is a significant part of this environment and can in most cases be affected by the company in question. However, analyzing and altering customer relations lies outside the scope of this study and will therefore be regarded as part of the context and thus only frame and limit the future state recommendations in chapter 6.

Market and Customer Characteristics

The market is highly consolidated with three major actors and first tier suppliers to the aircraft industry. These actors are as mentioned above, Pratt & Whitney, Rolls-Royce and General Electric and are all significantly larger, in terms of turnover, than the second-tier suppliers to which GKN belongs. The power distribution between the first tier- and the second tier suppliers is therefore much skewed, to the advantage of the first tier suppliers. The relationships between GKN and their customers are thus to a large degree dictated by the customer[2].

The raw market demand for original equipment, i.e. aircrafts, is fairly stable and predictable for a significant time period ahead, while the demand for spare parts is somewhat more fluctuating. The aircraft *OEM* generally offers their products with two or more engine alternatives and to secure component presence, GKN supply components for both alternatives[2]. This implies that although the demand for the final product, the aircraft, is fairly easy to predict, it is more complicated to forecast the demand for GKN´s products since it is difficult to predict which of the engine alternatives the end customer will choose.

The product volumes in the aircraft engine market are fairly low, which also is the case for the Trollhättan site. For spare parts, mainly rotating components in GKN´s portfolio, the volumes are somewhat larger but still low in comparison with other industries[3].

Business Contract Characteristics

All GKN´s active business agreements are of mainly two contract types and in some cases a hybrid between these two. The dominant type, according to a Key Account Manager[2], is called *Risk and Revenue Sharing Programme*, *RSP.* This contract type implies that GKN takes part in an engine development project, and thus a partly ownership of the engine revenues. The share in the engine is determined by the amount of capital invested in the project. Besides owning part of the revenues, the contract also implies ownership of the risk and design responsibilities. This entails that GKN will get coverage for their costs first when the engine generates revenue, i.e. when a complete engine is sold and delivered. The margins for original components in these contracts are fairly low and at times negative while the spare parts generate very high margins, making these projects very lucrative in the long term. These contracts thus, in many cases, reach breakeven first after 20-30 years into the life time of the project and implies large risks and capital costs[2]. The basic agreement is that GKN will deliver whenever and whatever volume is required and is not guaranteed any order volumes.

The second type of contract is called *Long Term Agreement*, *LTA*, even though it is not as long term as the *RSP* contract. These agreements essentially entails that GKN

[2] Key Account Manager [Business Management]. Interviewed by the authors 2015-03-31.
[3] Logistics Developer 2 [Logistics Development]. Interviewed by the authors 2015-01-23.

sells machining hours to the customer. The contract is set up in a way that GKN takes on to be able to deliver a pre-determined volume of a certain component on a yearly basis. The company is guaranteed to get paid for that volume regardless of it is collected or not[2]. This contract type implies thus less risk than *RSP* agreements but also lower margins and is therefore not as lucrative. GKN has no design responsibility and produce according to specifications set by the customer[2].

Quoting and Ordering Procedure

For *RSP* agreements there is no real quoting procedure. The candidate companies are evaluated on their product development and production capabilities and on how much capital they are willing to invest in the project. The most capable company gets the business. Neither is there a real ordering process, since GKN has to ensure availability and be able to deliver required volumes at any time, production has to be based on forecasts[2]. The customer typically shares their estimated demand three years into the future and the estimated volumes are updated daily through direct contact. These forecasts are derived from the customer´s expected market share and thus their estimation of to what extent their engines will be chosen by the end customers. Since GKN oftentimes has components in both engine alternatives, they are often faced with a situation where e.g. both customers expects a market share of 60% due to overestimations. GKN therefore employ their own analysts who give their opinion on the forecasts[2]. The forecasts are refined as the delivery period approaches. For a typical product program the demand three years ahead is given on a yearly basis, demand for next year is given on a quarterly basis and for the next quarter demand is forecasted weekly. Eventually GKN deliver against *Purchase Orders* that can be received from the customer everything from a week before the delivery date to a couple of months ahead. Up to the point when the *PO* is received, there are no firm customer orders or guaranteed volumes[2].

For the *LTA* agreements the quoting procedure is fairly standard. Among the companies that can deliver according to specifications and tolerances, the company that can deliver in time for the lowest price gets the business. The base for GKN´s quoted price for a potential component is the routing time and the assigned time cost for the involved resources. The resources are grouped together according to their function and charged accordingly, i.e. one hour in a lathe always cost a certain amount regardless if one of the lathes is more expensive to run than the others[2].

When a business deal has been won, the contract states a certain production volume on a yearly basis. The demanded volume is spread out over the year with large variations from month to month. GKN strives to achieve a more stable production and develops a demand plan that is more level on a monthly basis and has to be approved by the customer. This demand plan states an upper level of what the customer can expect to have available and is the base for long-term planning of production capacity. The demand plan can be changed throughout the entire lifetime of the project. GKN delivers, as for the *RSP* agreements, against *POs* that not necessarily live up to the demand plan volumes in the short term[2]. GKN is obligated to ensure the availability of the demand plan volumes at all times while the customers are not obligated to pick up those volumes but are obligated to pay for the agreed upon yearly volume. The demand plans are the base for the development of master production schedules, covered in the following sections[2].

4.2 Design

In this section, parameters in the set-up of the production system that the company has influence over, and that are within the scope of this thesis, are covered. It contains all aspects from the business model and production steering to the physical layout of the production resources.

4.2.1 Business Governance

Current governance of the production system and the parameters through which performance is measured is centered on time and cost. Everything is measured in time and transformed into costs by time cost factors, both on a system level as well as on a local level in labor reporting. The way the system is governed and evaluated has a significant impact on how the production is run, as will be discussed in chapter 5.

Performance measurement

Delivery performance is measured by how much an order is delayed in relation to the delivery date set in the *MRP* system by the demand plan. It is measured internally and customer satisfaction is not formally or continuously measured.

All business deal calculations are based upon the time cost of production. When doing this calculation the starting point is the routing time for the product in question, i.e. how much time said product will require of the production resources and how much each time unit cost. This routing time thus then become a performance measure for the production system. The routing time is expected to decrease after the ramp up due to learning effects and this is taken into account when evaluating a deal. If the routing time does not decrease according to plan or even increase, it is seen as a sign for poor performance since the deals turns out to be less lucrative than expected. This way of measuring has a great impact on the load and capacity planning. Furthermore, all work centers are judged based upon how much operating time they report. If it deviates from their capacity it is seen as a sign for caution and potential rationalizations. According to Logistician 1[4], this results in an unwillingness to ask for help from other work centers if in trouble, since all work centers strive to keep their budget portion and staff.

There exists a general goal of reaching what is called the *1000 SEK factory,* which implies an effort to reduce the time cost to 1000 SEK per hour in the entire factory. This is supposed to be done by increasing the local efficiency, i.e. to handle more products per time unit and thus spreading out the time cost on more products.

[4] Logistician 1 [Logistics C-shop] Interviewed by the authors 2015-02-19.

Labor and Responsibility structure

Even though many of the operations are automated, the tight tolerances and advanced methods require skilled and educated operators, especially in welding and heat- and surface treatment, which require certain certifications. The learning period for newly hired operators is thus long and expensive[5]. Therefore are the operators educated for only one, or in some cases, a few operations, which implies difficulties in moving or lending operators from one work station to another if required for e.g. covering sick leave, change in demand or other capacity situations[6]. The operators are responsible for the quality of their work and due to the tight demands for traceability in the aircraft industry, all involved operators in the production of a product are logged[7].

Production planning and logistics are carried out by logisticians. The factory is divided into functional process areas and a logistician is responsible for one, or in some cases a few, of those areas. A product passes over such responsibility borders several times on its way through the production system. Communication between the involved logisticians in such a crossing is arbitrary and non-formalized and the products are handed over when finished in the upstream responsibility area.

Labor Reporting

The operators at the different work centers report the operating times manually. This operating time is the only time that is reported from the work centers. This implies that it should include all handling of a product at the work center. However, since the time is reported manually, there is a significant possibility for the reporting to be done differently at different work stations and by different operators. This is actually also the case, some operators report the entire time from when they pick up the detail to when it leaves the station while others start the time when the machine is turned on[6]. There is thus no real data on set-up times or handling times available. This together with the fact that the reported time is used as a base for individual evaluation of the operators creates a situation where the risk for corrupt data is high and where the operators are unwilling to help each other[8]. According to Logistician[6], it also creates a situation where problems are passed around between the divisions and responsibility is avoided since no one wants to take care of it because then they will not be able to keep up with their schedule. Furthermore, this creates an incentive for the operators to log more work than the true amount since they want to show that they are crucial for the company. This at the same time as the logistician wants to keep the routing time as low as possible as it is used in their evaluation. This creates a time gap between reported time and how much e.g. an operation should take that should be seen as waste and a sign for caution. This time gap however is not analyzed in any formal way and left unhandled.

[5] Logistician 2 [Logistics A-shop] Interviewed by the authors 2015-02-11.
[6] Logistician 1 [Logistics C-shop] Interviewed by the authors 2015-02-04.
[7] Logistics Developer 3 [Logistics Development] Interviewed by the authors 2015-03-24.
[8] Logistician 3 [Logistics X-shop] Interviewed by the authors 2015-02-06

4.2.2 Production Layout

GKN Aerospace Engine Products' factory unit in Trollhättan, Sweden, consists of multiple production facilities. Generally, all production shops have a functional layout, where similar machine types and resources are grouped together in job-shops in each production-shop. Machines have been grouped into this layout for many years and there are several reasons for this:

1. The factory has many processes that require specialist support, e.g. heat and surface treatment.
2. There are processes, e.g. machining processes that require the same technical competence from specialized operators.
3. Some processes, e.g. grinding, control and washing are shared amongst all products and will therefore have high machine utilization.

The majority of the machines in the factory cannot be moved. Most of the machines are large and bulky and also lowered into 2-5 meter deep pits. In these pits, some machines rest on airbeds and are fixed into the ground. This is required due to the fact that a high amount of the operations have very tight tolerance limits, e.g. in some cases within ±0.001 mm. It is therefore crucial that the machines micro movements are minimized. However, even though the machines are fixed, many operations are still sensitive to micro movements. According to Logistician 1[4], different seasons and temperatures can actually affect the movement of machines and cause deviating results in operations leading to quality issues.

The compartment of the shops is a factor limiting machine re-allocation in the factory and shops[3]. Another reason for the machines are lowered into the pits is that the height to the ceiling is often too low. Investing in new or moving machines demands therefore advanced and extensive planning since pits need to be dug. An example is work center 9767, an automated multi-functional cell that is lowered into a pit with a size of a 50 meter long swimming pool. Thus, reorganizing machines is practically impossible[3].

The products GKN manufactures are characterized by low volumes and according to Colliander[3] is a product with a yearly volume of over 100 units considered as a high volume product. Even though the product mix is not changed, as often as in other manufacturing industries, there are too high costs to change the production layout due to changes in the product mix due to earlier mentioned physical constraints.

Buffer, Storage & Transport

Buffers are used in the production system at strategic places in order to ensure that some machines always have work to do even if there are fluctuations. This was confirmed by observations in two separate cases. A preceding work center lacked personnel for a certain period of time, which meant that the subsequent work center processed orders from its queue that work day. In the other case, a subsequent work center lacked manpower. The logistician then ordered the operator to start building buffers.

Orders that are in production and between operations can be stored in a couple of different ways while waiting for e.g. a machine to be available or a decision to be taken. A single work center can have their own space in front of the machine that is reserved for in and out traffic, where pallets are placed in the square for incoming or outgoing goods by a forklift or in some cases manually by the operator.

However, for many of the work centers there is not enough room for the forklift to deliver the order directly to the machine. Many of the resources instead share a storage area for incoming and outgoing orders. The area is dedicated to the resources in the nearest surroundings. These are located all over the shops ranging from room for a couple pallets to approximately two dozens. The operator then drags the material to the machine with a manual forklift and back when the operation is finished. A brief overview of the storage process is described in Figure 16.

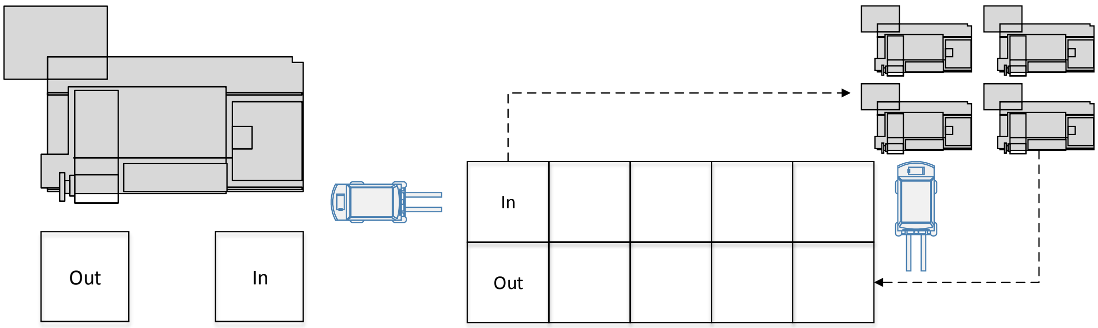

Figure 16: Loading and storage areas that can be found in all shops.

If there is no room in the shared storage area in the shop, due to e.g. long queues to machine resources, the order is sent to intermediate storage in the warehouse in connection to primarily X- or C-shop. Berglund[8] states that orders that have finished an operation in e.g. C-shop might be transported to the intermediate storage in X-shop and back again if there is no available room in C-shop.

Transports are performed either by forklifts or manual forklifts between machines and storage areas. The forklift activities are outsourced to *Coor Service Management*, who operates all forklifts. Manual forklifts are however managed internally. The quantitative data analysis showed that the transport times are short in relation to total operation and waiting times, which is why these are disregarded in the further data analysis. However, the waiting time for transports varies more and affects the production flow. Berglund[8] states that it is very common that one has to wait for hours to get a forklift transport after it was automatically requested after the operation was finished.

4.2.3 Production Planning & Control

This section covers the logic behind the production system at the Trollhättan site, which includes how customer orders are translated into production orders, loaded onto the system and released to the shop floor. It also includes how available capacity is determined and planned, how production orders are controlled and steered after being released and how changes in demand are allowed to affect the production system.

Logic Structure

The company uses a traditional *MRP* system that pushes materials through the production to meet the customer demand. The raw customer demand is translated into a demand plan, as described in section 4.1.2, which is what the company states that they can deliver. From this demand plan the logisticians work out a master production schedule that will meet the deliveries promised to the customers[9]. This master production schedule is what sets the due dates for the final product in the *MRP* system, and will be covered in more depth in the load planning section.

When the due date is set for the final product, the *MRP*-system uses a *Bill-of-Material*, *BOM,* as input to assign individual dates to all ingoing components for when they are needed in the production. The system then counts backward one lead time for each material from the due date in order to assign the date for when the specific material at the latest should be released onto the shop floor to meet the due date, as in equation 17.

$$Start\ date = Due\ date - (Routing\ time + Supplier\ lead\ time) \quad (17)$$

The lead time for each material is determined by the routing time, presented in section 4.2.1 that represents the total supposed operating time, the expected queue time in the system and the lead time for the incoming raw material from suppliers.

Load Planning

The demand plan is derived from the raw customer demand and is what the company undertakes to deliver. The raw demand varies greatly from week to week implying that some weeks the demand will be higher than what the company can meet and the demand plan is therefore more level than the raw demand[7]. Even if the demand plan is more level, it still varies more than what is desirable in the production. In the master production schedule, the logistician therefore tries to achieve a stable production rate that can be maintained at least in the short term[9].

When developing the master production schedule, the logisticians compensate for inaccuracies in the stated routing time by adding time so it matches their experiences regarding the specific operation. Due to how the routing time is used as an indicator of local production performance, changing the routing time upwards is thus avoided since it is seen as a sign of poor performance[8].

In the long term, the master production schedule does not take the available capacity into consideration since the logic of an *MRP* system, as stated in section 2.3.2, implies infinite loading, which can be can be visualized as in Figure 17.

---

[9] Logistician 4 [Logistics Disc-shop] Interviewed by the authors 2015-02-11.

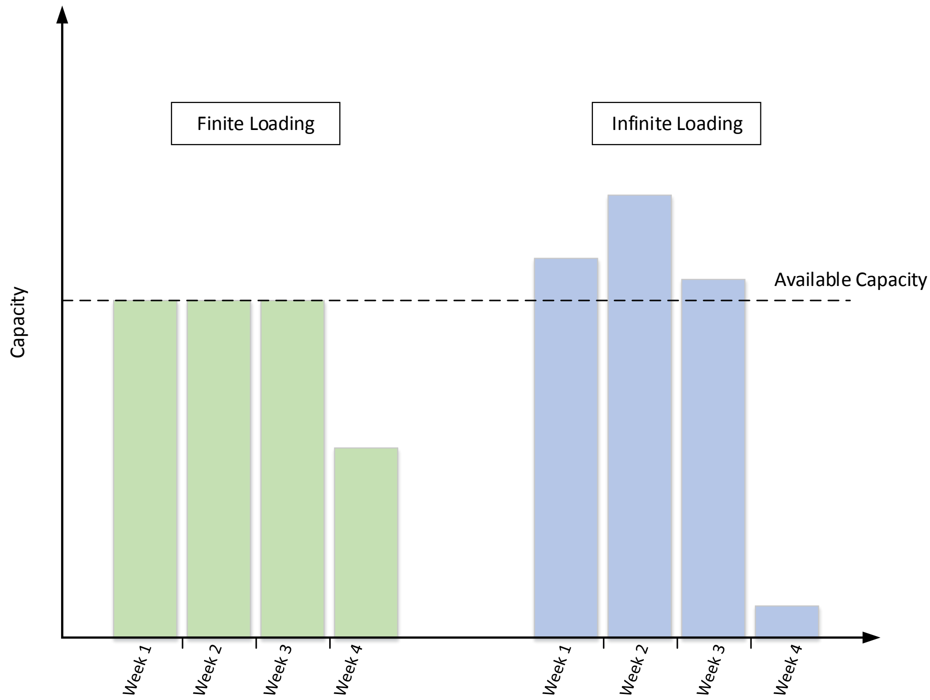

Figure 17: Comparison between finite and infinite loading logic.

The master production schedule is loaded onto the shop floor according to the product routing. Each product is assigned an individual routing, which includes its operation list and information regarding on what work center the operations should be carried out. The material number and operation number is combined to a unique operation code. This code is static and the same for each product of that type that enters the production. The assignment of operation codes to work centers and machines is dependent on a number of aspects. Firstly, the operation code must be able to be carried out in the machine, i.e. the product must fit and the machine must capable to perform that operation e.g. lathing, welding or milling[6]. Secondly, the machine must perform according to the tolerance limit of the product in question. Finally, the machine must be certified to perform that operation code[8]. The fact that the majority of the products are critical components in the engines they are mounted in and that the aviation market is exposed to high safety requirements, imply that the customers must be able to rely on the product quality that GKN delivers. The machines and production methods must therefore be certified by the customers for them to be allowed to operate on the customer´s product.

The machine park consists of mainly NC-machines that are highly automated. This means that when allocating a product to a machine, a NC-program has to be created, tested and calibrated for that specific machine-product combination. This requires programming, production technicians and testing resources. The situation at the company today is that it takes at least eight weeks for a NC-program to be created and another couple of weeks to calibrate the machine before it is capable to handle the new product[4]. Since the machine also has to get a customer certification implies that it can take up to half a year to allocate a product to a new machine and that the product routing thus is rather fixed in the short-to medium-term[10].

There is a lack of formalized documentation regarding the capabilities of each machine, apart from the specifications from the supplier. The tight tolerances imply that two machines of the same make and specifications might be too different for them to be able to perform the same operations. Such information is often tacitly owned by the responsible logistician or production technician and thus not available for other decision makers[6].

The master production schedule is realized through order release according to the production start date, generated by the *MRP* system. However, as stated in section 2.3.2, this date is only considered as a latest start date, leaving room for individual decisions for the logistician. The decision logic is such that if the latest start date for an order has passed, that order will always be released as soon as there is available material and available capacity at the gate-resource. If the start date lies in the future and there is an available gate resource, it is up to the logistician to decide whether to release the order or not. Due to the set-up of labor evaluation, discussed in section 4.2.1, the logistician is pressured by the gate-resource to release the order, which oftentimes is the result.

Order Prioritization

When released onto the shop floor, orders are controlled and prioritized mainly according to their due dates. This is the case for all work centers but the *Shared processes*. When arriving at such a work center the order in queue, if several, with the nearest due date will be handled first. At a shared resource however, the prioritization rule is *First-In-First-Out*, *FIFO*. There are however exceptions to these two procedures. One exception is that each logistician responsible for a part of the production system also has the possibility, in a specific time period, to mark a fixed number of orders as *Prioritized Orders*. Such an order will always be handled as soon as possible, regardless of its set due date[5]. Furthermore, in certain parts of the shared resources, e.g. heat treatment, efforts are made to maintain a pre-determined production rate for certain high-volume product flows and resource load is planned accordingly. However, due to variations in the arrival rate in said product flows, the capacity initially planned for the rate-steered production sometimes is available for *FIFO*-steered production. This irregularity causes conflicts between the different prioritization rules and requires constant actions from the responsible logistician[5]. Moreover, for some resources advanced algorithms for optimal prioritization have been developed in order to ensure optimal use of capacity and are to be implemented. An example is the multi-task cell in the X-shop and for such resources, the other prioritization rules will be overruled by the algorithm[11].

---

[10] Logistics Developer 1 [Logistics Development] Interviewed by the authors 2015-02-12.
[11] Logistics Developer 4 [Logistics Development] Interviewed by the authors 2015-03-16.

Daily Planning & Control

Much of the daily planning and control is done through meetings every morning where the responsible logistician and shift heads are present. The meeting is held around large manual visualization boards with information regarding planned load, manning, production queues, quality issues, planned maintenance and health and safety.

During these meetings the product queues are gone through and the operators get information on whether they can use the standard prioritization rules or if there are any exceptions. It is now jobs are actually loaded onto specific machines in order to fulfill the planned load from the master production schedule. There are big differences between the different shops and responsibility areas in how this is done.

There are, as mentioned, widespread efforts to achieve stable production rates both when developing the master production schedules and in the daily planning and control. In most responsibility areas, the logistician load the machines with orders according to their MRP-dates set by the master production schedule. The visualization boards are matrixes organized by products with columns for the operations performed in that specific responsibility area and magnetic bars representing orders are placed in rows according to if they are in operation, waiting or if they have some sort of quality problem[4].

However, in some areas such as the new fabrication area in the A-shop, the objective to ensure a stable production rate is stricter. Here each product type has fixed time slots at the incorporated work centers that are the same for each week. If there are no orders of a certain product type available at a work station when the time slot for that product starts, the work station will be empty even if there are other orders that could be handled by that work station available[4]. Here the board is organized by workstations and each time slot has its own color. The magnetic order bars are placed on such a time slot when they are active. The fixed production sequence in this area overrides the master production schedule to some extent since the time slots are not updated continuously, making this area not as responsive as other areas where these decisions are made daily as the program changes.

Capacity Planning

The infinite loading logic of the *MRP* system implies that capacity is seen as relatively flexible in the medium-to-long term, as stated in section 2.3.2. Thus, a capacity-analysis tool has been developed outside of the information system for comparing the future planned work content, i.e. master production schedule, with the available capacity for that specific time period[3]. The resolution of this analysis is at a work center level and the purpose of this tool is to identify where there are gaps between load and capacity and thus where capacity has to be increased. Since machines are complicated and specialized, the alternative for capacity increase in near time is overtime or borrowing resources from other work centers[7]. Task specialization requirements limit the possibility for expanding the work force, thus leaving resource borrowing as the major alternative for capacity increase[5].

The capacity is calculated as

$$Available\ capacity = CAP_{100} - Productivity\ loss \quad (18)$$

Productivity loss is based on the logisticians experience and represents the capacity loss due to e.g. maintenance, sick leave and quality issues, and is most often considered to be 20 % of *CAP100*[7]. In the equation below, *Cap100* represents the theoretical maximum capacity and is determined by

$$CAP_{100} = \left(\frac{N_A}{N_O} \cdot H_S\right) \cdot L_M \cdot F_M \quad (19)$$

where

$$\frac{N_A}{N_O}: Shift\ fulfillment$$

$$F_M: Machine\ factor$$

$$L_M: Manning\ level$$

$$H_S: Shift\ hours$$

Each work center runs according to a predetermined shift and the S*hift fulfillment* is the ratio between the actual staffing at the work center and the staffing stated in the current shift.

There are currently several different types of shifts running but efforts have been made to reduce the number of shifts and run a shift type called *GKN3* wherever possible[8]. This shift is divided into three work shifts each day and operates 107 hours per week. Larger shifts are run where the workload is high, mainly in the heat treatment and other shared resources[5].

The *Manning level* is the result of the capacity adjustment by resource borrowing as stated above. If the work center borrows capacity the *manning level* is greater than 1 and vice versa. The *Machine factor* originates from the fact that some machines do not require operator presence during the full run time and the operator can serve another machine in the meantime[6]. As the capacity is based on man hours, this has a significant impact. If the machine requires full staffing the *Machine factor* is 1 but if the machine can run e.g. 20 % of its operating time without an operator present, the *Machine factor* will be 1.2. This implies that the available capacity increases by 20 %.

Short Term Demand Changes

Since the company has undertaken to deliver what the customer needs, the demand might change more frequently than the quarterly updates of the demand plan, which implies that it might be outdated and inaccurate.

A rough-cut capacity planning tool has therefore been developed in order to more frequently update the master production schedule and detect the potential capacity gaps. The tool lies outside of the *MRP* system and is owned by the logistic development department. The use of the tool is initiated by a change in demand and a request to change the demand plan from the responsible business coordinator. The request is handled by the responsible logistician who translates the new volume into a master production schedule and tests it in the rough-cut tool to see if it matches with the available capacity[9]. If there is enough capacity to handle the updated volume, the new master production schedule and corresponding delivery schedule is discussed either in weekly meetings on the shop floor, if it is a minor, change or in larger meetings with the market department and involved managers if the change is significant. If the delivery schedule is approved, the logistician manually plugs in the data from the rough-cut tool into the *MRP* system, enabling a new load planning[9]. If the available capacity is insufficient, a decision is made whether the new volume will be accepted and realized through overtime or borrowing capacity from other parts of the factory, or if the change request is to be turned down.

Regardless if the new volume is approved or not, the information from the rough-cut tool is loaded into another tool that consolidate information regarding the changes in the delivery schedule and the production system´s capability to fulfill it, i.e. finished goods, *WIP* and raw material stock. This information is shared with the customer as a status update[7]. The process for changing production volumes is presented visually in Figure 18.

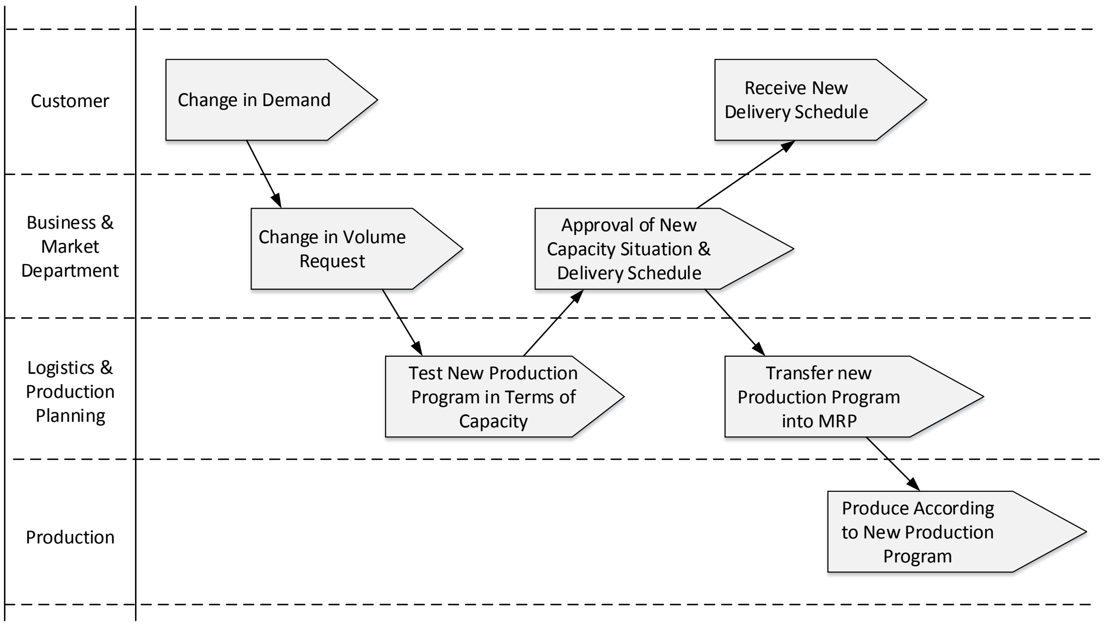

Figure 18: Process for updating demand plan in near time.

Quality Inspection

Due to the close tolerance requirements there is a need for tight quality control throughout the production chain. Each product is controlled several times and with several methods on its way towards completion. Sophisticated measuring equipment is used for detecting dimensional errors both at each work station after each operation and at dedicated control stations. All operators have the responsibility to report every deviation from the specifications they encounter. X-ray equipment is used to detect internal tensions and cracks in the materials that can occur due to machining and temperature changes caused by welding and heat treatment. For detecting surface deviations in welds and material that are not visible to the naked eye, a method called *Liquid Penetrant Inspection* is used. The product is covered in a liquid with very low surface tension and viscosity that penetrates into defects and imperfections. The product is then washed so that excess fluid is removed. The liquid is fluorescent so that when the product is then inspected with ultraviolet light, surface cracks and deviations are detectable[3].

When a deviation or defect is detected it is categorized according to the severity of the imperfection. This categorization states what type and how the corrective action should be performed. For smaller defects such as dimensional errors where the dimension of the product is larger than in the specification, internal production technicians can make the decision to simply machine the product down to the specified dimension. If a more severe defect such as internal strains or cavities is detected, the customer has to decide what corrective action should be performed. This is due to the same reason that machines have to be certified which is stated above. The customer can either decide that the product may be repaired or that the defect is so severe that the product has to be scrapped. For products that may be repaired, the customer will advise upon how the reparation is to be done[3]. The repair operations are added to the product´s operation list under the code *ZP03*[3]. The decision can take everything from a couple of days to weeks and the product is taken out of production until the decision has been made.

When the corrective action has been decided upon, the product will re-enter into the production at the stations where the *ZP03*-operations will be performed. *ZP03*-operations are in the same product flow as all other products, which can cause obstructions and delays.

4.3 Effect

In this section the effect, in terms of company- and production performance, of the context and design will be presented. The data presented is the result of the quantitative part of this study. All graphs, for which there are no source references, have been developed and put together by the authors. The data presented in all figures and tables have been masked due to confidential constraints. However, the masked data presented is fully validated and meaningful, and therefore do not change any of the identified findings.

4.3.1 Delivery precision

As stated in chapter 1.1, customer satisfaction is the ultimate goal of any profit driven organization and that delivery precision, for the customers, is a major part of the sense of product quality. Problems with poor adherence to delivery dates are, as mentioned, also the main reason for why this study is being conducted. Figure 19 shows the delivery precision for 2014 for engine products in GKN´s Trollhättan site. The measure is defined as the ratio between deliveries in a certain time period and the agreed upon customer demand in the demand plan for the same period. Orders that are delivered ahead of schedule are included in the numerator and orders that are already delayed are added to the denominator. The measure measures thus in fact the fraction of orders that are not delayed instead of deliveries according to schedule.

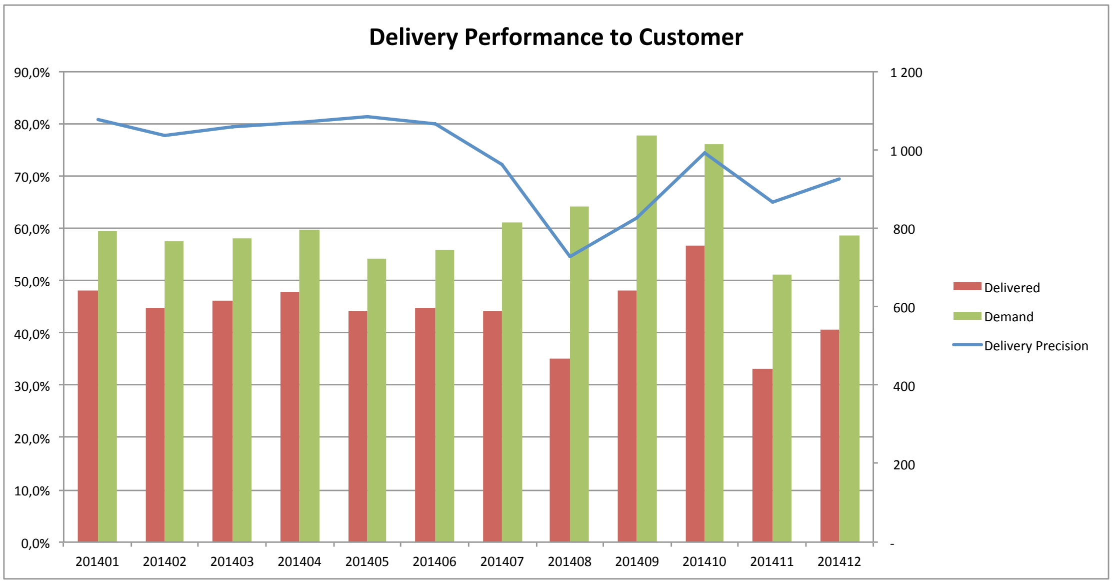

Figure 19: Delivery precision to customer by Engine Products during 2014.

The general view throughout the company is that a 70 % delivery performance, which is the average through 2014, is rather good. According to Logistician[2], the customers are also fairly satisfied with this performance with the motivation that the demand is uncertain and definitive *PO* is given close to the expected delivery date. However, the delivery performance is the main selling point alongside price when signing new business deals. The importance has increased over time and is expected to increase further in the near future[2].

The delivery performance is mainly affected by delays in production, which can be seen in Figure 20 below. The delay is presented for the logistical responsibility areas. It is measured in reported time against the time that should have been reported, if the area should be able to fulfill the master production schedule.

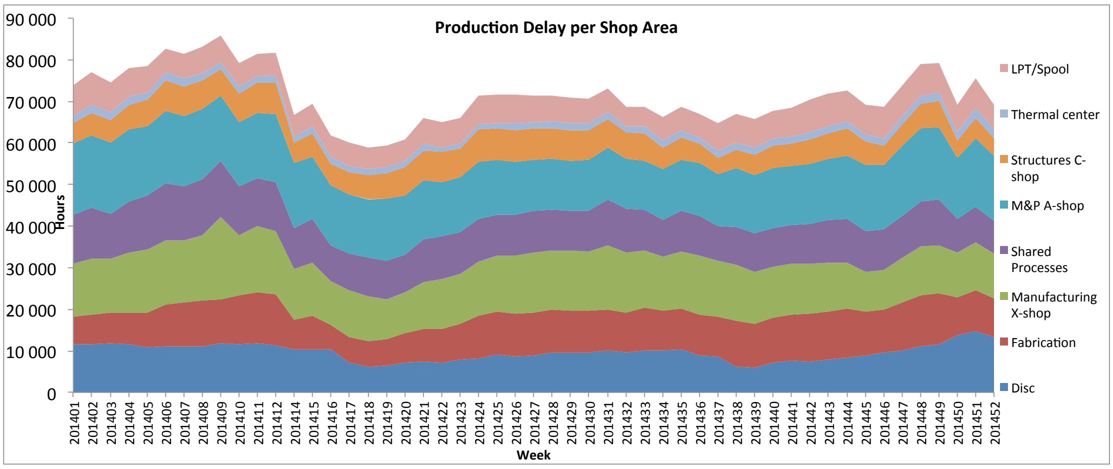

Figure 20: Accumulated production delay in relation to master production schedule per production area during 2014.

Both these measures are fairly stable over time with exceptions for dips and peaks in conjunction with the major vacation periods and periods in which the factory has worked overtime in order to catch up on their schedule[7].

All data is structured per responsibility area and to make the analysis as clear and interesting as possible, two areas have been picked out and are covered further. These two areas are seen as the most interesting with big differences, in terms of capacity utilization, situational complexity and performance, between them. They are *LPT/Spool* and *Shared Resources.*

4.3.2 Capacity utilization

The work centers capacity utilization is calculated for the majority of the work centers in the factory. Figure 21 and 22 illustrates the declared capacity for each work center in the two different areas, *LPT/Spool* and *Shared Processes*, and how the hours utilizing this capacity are distributed. Figure 23 illustrates the capacity utilization per machine category for the whole production system. The productivity is currently based on the logisticians' estimations, experience and assumptions. However, in order to conduct an accurate analysis, the productivity loss is in this study based on actual past data and therefore defined as

$$Productivity\ loss = Unplanned\ maintainence + Planned\ maintainence + Allowances + Quality\ losses \quad (20)$$

The definition of the data, assumptions and how it has been accessed is described below. Reservations are made for the work centers, primarily the rightmost work centers in Figure 21 and 22, with few reported hours since the data may not be fully reliable.

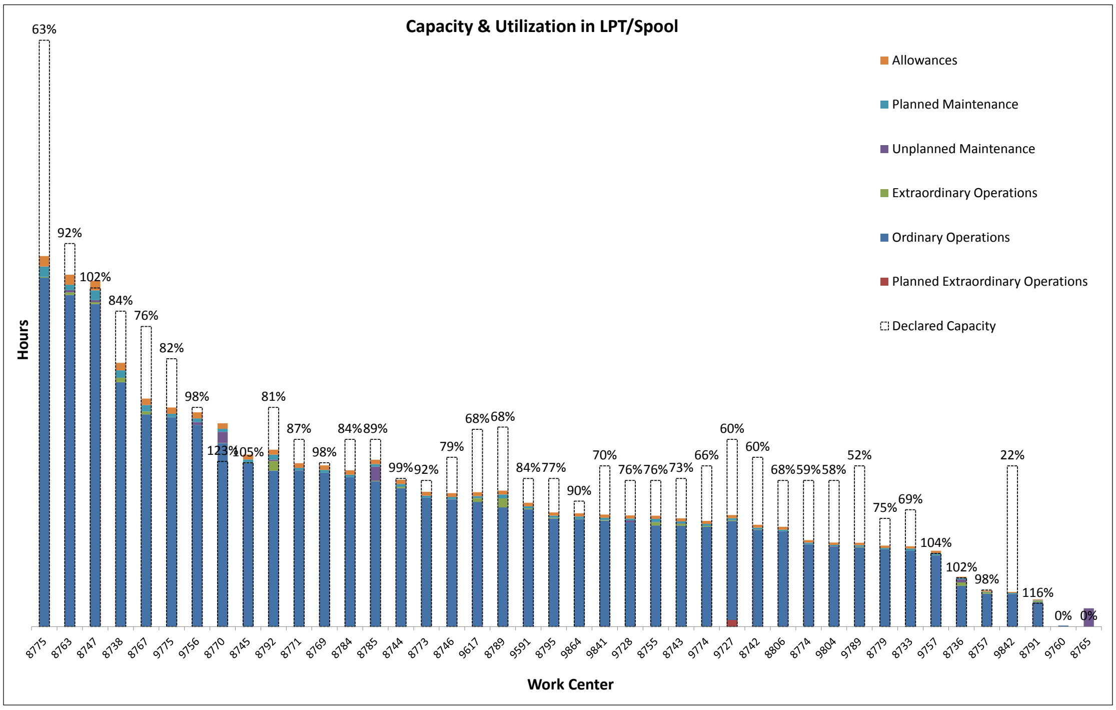

Figure 21: Capacity utilization per work center in LPT/Spool during 2014.

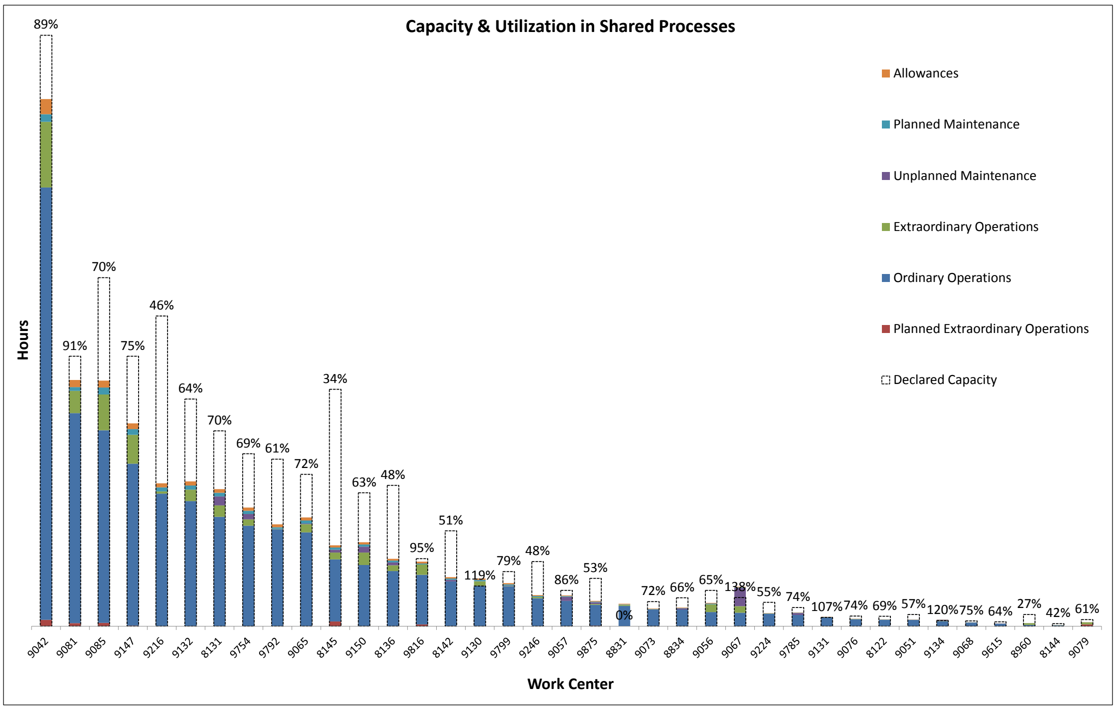

Figure 22: Capacity utilization for each work center in Shared Processes during 2014.

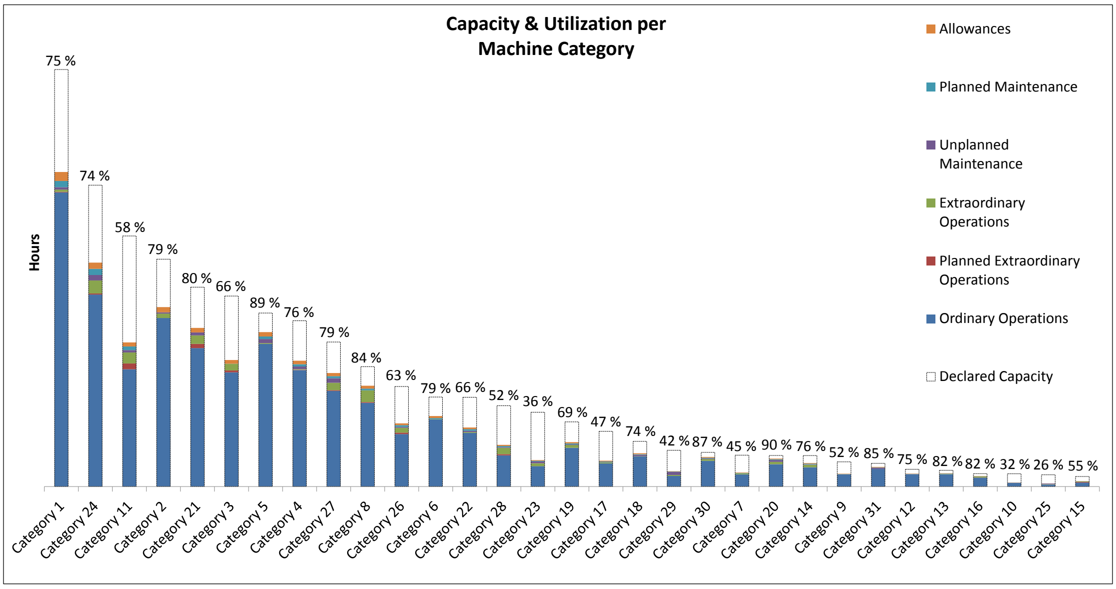

Figure 23: Capacity utilization per each machine category in the production system during 2014.

Declared Capacity

The declared capacity is collected from the Logistics Support department's capacity analysis-tool, where each work centers´ capacity is defined. The capacity is determined through a process in which the logistics developer together with the responsible planner decides what the actual capacity is, given the up-to-date machine, manning and shift levels and an estimated productivity loss. The value defined as declared capacity is *CAP100*. The *CAP100* value becomes the stated maximum for how many hours the work center can produce without any productivity losses.

Ordinary Operations

An ordinary operation is initially scheduled on an order's list of operations. It is required to fulfill the order. The reported hours for each work center consist of the operating time including set-up- and teardown time, manually reported by the operators. This data have been accessed through the information system.

Extraordinary Operations

Extraordinary operations are reported as *ZP03* in the information system. These operations are non-value adding, since they are corrective of previous operations or incoming material and therefore separated and added on top of the ordinary operations as a productivity loss. As with the ordinary operations, the operating times are reported manually by the operators.

Planned Extraordinary Operations

Planned extra ordinary operations are reported as *ZP11* in the information system. The operators also report these operations manually. These operations are non-value adding as *ZP03*, due to the fact that they are of a corrective nature. The difference is that they are added to the original operations list for an order since it is known that they are needed for a large portion of orders. The number of hours consisting of planned extra ordinary operations is added on top of the ordinary operations since they are a productivity loss.

Unplanned Maintenance
The data concerning unplanned maintenance have been accessed from a database in the information system with support from the Machine Maintenance Engineering department. For each unplanned maintenance that is required from e.g. an unplanned breakdown, a repair request is issued by the operator and sent to the technicians, who then responds to the request, repairs the machine and finally reports how many hours the machine stoppage was in total. The amount of machine stoppage hours for each work center is added to the rest of its reported hours and thereby illustrating productivity losses.

Planned Maintenance
The amount of hours representing planned maintenance consists of two weighted factors from two different data sets. The first one is based on actual reported planned maintenance hours for each machine from the technicians. However, Forsed[12] states that this dataset is not fully reliable as with the data for unplanned maintenance. The factor is therefore weighted to two thirds, since it still is the most reliable source available. The second dataset estimates the planned maintenance to be 1 % of all reported production hours, but this set does not contain all work centers. Thus the 1 % serves as a template. This factor is then weighted to one third to serve as a weighting neutralizer since the first factor cannot be fully reliable. The value is added on top of all reported hours and illustrates the productivity loss.

$$Planned\ Maintenance = \left(\frac{2}{3}\right) \cdot Reported\ Maintenance + \left(\frac{1}{3}\right) \cdot 0{,}01 \cdot Reported\ Production\ Hours \quad (21)$$

Allowances
Allowances are estimated to be 3 percent of all hours reported on each work center and added on top of all hours reported. There is no available data for this measure and therefore replaced with a broadly used template of 3 percent.

4.3.3 Variability
This section describes the variability in the production system and its effects in terms of queue time.

Process & Flow Variability
Figure 24 and 25 show the coefficient of variation for each work centers' process time, queue time and arrival frequency per time unit in *LPT/Spool* and *Shared Processes*. The horizontal lines illustrate the limits for the classes of variability explained in chapter 2.1.1. The levels are primarily defined for arrivals per time unit

[12] Maintenance Technician [Machine Maintenance Engineering] Interviewed by the authors 2015-02-18.

but are also used for classifying the variation for process time and queue time. It should be understood that the limits set at 0.75 and 1.33 are relatively high for evaluating the variation of process and queue time and that these normally would be lower (Hopp & Spearman, 2011).

The coefficient of variation for process time is calculated by dividing the average by the standard deviation of the reported operating times at each work center from the information system. The variation for the queue time was calculated similarly but it is retrieved from another database logging the queue time before each work center. The queue times in this case are computed from when each order, i.e. a single product unit, arrives to the work center's load and unload area, until the operator registers that the operation is started. The number of arrivals is computed for each work center per workday. The coefficient of variation is then, as previously described, derived from the average and standard deviation of number of arrivals per workday.

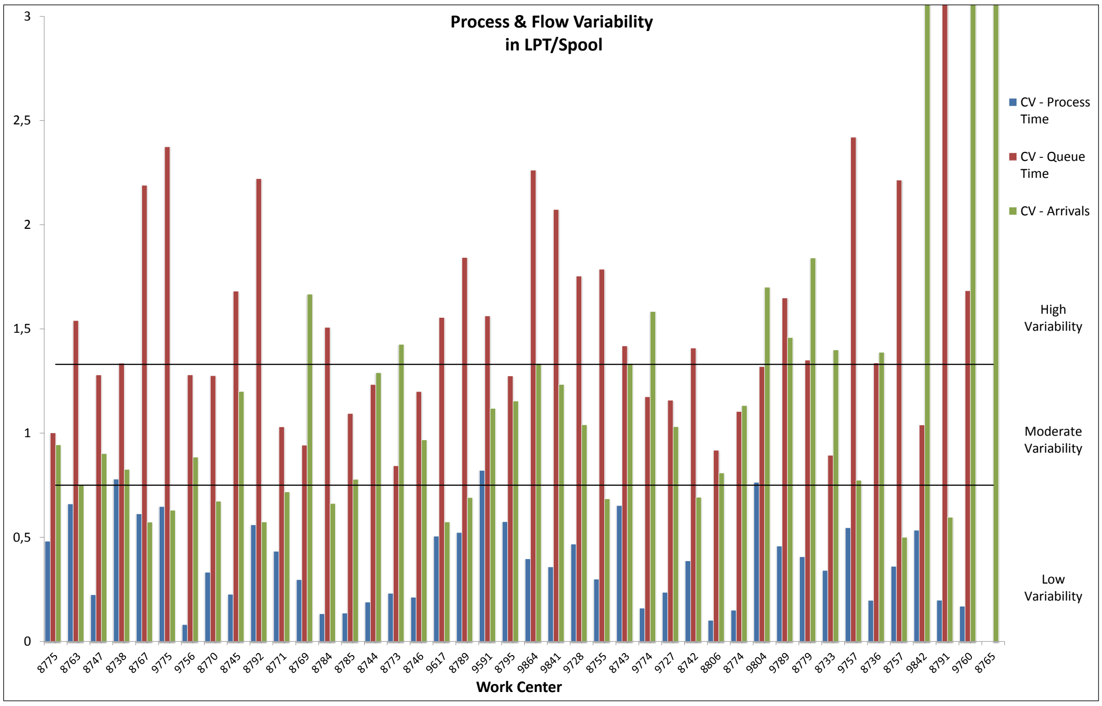

Figure 24: Process and flow variability for each work center in LPT/Spool during 2014.

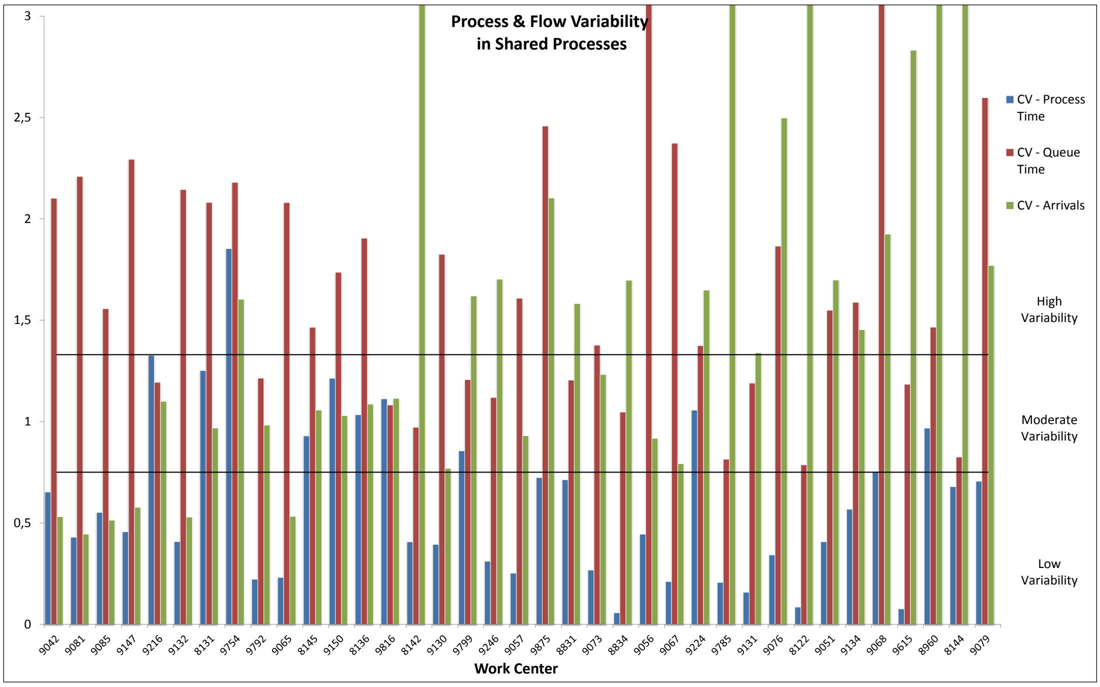

Figure 25: Process and flow variability for each work center in Shared Processes during 2014.

Expected Queue Time

Figure 26 illustrates the expected queue time for each work center in *LPT/Spool*, derived from Equation 9, given the coefficient of variation for process time and arrivals per time unit. The figure models how the expected queue time increases as the utilization at the work center increases. It is important to note that the expected queue time is a theoretical value derived from the real measured values of variability.

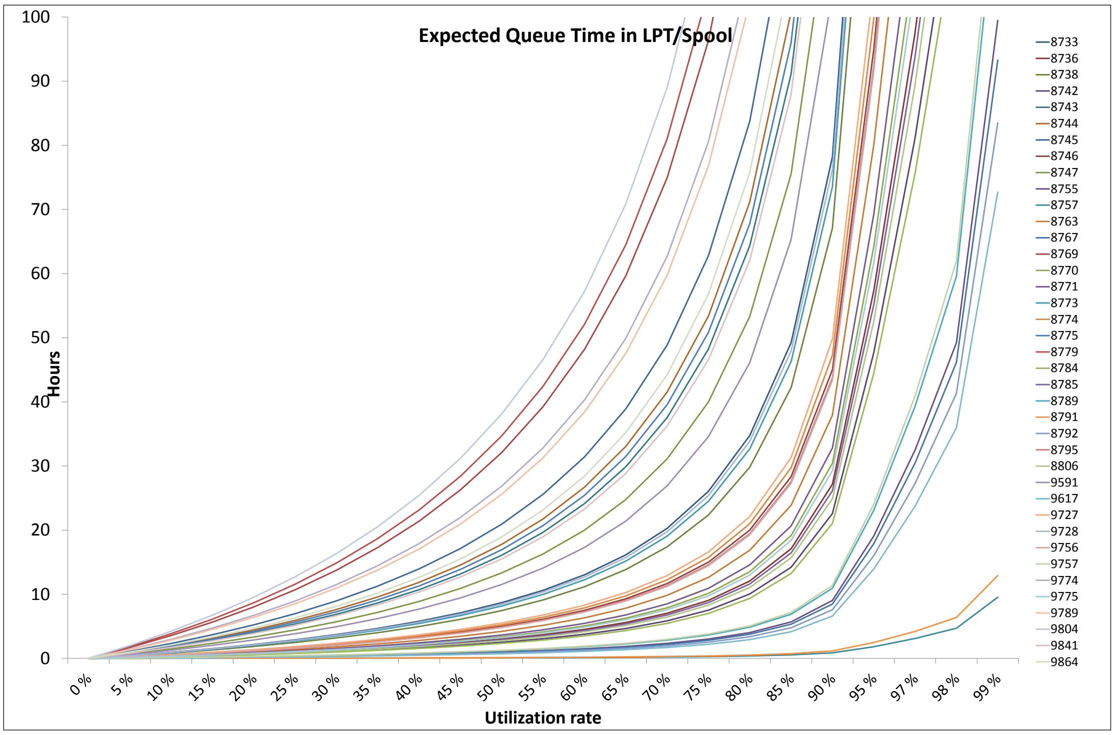

Figure 26: Expected queue time for work centers in LPT/Spool during 2014.

The model becomes less accurate when it approaches extreme values, which is why it is generally more accurate to the center of the figure. Table 2 shows that the model is fully accurate for approximately 30 % of the studied work centers when compared to the real average queue times and capacity utilization levels. Furthermore is the model consistent but not accurate for another 20 % of the work centers, which is caused by extreme values in the model. In cases where the model does not give an accurate result compared to real values, the model always predicts a lower expected queue time than what the actual values are. This shows that the model is not capable of capturing all existing complexity in the production system but is accurate enough to use for approximations and conceptual discussions.

Table 2: Real queue time and capacity utilization.

	Work Center	Real Average Queue Time	Capacity Utilization
Accurate	8738	39,6	84 %
	8743	65,5	73 %
	8771	35,8	87 %
	9591	54,4	84 %
	9756	70,4	98 %
	9774	50,1	66 %
	9775	17,4	82 %
	9841	43,8	70 %
Consistent but not accurate	8736	35,2	102 %
	8747	74,0	102 %
	8757	22,8	98 %
	8769	90,6	98 %
	8770	46,4	123 %
	8791	11,7	116 %
	9728	57,5	76 %
	9757	25,3	104 %

4.3.4 Throughput and Work-In-Process Performance

As mentioned and discussed in chapter 2, the lead time is a central measure when evaluating the company performance, it e.g. steers the position of the *CODP*, how responsive the company can be to changing customer demand and has a great impact on the companies chances to make new business deals[2]. In Figure 27, the lead time, and how it varies, is presented for a selection of GKN´s products. It shows the 30 products with the highest production volumes in 2014. The products presented are all parent products, i.e. final products sold to customers, which implies that the lead time for these products contain the lead time for all their sub-assemblies which also have their own material number in the *MRP*-system. The measure uses thus the tree-structure of the information system, *BOM,* and goes through all ingoing materials and adds their lead time to the total sum. The measure contains all activities from the product entering the system to when it is ready for delivery, such as operation time, queuing, waiting for transportation, extra activities caused by quality issues and time spent in intermediate storage. Transportation is also included but is negligibly small in comparison. The bars show their average lead times in 2014. Furthermore, the thin lines in the figure depict the variation, and show the area that is one standard deviation in both directions from the average.

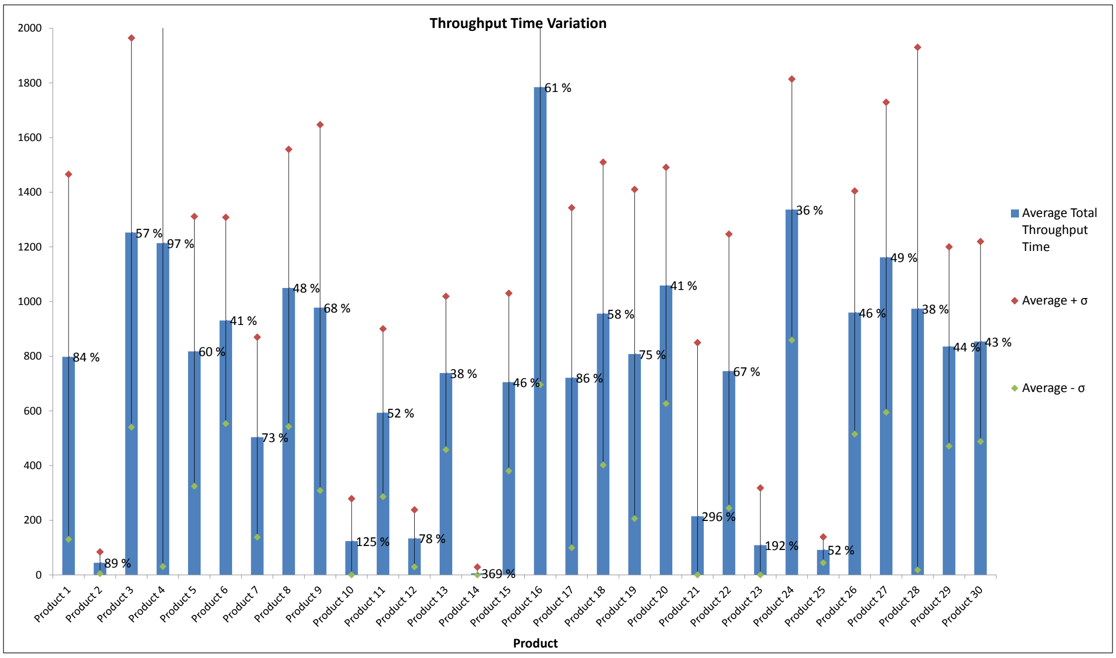

Figure 27: Lead time and lead time variation for the 30 most produced products during 2014.

Lead time also has a direct impact on how much capital the company has to invest in keeping stock, both in the production as *WIP* and before and after production as raw material and finished goods. The capital tied up throughout 2014 can be seen in Figure 28.

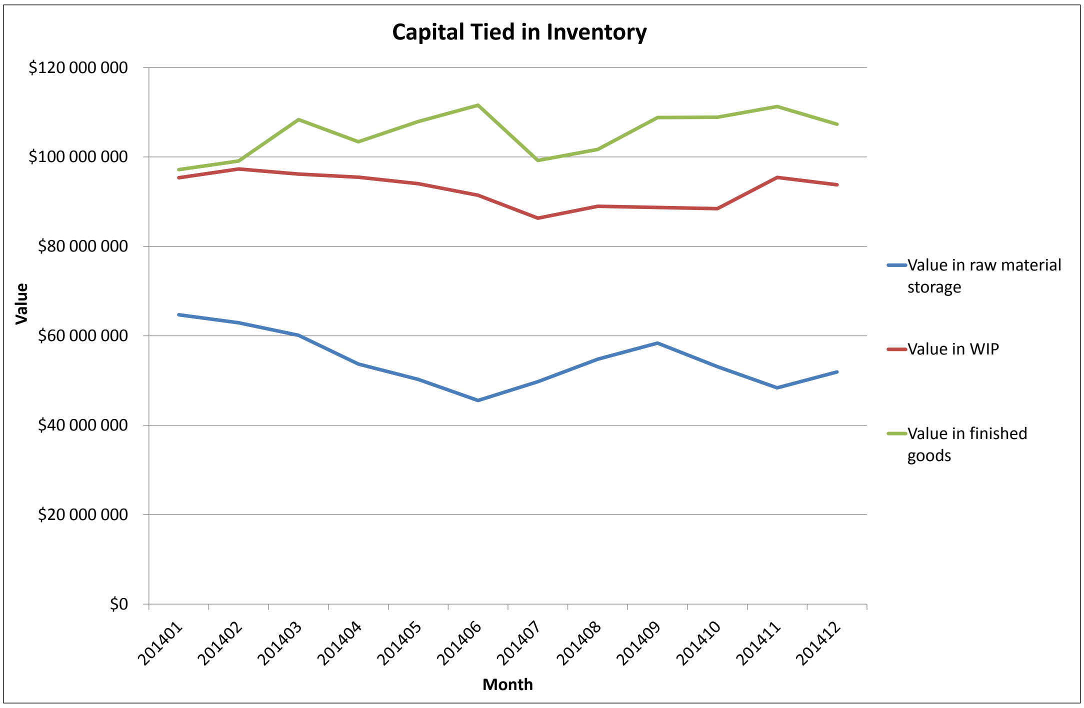

Figure 28: Capital tied in inventory.

The lead time, as mentioned, contains all activities from start to finish, but not all activities are value adding. Non-value adding activities are such that they can be removed without reducing the value of the final product. What makes up the lead time for the selected products, and the extent of these parts, can be seen in Figure 29. The queues contain both the queue before an operation and also the time after the operation before the next, i.e. waiting for transport, transport and intermediate storage. The planned queue time is entered into the information system and represents the time the company expects it will take to prepare the order for an operation after the up-stream operation has finished. This time is taken into consideration in the master production schedule and affects the *MRP*-dates. All queuing is however considered non-value adding since it implies no alteration of the product.

The remaining part of the lead time is made up of operating time, i.e. reported time. However, not all of this time is considered value adding, some part of the reported time is used for correcting quality issues, i.e. *ZP03* and *ZP11* operations, and these can be seen as *Non-value adding extra operations*. Furthermore, activities such as set-up, change over and administrative operations are deducted from value adding activities. According to a study made in the X-shop by Berglund (2014), approximately 44 % of what is left after deducting the extra operations is used for value adding purposes. The true value-adding time is in the graph presented as *Value added time*. This percentage is used as a standard value for all products since no data for other products is available and time studies were not an available tool for this study. The reported time also contains operations such as quality control, which, in a strict sense, are not value adding. These are not illustrated in the graph, which implies that the true value might be even slightly lower. The value added ratio for the selected products varies from around 15 to as low as 2-3 % with an average of 4 %.

Figure 29: Lead time partitions for the most 30 most produced products during 2014.

5 Analysis

The analysis of the empirical findings is structured according to the model presented in Figure 30. The chapter starts with analyzing the effect of the company's design and context in order to understand how much of the capacity is lost in the current state. Subsequently, the analysis continues with interpreting the symptoms that are leading to the effect of lost capacity in the production system. Lastly, the root causes of the symptoms are presented. Since the connections between the symptoms and the effect have been covered, the impact of the identified constraints will only be linked to the symptoms. For the constraints that cannot be directly related to the symptoms, their impact on the lost capacity will be covered when presented. An overview of the major cause and effect relationships is presented in Figure 31 in section 5.4.

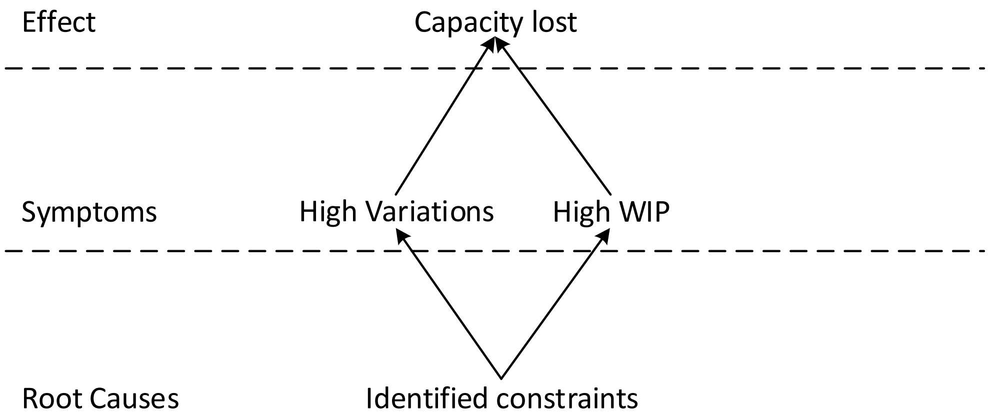

Figure 30: Structure for the analysis.

5.1 Capacity Loss

The production system with its current set-up is considered to be loaded to its upper limit since the company struggles to meet the set delivery dates and that delays are stable over time. Figure 19 shows that the delivery precision is on average 70 % and that the production delay, shown in Figure 20 is stable over time around 60 000 hours. Since the system is loaded to its limit, it is reasonable to expect that the company is using most of its available capacity. However, this figure is in GKN´s case is only 72 % which implies that out of the available capacity GKN loses 28 % that could have been used to improve the delivery performance and reduce the production delay.

This measure is calculated by summarizing the reported operation time for 2014 over all active work centers in the production system, and then divided by the system´s total capacity. This after deducting for all foreseeable and unforeseeable productivity losses covered in section 4.3.2. The procedure is thus the same as for developing Figure 21 and 22, with the difference that it is done for the entire production system instead of for some selected areas.

The capacity loss indicates that there are issues with the production system´s current set-up and that there is potential for increasing the throughput by being able to utilize more of the existing capacity. There is clearly a need to increase throughput in order to reduce the delay and increase the delivery precision, but it is not reasonable to expect the company will be able to use all of the 28 % lost capacity to increase the throughput.

As can be seen in Figure 23, some resource categories are utilized to a higher extent than others, e.g. *Category 5* machines have a utilization rate of 89% while *Category 1* machines are utilized to 75 %. This difference is due to that the current product mix implies a skewed demand for the different resource categories in relation to their available capacity. This implies that a share of the 28 % capacity, that is considered to be lost, is made up of capacity that the product mix and volume does not require. Due to the complexity and poor transparency of the production system it is difficult to estimate how much of the lost capacity can be attributed back to the skewed demand. Hence, it is not feasible to access all of the 28 % lost capacity by changing the setup of the production system.

Furthermore, the insight of not being able to access the entire 28 % is in accordance with the non-linear relationship in Figure 1 presented by Hopp and Spearman (2011), which shows a clear trade-off between high capacity utilization and short lead time. To be able to estimate how much of the lost capacity can be accessed, the company must determine the longest lead time it can accept. Figure 26 shows that there also is a wide spread between the work centers in at what utilization rate the queue time increases exponentially, which is dependent on the process and flow variations they are exposed to. Thus determining how much of the available capacity can be utilized has to be done on an individual work center level, which is beyond the scope of this study. However, it is not reasonable to expect to get access to more than 90 % of the resources since Figure 1 shows that even with minimized variations, the lead time will increase exponentially when approaching this limit. The total potential for increasing throughput is therefore based on a maximum increase in available capacity of 18 %. For the purpose of a clear analysis a conceptual maximum lead time is defined here and will be used hereinafter.

TOC states that capacity should only be utilized if it contributes to the productivity of the system (Rahman, 1998; Lockamy & Cox, 1991). This implies that a share of the capacity loss of 28 % is due to that the system is limited by some constraint and should therefore not be fully utilized. However, by appropriate management of these constraints, both production and managerial, they can be elevated which gives room for an increase of utilization for productive purposes and implies that reducing these 28 % still is feasible and desirable. The identified constraints are presented in section 5.3 and how they should be appropriately managed is covered in chapter 6 *Recommendations*.

Besides getting access to more of the existing capacity to load with operating time, there is potential for increasing productivity by a more optimal utilization of the operation time itself. Figure 29 shows that only a part of the time reported as operating time is value adding, and thus potential for improvements. If the *OEE* of the resources, as defined in section 2.2.1, is increased, more of the capacity can be used for productive purposes, thus increasing the throughput.

The question of why the company loses 28 % of its capacity still remains. The clear connection between capacity utilization and lead time is an important realization. The impact of variations and high *WIP* levels on this relationship, shown in Figure 1 and by *Little´s Law,* make them key candidates as causes for the problems the company currently faces. These two are referred to as symptoms in the *Analysis Structure Model* and their part in the capacity loss is explained and discussed in the sections below. With the available data it is not possible to determine how much of the lost capacity is due to variability, overproduction and natural capacity constraints that cannot be elevated without further investments. It is however, as mentioned, clear there is a substantial potential that can be fulfilled without investing in adding extra capacity to the factory.

5.2 Symptom

This section discusses the symptoms causing the effect of capacity losses. The symptoms identified to have the highest impact are high variability and high levels of *WIP*.

5.2.1 Variability

By studying Figure 24 and 25, it is evident that the company currently experiences high levels of both process and flow variability in the production system. The high variability degrades the performance of the whole production system as stated by Slack et al. (2010), Schmenner and Swink (1998), Hopp and Spearman (2011), and therefore holds for a significant share of the capacity loss of 28 %.

It is apparent that the production system buffers the variability with inventory, capacity loss or throughput time, in accordance to Hopp and Spearman's (2011) *Law of Variability Buffering* and as visualized in Figure 1. Figure 29 illustrates that, for the 30 products with highest production volume, a product spends approximately 90 % of the total lead time in queues, which results in long throughput times. Figure 26 illustrates that both the process and flow variability in the greater majority of the work centers lead to highly increasing throughput times already at a utilization level of 65 - 80 %. Many of the work centers have a higher utilization level than this span, which therefore leads to longer throughput times in reality, as seen in Table 2.

How the variability leads to capacity loss is observed in a comparison between *LPT/Spool* and *Shared Processes*. Since all products passes through *Shared Processes* and only similar products are processed in *LPT/Spool*, the work centers in *Shared Processes* have a generally higher level of process time variability than in *LPT/Spool*, which can be seen when comparing Figure 24 and 25. Furthermore, there are two times more work centers classified as having high arrival variability in *Shared Processes* than in *LPT/Spool*. This should then, according to Hopp and Spearman

(2011) and Slack et al. (2010), mean that *Shared Processes* must have generally higher capacity losses than *LPT/Spool*, which also is the case. A comparison of Figure 21 and 22 shows that *Shared Processes* generally has lower capacity utilization than *LPT/Spool* while Figure 20 show a, on average, 66 % larger production delay for *Shared Processes*. This indicates that *Shared Processes* cannot access the capacity it needs.

There are more indications of that variability affects the performance of the current production system. An analysis of the correlation between the variation in queue time and arrivals per time unit shows a moderately strong negative relationship (-0.52) between them. The logical explanation behind this is if the variability in arrivals decreases, implying a more even flow of materials to the work center, the variability in queue time increases since the queue becomes shorter. An absolute change in queue time gives greater relative change in a short queue than in a long queue, which explains the negative correlation. There is also a weak positive correlation (0.33) between the variation in process time and queue time. Given a constant arrival rate, it is most likely that variability in process time gives rise to variability in queue time and a more un-even and disrupted flow in the production system. This is in compliance with the theory presented by Hopp and Spearman (2011) and Slack et al. (2010).

In short, it is clear that the symptom of high variability gives rise to capacity losses, queues and long unpredictable throughput times. Figure 27 shows that the total throughput time for an order highly varies and it is therefore understandable that the delivery performance is at a low level, which in the end affects the customers.

5.2.2 Work-in-Process

Figure 28 shows that the company has high and fluctuating *WIP* levels in absolute terms. When combining this information with the conclusion from Figure 29 that the lead time consists of 90 % queue time it is evident that the *WIP* levels are unnecessarily high.

The linear relationship between the *WIP* level and throughput time from *Little's Law* (Slack et al., 2010) implies that a reduction of the *WIP* will reduce the throughput time. This is interesting since this linear relationship has an impact on the non-linear relationship between lead time and capacity utilization. Assuming that the maximum acceptable lead time is reached, it is not possible to improve throughput by increasing the resource utilization since this would only increase lead time in accordance with Figure 1 and 26. In this situation, a reduction of *WIP* would shorten the lead time and thus enable more of the capacity to be used and still lie within the acceptable lead time and give room for a higher throughput. This is, in accordance with *TOC,* of course only true if there are *WIP* levels that are higher than what the capacity constrained resource in a product flow can process within one standard lead time.

Since the capacity constrained resource per definition is the weak link in the product flow there should be no queues for the rest of the resources in mentioned flow, if the *WIP* level is appropriate. Unnecessary *WIP* creates new capacity constraints that reduce the amount of existing capacity the company can access. An example of such a situation is the long waiting time for transports, as described by Berglund[8]. The reason for this is most likely the high level of *WIP* in queue that causes lack of space capacity in front of a work center or in the shared storage area among work centers. This means that material must be transported e.g. from X- to C-shop and back again for intermediate storage, which requires added transport and potentially making transport a constraint.

The long and highly varying lead times illustrated in Figure 27 are creating problems for the company. Due to this situation a great extent of the production must be based on forecasts instead of actual demand. The long lead times results in poor responsiveness of the production system which, together with uncertainty in forecasts, creates the need for high levels of WIP in terms of overproduction, buffers and safety stock in order to be able to meet the actual demand. This is consistent with Schönsleben's (2007) theory, which implies that production volumes deviating from the actual demand only create waste. The capacity cannot be used optimally since it is used for products that are potentially not needed instead for product actually demanded. This means that capacity actually needed is lost. If the lead times would have been shorter, more of the production could be based on actual demand, which would reduce uncertainty and thereby *WIP* levels.

Figure 24 and 25 show high variations in queue time for almost all work centers, which implies there are queues in front of several resources in the same product flow. This indicates that the company has unnecessary high *WIP* levels and that there is room for reducing the throughput time without reducing the load on the *CCR* and thus increasing the throughput. The vast production delay and the poor delivery precision indicate that the maximum acceptable lead times have been reached which further confirms that the company's high *WIP* is one of the major reasons behind the capacity loss.

5.3 Identified Root Causes

This section discusses the identified constraints that give rise to the symptoms variability and *WIP*. The identified constraints are divided into physical, managerial and situational.

5.3.1 Physical Constraints

The high level of *WIP* obstructs the identification of production bottlenecks due to existing bottlenecks dynamics. In conformance with Shen and Chen (2010), this means that the bottlenecks move around and shift place in the system due to environmental changes and issues that are obstructed by the high levels of *WIP*. Since the company's production system is exposed to large complexity and high variations. Shuch et al. (2012) mean that the only way, in this situation, to identify a potential

CCR is to use continuous real-time methods or simulation modeling, which are not currently available to the company. This implies that to be able to identify physical bottlenecks the complexity, system variations and *WIP* first have to be reduced, in accordance to what is discussed in section 2.2.3.

Moreover, the high level of *WIP* creates new constraints that normally are non-constraints and should therefore not affect the production performance negatively. Another factor that limits the production system is the lack of NC-programmers. NC-programmers, who influence the production system, would be considered as a physical constraint according to *TOC* (Cox & Scheilier, 2010; Schönsleben, 2007). The lack of programming resources leads to the current set-up of the production system has lower flexibility than the theoretical one. This implies that there actually is capacity available that could relieve a constrained product flow, but cannot be used since potential machines are not programmed or certified to perform the needed operations. The direct effect is that the current set-up therefore generates more complex product flows than necessary, which results in higher variability leading to capacity loss. The lack of flexibility also makes it harder to simplify the product flows using new routings or by a cell layout.

Since the logisticians' responsibility is divided into functional areas and does not cover a full product flow, this leads to higher variability and a disrupted production flow. Prince and Kay (2003) implies a functional responsibility structure means logisticians only focuses on managing their own processes instead of influencing and managing the whole production performance of products. This also seems to be the case at GKN. The fact that each area has different production rates[4], tries to keep as much production hours themselves, and a non-formalized communication flow for handovers, result in an increasing variability in the production flow. This clearly generates a sub-optimized behavior.

Today´s functional layout generates a complex product flow, but given the history of the factory and the fact machines are hard to move, it is currently the only available physical layout. Slack et al. (2010) points out that this enables high machine utilization, which is what GKN actually strives for in order to spread out the time cost[4]. However, this leads to long lead times and high WIP, which is in full compliance with Slack et al. (2010) who states that the company must accept these effects if they want to maintain a high utilization.

GKN's production qualifies only for two out of four criteria for using a functional layout stated by Slack et al. (2010). The company is not in a *MTO* situation or has a wide range of products. According to the theory, this implies that a functional layout, disregarding the given constraints, is not the optimal layout. There are indications that the current functional layout and push system is not fitted to GKN's business model. A functional layout and push system striving for high utilization is not responsive enough to meet the requirements of the business situation the company is in, since the production system should be able to fulfill a *PO* on short notice[2]. With its focus on high utilization that causes high *WIP* and inertia, the production system does not match the demands of the business deals that premier high delivery precision and therefore responsiveness and flexibility.

The fixed product routing, where individual work centers are predefined[7], indicates that each machine is its own planning point. This is in compliance with Prince and Kay (2003) who states that a functional layout therefore requires a complex planning and control system. When the load on each machine has to be planned individually it gives rise to a pushing system, which itself breeds for long lead times and high *WIP* in order to meet the customer demand (Prince & Kay, 2003).

5.3.2 Managerial Constraints

In this section, constraints that have been identified to stem from managerial principles, policies and decisions are covered. This section is thus key in linking overhead business decisions to productivity losses in the production system.

Constraints in the Production Logic

It is clear that for both contract types, *LTA* and *RSP*, production does not take place in order to fill firm customer orders but to only ensure availability. This implies that the start- and finish dates in the *MRP*-system, which steers and controls the entire production system, does not really exist. They are decoupled from the actual customer demand and, as mentioned in section 4.2.3, derived from the master production schedule with the objective to achieve a stable production rate for all products. This creates unnecessary turbulence in the production system since prioritization among orders is made according to the dates and not the actual demand. Resources, as seen in e.g. *heat treatment*, are idle waiting for a prioritized order to be finished at an upstream resource while they have other orders in queue, and valuable capacity is lost.

The start dates only represent, as mentioned, the latest start date of an order which gives room for the logistician to release the order early which at length will lead to unnecessarily high *WIP* levels. The incentives for the logisticians to actually do so are covered further down. Furthermore, the infinite loading logic that is unavoidably inherited from the *MRP*-system also leads to overloading the *CCR*, which according to *TOC* creates unnecessary physical constraints that limits the system and leads to capacity loss. This can be seen as the master production schedule, in the long term, does not take the actual capacity into consideration when loading the system.

Business model and governance

The routing times, as defined in section 4.2.1, are the base for cost calculations when evaluating potential and existing business contracts. Since adherence to this routing time is used to evaluate logisticians and production performance it creates incentives for the logisticians to not increase the routing time even if they know it is too short. This implies that the logisticians add a markup to the routing time when planning the load which gives room for variations, both between logisticians and in decisions from time to time. This causes disruptions and differences in production rate in the production flow and, according to what is discussed in section 2.1.2, at length arrival rate- and queue time variations. Using too short routing time when matching load to capacity leads to a higher load than the *CCR* can handle and unnecessary high *WIP* levels, which can be seen in Figure 21 and 22 with utilization rates above 100 % and in Figure 24 and 25 with high queue variations for almost all work centers. This and the *1000 SEK factory* goal are good examples of the, according to Abdel-Maksoud et al. (2005), inappropriateness of using traditional performance measures based on accounting and time costs when evaluating production performance. *The 1000 SEK factory* goal uses local efficiency as objective function and not overall product flow. This creates incentives for increasing capacity utilization without it leading to increased productivity, and according to *TOC*, this leads to increased *WIP*. It also causes sub-optimization and rivalry in the interface between different responsibility areas, which will cause variations and disruptions in the product flow.

The same philosophy can also be found in the budget evaluation of the responsibility areas. If the area reports fewer hours, in routing time, than what they should according to the master production schedule it may be seen as they have overcapacity and will be a candidate for rationalizations. Since the routing time is wrong, the area will look bad even if they are fully loaded or run into problems. This adds on to the incentives to keep all the work in the area, caused by the functional layout, even if they have problems and cannot produce, instead of seeking help from other areas where excess capacity might be available. This is a direct loss of available capacity and will also create disruptions and flow variations that will spread and amplify downstream in accordance with Hopp and Spearman (2011). Furthermore, it causes the need for gate resources to pressure the logisticians to release orders that are not due to be released since they feel the need to report the hours. This is possible since the *MRP* has no earliest release date and it is thus up to the logistician to decide. As mentioned in section 4.2.3 this often occurs, which results in a load that is too large for the *CCR* to handle and thus too high *WIP* levels in the system.

The evaluation of individual operators also creates problems. Since the operators report the operating time themselves, and are evaluated based on the time reported, it creates incentives to exaggerate this time in order to show that they are needed. This implies that it looks like the capacity utilization is higher than it actually is and that available capacity might remain unloaded and thus lost. This operating time contains both the runtime in the machine and the set-up- and teardown time. It implies that it is not possible to determine how much of the reported operating time is lost to non-value adding activities such as set-up, in accordance with Slack et al. (2010) and what is discussed in section 2.2.1.

Since set-up- and teardown time are not measured, it is impossible to improve the *OEE* and create standardized methods for set-up and tearing which, since the machining operations are highly automated, causes much of the process variations. By not measuring all parts of the reported operating time, the productivity loss calculation has to be based on assumptions and template values and not actual data. This imply a risk of overestimating capacity, which in Figure 21 and 22 can be seen to be the case with utilization above 100 % in several work stations, and thus overloading the system with too high *WIP* levels.

Furthermore, much of the knowledge of the machines´ capabilities is tacitly owned by the logisticians and not available for routing planning. This leads to lost flexibility and capacity since decision makers, who could have re-routed products to existing idle resources and relieve an overloaded resource, do not know of the possibility.

Planning and Control

Logisticians are as mentioned pressured to release orders early and do so if incoming material is available. This implies that they are loading non-capacity constrained resources over their productive capacity, as discussed in section 2.2.1, and thus according to *TOC* only create queues and high *WIP*. This since the capacity of the *CCR*, the resource setting the pace for the entire product flow, is not taken into consideration. It also creates flow variations since the logistician prioritizes to keep the gate resource busy above ensuring a stable production rate. According to the *law of variability placement*, discussed in section 2.1.2, introducing variability early in the flow will create amplifying variations downstream in the production system.

These issues commonly occurrs in pushing production systems (Schragenheim et al., 2009) where materials are handed over to the downstream work station or responsibility area when they are finished in the upstream resource, without an actual need for it. This issue is noticeable in the way the daily planning and control is done at GKN. Each logistician and responsibility area has their own procedure and logic when loading actual jobs onto the work stations. For example, the fixed schedule and production rate in the fabrication shop does not take into consideration the downstream needs or the upstream capabilities. The resulting separation from the rest of the product flows gives rise to disruptions and high flow variations, as it is seen in Figure 24 and 25.

The fact that the logisticians are not responsible for entire product flows also leads to different prioritization rules in different areas. The planned advanced algorithms for some resources will lead to further disruptions and sub-optimization since the objective function is local efficiency and not global efficiency which would, according to *TOC*, imply subordination to the *CCR*. Conflicting prioritization rules and the option to mark orders as *prioritized orders* implies variations and disruptions in the production rate in the interfaces between different responsibility areas. This creates overloading and queues or material shortage and idle capacity. Buffers and higher *WIP,* seen in Figure 29 where queues make up around 90 % of the lead time for all products, is used to compensate for the variations in the product flow.

5.4 Analysis Overview

In Figure 31 below, a visualization of the identified cause and effect relationships between constraints, symptoms and effect is presented. This figure is a further development of Figure 30 and can also be seen as a summary of the analysis.

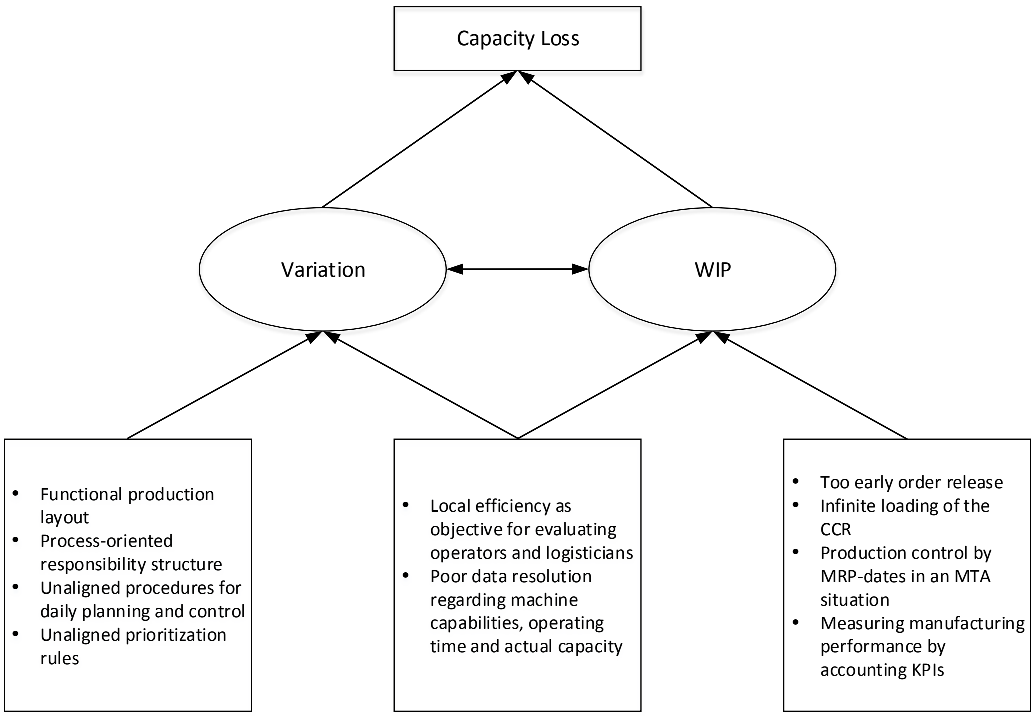

Figure 31: Visualization of cause and effect relationships and analysis summary.

6 Recommendations

This chapter describes and explains recommend manufacturing actions for GKN necessary to adopt to increase throughput in production. Figure 32 illustrates the framework that defines the recommended solution. The roof symbolizes the goal, *increased throughput* through *reduced variability* and *WIP*, and it rests on the block of a *pulling product flow*. This is supported by the required pillars, *product flow* and *production planning and control,* which are made up by a number of principles and methods. These are supported by a *governance and performance measurement* culture that corresponds to the company's *manufacturing situation*. The recommendations are summarized as an action plan, divided into short and long term actions, at the end of this chapter.

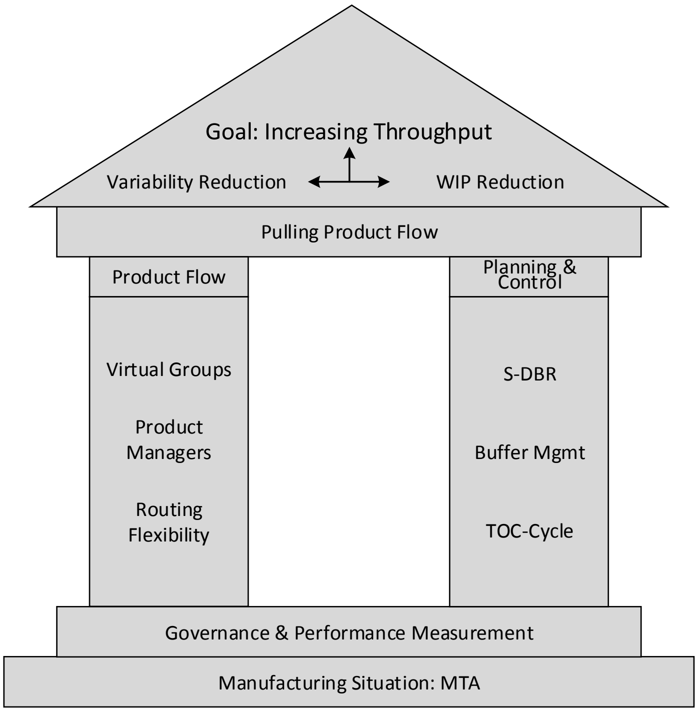

Figure 32: Framework of the recommended solutions.

6.1 Product Flow

Current product flows in the factory have high complexity that must be reduced to access the lost capacity. A new flow layout is therefore recommended. The new layout is based on three different parts: *virtual groups*, *product management* and *routing flexibility*.

6.1.1 Virtual Groups

Since today's functional layout and push system do not correspond to the demands from GKN's business model and according to Slack et al. (2010) is not the most optimal one, it is recommended for the company to strive towards a cell-oriented product flow where focus lies on reducing lead time instead of high machine utilization. However, given the physical constraints in the factory, a physical cell-oriented layout is not possible. It is therefore recommended for the company to strive towards an implementation of *Virtual Groups* (Prince & Kay, 2003), as explained in section 2.4.2. The forming of *VGs* is done by using the *EPFA*-methodology described in Appendix I.

The concept of *VGs* fits GKN's situation and characteristics well. This cell-oriented arrangement supports a simplified planning and control system (Prince & Kay, 2003), needed to reduce the complexity of the production system. The use of *VGs* enables a planning point for a whole cell or product flow instead of individual machines, which reduces complexity and supports the appliance of a pull system.

The use of *VGs* will group together similar products and flows and therefore result in reduced process variability. Reduced flow variability is also achievable since *VGs* will lead to more even and less intersecting product flows. GKN will be able to maintain a process orientation if needed, e.g. in the *heat treatment shop,* since the concept of *VGs* allow machine sharing (Prince & Kay, 2003). In short, *VGs* will enable the company to move towards point C in Figure 1 (Slack et al., 2010), higher utilization and shorter waiting times, since process and flow variability is reduced. Reduced variability and complexity in the production system gives shorter lead times and also leads to reduced levels of *WIP*.

6.1.2 Product Managers

It is recommended for the company to change the responsibility structure for the logisticians in order to reduce variability, handovers and disrupted flows. Since *VGs* focuses on the management of products (Balakrishnan & Cheng, 2005), even though in a physical functional-oriented layout, logisticians will focus on the production of products instead of managing processes.

Having product managers instead of process managers will lead to less disruptions and sub-optimization since department borders are erased and focus lies on the total throughput for a whole flow instead of production hours within a department. The existing protectionism among the departments is therefore likely to be decreased. This new responsibility structure enables a more even product flow from reduced flow variability since the number of handovers and barriers are reduced. Load planning and prioritization rules will be covered in section 6.2.

6.1.3 Routing Flexibility

It is recommended for the company to increase the routing flexibility in the production system since capacity available is constrained by low machine flexibility. Improved machine flexibility is also required to provide as many options as possible when sorting machines into virtual groups so that an optimal layout, in terms of complexity and flow, can be reached. This requires more capacity in terms of NC-programmers or a more streamlined programming process. More options when designing new product flows reduces intersecting and conflicting flows, constraints and thereby variability and queues. Some operations performed in one resource category might also be possible to perform in another, e.g. some lathing operations can be performed by milling machines. This implies that increased routing flexibility also reduces the gap in utilization rate between different resource categories, thus countering the issue with the product mix´s skewed resource demand.

6.2 Production Planning and Control

Since the way the production is managed has been identified as cause for a large share of the lost capacity, new planning and control procedures more appropriate to the company´s situation is therefore recommended.

The company is recommended to implement an *S-DBR* logistical system that, in accordance to what is discussed in section 2.3.2, facilitates a pulling product flow. *S-DBR* will reduce variations (Hopp & Spearman, 2011) and cut lead time and *WIP* since production volumes are coupled to the actual customer demand, (Schragenheim et al., 1994), which is made possible by the transition to a *VG*-layout. As discussed in section 2.3.2, a *JIT*-system would also facilitate a transition to a pulling production system but does not fit the company's situation. The corresponding *Kanban* system is best suited for high volume production with fairly few product variants, which is not the case for GKN. Furthermore, it fails to recognize that the market is a constraint in the same sense as internal constraints. An *S-DBR* system is thus recommended since it fits the company situation and implies the biggest productivity potential.

As discussed in section 5.3.2, it can be concluded that GKN´s production is to ensure product availability, with all *WIP* tied to specific customers. This clearly places the company in an *MTA* manufacturing situation. As covered throughout section 2.3, the *S-DBR* system plans and controls the production after buffer levels and not set dates as in the *MRP*- system. Since the company cannot completely reach an *MTO* situation, in which there are fixed delivery dates, the new system will solve the identified constraints caused by the *MRP* dates. Some aspects of the existing *MRP*-system are however still necessary. The *BOM*-structure is still needed to develop the procurement lists and the connections with the suppliers. The resulting system with the combination of the existing *MRP* logic and the new *S-DBR* production system is similar to the *Demand Driven MRP*, described in section 2.3.2.

6.2.1 Load Planning

As stated in section 2.3.2, it is reasonable to set the initial lead time when transitioning to an *S-DBR* system, to half the lead time in the current system. This seems feasible since Figure 29 shows that, on average, 90 % of the current average lead time consists of waiting time. Due to the linear relationship between throughput time and *WIP*, presented in Equation 1, this also implies a 50 % reduction of *WIP* to a value of around USD 40 million. Shorter lead time implies the *CODP* is moved upstream and that less production will be based on forecasts which, as covered in section 2.3.2, will enable the use of simpler planning and control procedures. This implies more stability and predictability in the production system. The active forecast period would be closer in time and therefore more accurate, which will decrease the problems of capacity being used for unnecessary production. This frees up capacity that can be used for products for which there is an actual demand.

In accordance with *TOC* and as made possible by the product orientation of the *VG*-layout, the *CCR* is the only planning point in a product flow. Bottlenecks are, as discussed in section 2.2.3, shifting but since variability and *WIP* are reduced, discrete methods based on past data can be used for *CCR* identification. Frequent runs of the *TOC-cycle* should be used to identify the current bottleneck and thus where load should be planned, and where improvement efforts should be directed. The *TOC-cycle* is, as discussed in section 2.2.2, a structured method for collecting and analyzing data in order to continuously improve the productivity of the *CCR* and thus the production system. By using it, the need for better data quality, e.g. the need for measuring set-up times, will become evident which will solve the identified issues with lacking or poor information.

The identified *CCR* should be loaded up to *CAP100* minus the productivity loss as in section 4.2.3. and a safety margin, as discussed in section 2.3.2. There is no fixed schedule but only booked time slots as described in section 2.3.2 and it implies a finite loading philosophy which solves the issues with overloading, queues and artificial constraints. This suits, in accordance with Schragenheim et al. (2009), GKN´s situation with expensive machines and specialized work force. Productivity loss based on actual data also reduces the identified risk of overestimating capacity and overloading the *CCR*. If the *CCR* is shared by two or more product flows, the product managers will all book the time slots until the load limit is reached, hence requiring communication between them.

6.2.2 Production Control

Production control and all decisions should, as mentioned, be based on the buffer status for the product in question, as described in section 2.3.2. Buffer management is a three-step process in which the first step is to determine the buffer target value, i.e. buffer size. Product 21 is used as an example.

The current average lead time for said product, as can be seen in Figure 27, is 214 hours. Applying above logic, the lead time in the new system will thus be 107 hours, which corresponds to a full week with the most common *GKN3* shift. The buffer target value is calculated by applying Equation 15. The maximum demand, in the demand plan, within one lead time during 2014 was 4 products. The uncertainty factor is derived from the standard deviation of the lead time, assuming a variability coefficient of 0.5, and added 2 products to the target value. With the recommended layout and production steering, the variations are assumed to lie in the *Low* section so a *CV* of 0.25 is seen as reasonable. The target value of the production buffer, made up by *WIP* and finished goods, should thus be 5. This target value does not reduce the *WIP* levels for said product but does ensure a 100 % availability and delivery performance by reducing the cycle time and increasing the throughput. The delivery performance was, during 2014, averaging around 50 %, which implies that a doubling of the performance is possible without large increases in *WIP*. This is a crucial improvement since delivery performance is, as discussed in section 4.1.2, a central aspect for reaching new business deals. A similar buffer system should be implemented for serving the raw material storage, but this require deeper analysis of the supplier network and is thus left to further research.

The second step is to operationally manage the buffer. If the buffer level for *Product 21* falls below 5, a new order should be released into the system. However, orders should be released only if there is an available time slot at the *CCR* within one standard lead time, as in depicted in Figure 8. The main logic behind *TOC* states that in a product flow, there should only be a queue in front of the *CCR*, which implies the resources before the *CCR* will only pass on material to the downstream resource if it is available. An order is thus entered into the system first when the gate resource is available. This will also reduce the variations early in the product flow, which according to Hopp and Spearman (1996), will have a great positive impact on the variations in the entire flow. The order release is now tightly paced by the *CCR* and thus the product flow´s capacity. This solves the issues with overloading caused by the *MRP* logic and individual decisions made by the logisticians. Orders are now pulled into the system rather than being pushed out.

Prioritization at the *CCR* should be done according to the buffer status, calculated by Equation 16, which solves the issues with conflicting prioritization rules. An order with buffer status red has highest priority and yellow orders will be prioritized over orders in the green zone. Local algorithms can be allowed if their objective function is to maximize product output instead of local efficiency. These are also only in question for the *CCR* since no other resource should have multiple product units in queue at any given time. The simplified prioritization procedure is supported by the product orientation in the *VG*-layout.

These buffer management systems completes the transition away from the derived start- and finish dates in the *MRP*-system and connects the production tightly to the actual customer demand. First when the customer withdraws a product from the finished goods storage, a new order is generated. This system, similar to a *CONWIP* system, fully recognizes the market as the superordinate system constraint and opens up for the benefits of a pulling product flow and facilitates a stable and predictable production.

The issues with local procedures when loading specific jobs in the daily planning and control, described in section 4.2.3, is resolved since the steering is organized around the buffer status of the products in question. There is no need to strive for keeping a stable production pace locally since the *CCR* sets the pace and one logistician is responsible for the entire product flow. This leads to fewer disruptions, less variation and less need for local safety buffers.

The third step is to adjust the buffer target value if needed. As mentioned in section 2.3.2, stagnation is undesirable. Target values should be lowered when the buffer status is stable in the green and increased if it is stable in the red are. The latter with caution due to reasons described in section 2.3.2. When the new system is settled in, it is reasonable to expect that in the long run, the lead time can be reduced beyond 50% and thus further reducing the *WIP* levels.

Beside lowering *WIP* and reducing variation, the new production steering logic and production layout will also create a predictable and transparent system where problems are easily detected and lead times are stable, which by looking at Figure 26 clearly is not the case in the current situation.

6.3 Governance and Performance Measurement

To facilitate the transition from the current state with the old design to the future state with above recommendations incorporated in the new design, governance and performance measurement must be aligned with the new production principles and philosophy.

A change in behavior is crucial for a successful implementation of the new design. The corporate culture must support collaboration, honesty and sharing of information. Such a culture change must start in top management and be spread to managers on all levels and requires a rich communication of the new values, both in volume and channels. The product-oriented philosophy must also be supported by a product-oriented incentive structure and performance indicators, discussed below. The focus in evaluation must shift from local efficiency to the performance of the entire product flow. Measuring and evaluating the hours an operator or production area produce creates, as discussed in section 5.3.2, only incentivizes to e.g. release orders early, report false data and avoid asking for help when in trouble. In accordance with the reasoning throughout section 5, this leads to high *WIP*, large variations and thus poor performance of the entire system.

As it has been pointed out, it is not recommended to use accounting principles as the base for evaluating manufacturing performance. The *1000 SEK factory* goal and using time cost as performance indicator should thus be discarded. Improvement efforts should be directed to the constraining resource and if local efficiency needs to be increased, in order to improve the throughput of the system, the focus should lie on reducing non-value adding time, improve quality and machine maintenance and thus the *OEE* of the resource. This will in fact reduce cost, but using cost as the objective function will only create incentives for producing as many hours as possible, which is counterproductive.

The company should implement a performance measurement strategy with KPIs mainly focusing on production flow and lead time, instead of local capacity utilization and efficiency. It is therefore recommended to use new capacity and utilization measures that support the flow theory, e.g. the productivity measures arrival rate r_a

and processing rate r_e described in section 2.1.1, instead of focusing on machine hours. This measure will reveal the actual capacity rate and utilization rate of a resource, given its current product mix, which makes it easier to visualize constraints in product flows. Moreover, the company should start measuring both process and flow variability to reduce their disruptive impact on the product flows.

By the same logic, logisticians should not be evaluated on how they adhere to the routing times since this implies them being evaluated by cost. The objective to reduce cost ought not to be achieved by simply shortening the routing time, since incorrect routing times disrupts the production planning. It should instead be based on real data from the production. The routing time can still be the base for estimating costs for quoting and evaluating business deals but the evaluation of these deals should be decoupled from the production steering. Since the logisticians will be responsible for an entire product flow, they should instead be evaluated on throughput and delivery performance, which supports a pulling flow and fits well to the *MTA* situation the company is in.

In summary, local efficiency and high capacity utilization is pointless and even counterproductive if it does not improve the productivity of the entire system. The above recommended governance, evaluating philosophy and corporate culture ensures that all improvement efforts have the overall system performance as objective function.

6.4 Action Plan

The recommendations are presented as an action plan with both short- and long-term actions. The short-term actions give quick results and do not require any capital investments. Actions on the long term are more complex and require larger adaptations and system changes. The short-term actions and their cause and effect relationships are visualized in Table 3. The long-term actions can be found in Table 4.

Table 3: Visualization of the recommended actions in the short term.

Time Frame	Action	Reason	Impact
Short Term	• Align planning, prioritization & loading procedures	• To avoid disruptions and sub-optimization	• Less disrupted flow, reduces variations and enables simpler planning and control principles
	• Develop new KPI- and performance measurement strategy • Decouple business performance evaluation from production management	• To focus on flow and throughput instead of capacity utilization and local efficiency • To avoid creating inappropriate incentives for improvement efforts	• Reduces *WIP* and variations and enables simpler planning and control principles
	• Finite loading of the resources • Do not release orders early	• To avoid overloading the production system • The non-constraint gate resource will not overload the product flow	• Reduces *WIP*, shortens lead time and enables simpler planning and control principles

Table 4: Visualization of the recommended actions in the long term.

Time Frame	Action	Reason	Impact
Long Term	• Invest in programming resources • Implement a *Virtual Group* layout	• To increase routing flexibility and enable a *VG* layout • Reduce complexity and support a product oriented responsibility structure • Enable simpler planning & control principles	• Less disrupted flow and reduces variations
	• Align the corporate culture to the new production philosophy	• To support collaboration and reduce protectionism • Enable better decision making through better data quality	• Reduces *WIP* and variations
	• Introduce a *TOC* philosophy • Replace the *MRP*-logic with *S-DBR* planning and control	• To align all decisions to support the market orientation • To support the *MTA* situation and avoid the *MRP*-date problems	• Reduces *WIP* and shortens lead time

7 Discussion

This chapter discusses the results of the analysis and the recommendations presented to the company. It furthermore explains the generalizability of the results and how other companies in a similar situation as GKN can benefit from the results. Lastly, discussion regarding how the findings contribute to the theoretic field, and suggestions for future research is provided.

7.1 Results

The results of this study and the empirical data presented in section 4.3 show a fairly expected behavior, based on the theory foundation in this field. The complex product flow that should be present in a functional layout can clearly be seen in the production system in question, and the expected process- and flow variations are found in Figure 24 and 25. The theoretical relationship between capacity utilization and lead time that is depicted in Figure 1and described by several researchers (Hopp & Spearman, 2011; Slack et al., 2010; Schönsleben, 2007), is also validated in the empirical data with Figure 26 accompanied by Table 2. The anticipated impact of variations on this relationship is clearly proven in the comparison between the two out responsibility areas in section 5.2.1. The notion that the pushing logic of an *MRP*-system with start- and finish dates based on forecast creates excessive need for safety buffers and incentives for overloading the system with *WIP* is shown to be true in section 4.2.3. The expected result with long and unpredictable lead times, large production delays and poor delivery performance is empirically found and shown in Figure 19 and 20 and in Figure 27 and 28. By showing that the production system in question actually behaves according to what is expected of it, it is reasonable to believe that the solutions for the resulting issues also are valid for the situation at hand. Therefore are these suggested solutions the base for the recommendations developed in this study, and the framework shown in Figure 32.

The analysis clearly shows that there is potential for increasing the throughput by changing the setup of the production system and without having to invest in more production resources. However, the study has not been able to isolate the effect of the identified constraints and how much of the 28 % that actually can be accessed. Since the constraints are interrelated it would require simulation or a similar method to isolate the effects, which is something that might be interesting to do before any major changes is undertaken. The recommendations give a holistic framework for how to improve the throughput at GKN, but do require large alterations of the production system and a major change effort. However, the principles and logic behind many of the recommended changes can still be applied to the existing system. By initiating the transition towards the future design by implementing these simple rules, a part of the improvement potential can be reached without the major effort.

Another interesting aspect that has not been taken into consideration in this study is how much the capacity has to be increased for the company to catch up on the production delay. By analyzing this, it would be feasible to answer the question if the 28 % would be enough or if it is inevitable to invest in more production resources in order to reach a 100 % delivery performance.

However, there are aspects of how empirical data have been collected and synthesized that can be questioned. The quantitative part of this study relies solely on secondary data and the quality of this data is therefore crucial for the reliability of this study´s results. The data is collected from several internal database sources and through triangulation and cross-referencing, it can be concluded that all measures and calculations in this study is based on the same, and the most updated data. The base for the results is past data from 2014, which naturally does not fully represent the current situation. However, data from only 2015 would be too small of a sample and natural fluctuations over a full year would be missed. Data from further back in time would only reduce the resolution and accuracy of the analysis since the production system is changed repeatedly.

Aspects that have come up in the qualitative parts of this study give reasons for questioning this data, e.g. that operators manually report the operating time and that the culture in the company gives incentives to adjust the data. Although this is the case, it has not been possible to collect primary data from time studies and detailed observations since there were clear indications that it would lead to a potential conflict with union representatives.

7.2 Contributions to Companies in General

Many of the constraints identified at GKN are due to the functional layout and to the *MRP*-system´s way of planning and controlling the production system. This set-up is commonly found in other companies so the recommendations given in this study might contribute to the result of other firms in the same situation. The reasoning around governance and performance measurement is applicable to companies even if they are not set up similar to GKN. Difficulties when transitioning from one design to another affects all firms going through change. This implies that the recommendations given around how the corporate culture has to be aligned with the new production logic and objectives can be applied broadly. However, there are certain characteristics of GKN´s production that are very specific. Firms with higher production volumes and less product variants are more suited for a *JIT* approach with a *kanban* logistical system instead of an *S-DBR* solution as recommended in GKN´s case. It would be interesting to do a collaborative change effort together with other companies in the same situation as GKN. This since the framework does not have so much empirical data to support the recommendations and that it thus probably require some further development during the implementation. By making a joint effort the participants would be able to exchange experiences and spread out the costs for implementing a system that is not fully developed.

The key for determining the applicability of the recommendations on another production system is thus how it is set up and if it is similar to the set up in GKN´s case.

7.3 Contributions to Academia and Future Research

Theory in this field is detailed and specifically concern a small fraction of the issues a company in this situations faces. The main contribution of this study is to connect different specific procedures and theories with a more comprehensive framework as a result.

Nomden et al.'s (2006) extensive taxonomy of *VCM*, including virtual cell-layouts and *VGs*, point out that previous research have focused on design and not empirical research, which also is the perception of the authors of this study. Most of the articles identified (e.g. Prince & Kay, 2003; Rehault et al., 1995; Kannan & Gosh, 1996) describe how companies with a functional-oriented layout, in a similar situation as GKN, can design a virtual cell-layout. However, previous research seems to lack substance on how the virtual cell-layout should be linked with actual steering principles of the production. This study shows how a company, with similar characteristics as described by Prince and Kay (2003) and other researchers, can address the problems that a constraining functional layout causes, by using a virtual cell-layout and connecting simplified steering principles, which generates a pulling product flow. The resulting framework connects the production layout with planning and control principles and describes how this fits to the surrounding environment. It is thus a comprehensive recommendation of how a company can reduce complexity and achieve a higher throughput while still having a functional layout and without having to invest in more production resources.

It is concluded that there is potential in the suggested framework for companies with functional-oriented machining plants. However, research within this area must further be strengthen by empirical data to show that the suggested framework is viable in reality. This can be realized by academia also joining the collaborative effort mentioned above. By doing so academia gets access to the required empirical data to validate the *VG* concept, a formalized implementation approach can be developed and at the same time this more holistic approach to support functionally oriented production systems when striving to increase their throughput and performance can be spread.

The next step would be to compare how the recommended system would behave in a simulation model compared to the current system. Thereafter, the next logical step is to introduce a pilot study in a small company or in a part of a bigger factory as final validation before implementing in the recommended system in the whole factory. The framework only covers company internal aspects and more research is therefore needed to connect it to the external supply chain, both upstream and downstream. The framework ensures that all decisions within the company have the same objective function but the entire supply chain must be aligned for the whole system to operate optimally.

8 Conclusions

This chapter concludes the thesis by answering the three research questions presented in chapter 1.3, thus fulfilling the purpose of the study.

What is the actual capacity of the production system with the current set-up?

Capacity utilization is tightly connected to throughput time and to be able to determine how much of the theoretical capacity that can be utilized, a maximum acceptable lead time therefore first has to be set. The wide spread in utilization rates at which the lead time drastically increases, means that a lead time limit would generate different utilization bounds for each individual work center. This adds further to the complexity of this question. However, it can be concluded by the poor delivery performance and vast production delay, that the system with the current set-up operates at its absolute limit and that it must be changed in order to increase its throughput.

How extensive are the capacity losses in the system and what are the main causes?

The production system loses in general 28 % of the capacity that is left after deducting for productivity losses. Large variations and unnecessarily high *WIP* levels are the main causes behind this capacity loss and are in turn made up of a series of both physical and managerial constraints. Variations, caused by an inappropriate layout and a complex production system, have a large impact on the lead time at any given utilization rate and thus result in capacity loss. High *WIP* levels, an effect of poor planning and control procedures, results in capacity loss through unnecessary bottlenecks, long lead times and an unresponsive production system.

Given GKN's situation, what is the potential for increasing the throughput and how can it be achieved?

Since the lead time, even with minimal variations, will grow exponentially when approaching 90 % utilization and that demand for different resource categories is skewed, it is not feasible to access all of the 28 % capacity lost. Thus is the base for increasing the throughput a maximum 18 % increase of available capacity. This potential can be achieved by an overall transition towards focus on product flow and throughput time instead of capacity utilization and local efficiency. A *Virtual Group*-layout, a product oriented responsibility structure and an increased routing flexibility, will reduce variations and complexity in the production system and enable the use of simpler planning and control procedures. Introducing a *TOC* philosophy together with simple *S-DBR* planning and control principles will shorten lead times, reduce unnecessary *WIP* and create a production system that is transparent and predictable. Finally, combining the new production layout and production management principles with an aligned corporate culture and performance measurement strategy will tie the production system tightly to the actual market demand and let it fulfill its throughput potential.

9 References

Abdel-Maksoud, A., Dugdale, D., & Luther, R. (2005). Non-financial performance measurement in manufacturing companies. *The British Accounting Review*, *37*(3), 261-297.

Anderson, E. W., Fornell, C., & Rust, R. T. (1997). Customer satisfaction, productivity, and profitability: differences between goods and services. *Marketing science*, *16*(2), 129-145.

Balakrishnan, J., & Hung Cheng, C. (2005). Dynamic cellular manufacturing under multiperiod planning horizons. *Journal of manufacturing technology management*, *16*(5), 516-530.

Baumol, W., & Blinder, A. (2011). *Microeconomics: principles and policy*. Cengage Learning.

Berglund, R. (2014). *Kartläggning av logistiska nyckelfaktorer vid uppstart av H-sektorverkstad*. Högskolan Väst.

Bergman, B., & Klefsjö, B. (2010). *Quality from customer needs to customer satisfaction*. Studentlitteratur.

Blumberg, B., Cooper, D.R. and Schindler, P.S. (2011) *Business Research Methods*. Third European Edition. Maidenhead: McGraw-Hill Higher Education.

Bryman, A & Bell, E. (2011) *Business Research Methods*. 3rd edition. New York: Oxford University Press.

Burbidge, J. L. (1991). Production flow analysis for planning group technology. *Journal of Operations Management*, *10*(1), 5-27.

Chopra, S., & Meindl, P. (2007). *Supply chain management. Strategy, planning & operation* (pp. 265-275). Gabler.

Cox III, J., & Schleier, J. (2010). *Theory of constraints handbook*. McGraw Hill Professional.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Czarniawska, B. (2004) *Narratives in Social Science Research*. London: SAGE Publications Ltd, 2004.

Denscombe, M. (2009). *Forskningshandboken: för småskaliga forskningsprojekt inom samhällsvetenskaperna*. Studentlitteratur.

Denzin, N. K. (1978). Sociological Methods. New York: McGraw-Hill.

Drolet, J., Abdulnour, G., & Rheault, M. (1996). The cellular manufacturing evolution. *Computers & industrial engineering*, *31*(1), 139-142.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review*, *14*(4), 532-550.

Ellegård, K. (1992). *Reflektiv produktion: Industriell verksamhet i förändring*. Volvo media.

Forsberg, C. & Wengström, Y. (2003). *Att göra systematiska litteraturstudier: värdering, analys och presentation av omvårdnadsforskning*. Stockholm: Natur och kultur.

Ghalayini, A. M., Noble, J. S., & Crowe, T. J. (1997). An integrated dynamic performance measurement system for improving manufacturing competitiveness. *International Journal of production economics*, *48*(3), 207-225.

GKN Aerospace Engine Systems. (2014). *GKN Aerospace Engine Systems.* [handouts].

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

Goldratt, E.M. (1988), Computerized shop floor scheduling, *International Journal of Production Research*, Vol. 26 No. 3, pp. 443-55.

Goldratt, E.M. and Cox, J. (1984), *The Goal: An Ongoing Improvement Process*, Gower, Aldershot.

Goldratt, E.M. and Cox, J. (1986), *The Race*, North River Press, New York, NY.

Göran, W. (1996). Vetenskapsteori och forskningsmetodik. *Göran Wallén och Studentlitteratur*.

Holme, I. M., Solvang, B. K., & Nilsson, B. (1997). *Forskningsmetodik: om kvalitativa och kvantitativa metoder*. Studentlitteratur.

Hopp, W. J., & Spearman, M. L. (2011). *Factory physics*. Waveland Press.

Huang, H. H. (2002). Integrated production model in agile manufacturing systems. *The International Journal of Advanced Manufacturing Technology*, *20*(7), 515-525.

Hyer, N. L. (1982). Education: MRP/GT: A FRAMEWORK FOR PRODUCTION PLANNING AND CONTROL OF CELLULAR MANUFACTURING. *Decision sciences*, *13*(4), 681-701.

Kannan, V. R., & Ghosh, S. (1996). Cellular manufacturing using virtual cells. *International Journal of Operations & Production Management*, *16*(5), 99-112.

Krishnamurthy, R., & Yauch, C. A. (2007). Leagile manufacturing: a proposed corporate infrastructure. *International Journal of Operations & Production Management*, *27*(6), 588-604.

Lee, J. H., Chang, J. G., Tsai, C. H., & Li, R. K. (2010). Research on enhancement of TOC Simplified Drum-Buffer-Rope system using novel generic procedures. *Expert Systems with Applications*, *37*(5), 3747-3754.

Lockamy, A. and Cox, J.F. (1991), "Using V-A-T analysis for determining the priority and location of JIT manufacturing techniques", *International Journal of Production Research*, Vol. 29 No. 8, pp. 1661-72.

Mankiw, N. (2014). *Principles of macroeconomics*. Cengage Learning.

McLean C. R., Jones T. A., Hopp T. H. (1983). The virtual manufacturing cell. Information control problems in manufacturing technology 1982. In *Proceedings of the Fourth International Federation of Automatic Control/International*. Pergamon Press: New York (1983), pp. 207–215

Mula, J., Díaz-Madroñero, M., & Peidro, D. (2012). A conceptual model for integrating transport planning: MRP IV. In *Advances in Production Management Systems. Value Networks: Innovation, Technologies, and Management* (pp. 54-65). Springer Berlin Heidelberg.

Nomden, G., & van der Zee, D. J. (2008). Virtual cellular manufacturing: Configuring routing flexibility. *International Journal of Production Economics*, *112*(1), 439-451.

Nyberg, R. (2000). *Skriv vetenskapliga uppsatser och avhandlingar: Med stöd av IT och Internet*. Lund: Studentlitteratur.

Olhager, J. (2010). The role of the customer order decoupling point in production and supply chain management. *Computers in Industry*, *61*(9), 863-868.

Patel, R., & Davidson, B. (2003). Forskningsmetodikens grunder. Att planera, genomföra och rapportera en undersökning.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health services research*, *34*(5 Pt 2), 1189.

Plossl, G. W., & Orlicky, J. (2011). *Orlicky's material requirements planning*. McGraw-Hill Professional.

Prince, J., & Kay, J. M. (2003). Combining lean and agile characteristics: creation of virtual groups by enhanced production flow analysis. *International Journal of production economics*, *85*(3), 305-318.

Rahman, S. U. (1998). Theory of constraints: a review of the philosophy and its applications. *International Journal of Operations & Production Management*,*18*(4), 336-355.

Rheault, M., Drolet, J. R., & Abdulnour, G. (1995). Physically reconfigurable virtual cells: a dynamic model for a highly dynamic environment. *Computers & Industrial Engineering*, *29*(1), 221-225.

Schmenner, R. W., & Swink, M. L. (1998). On theory in operations management. *Journal of Operations Management*, *17*(1), 97-113.

Schönsleben, P. (2007). *Integral logistics management: Operations and supply chain management in comprehensive value-added networks*. CRC Press.

Schragenheim, E., Cox, J., & Ronen, B. (1994). Process flow industry—scheduling and control using theory of constraints. The International Journal of Production Research, *32*(8), 1867-1877.

Schragenheim, E., Dettmer, H. W., & Patterson, J. W. (2009). *Supply chain management at warp speed: Integrating the system from end to end*. CRC Press.

Schuh, G., Potente, T., & Fuchs, S. (2012). Shifting Bottlenecks in Production Control. In *Enabling Manufacturing Competitiveness and Economic Sustainability* (pp. 505-511). Springer Berlin Heidelberg.

Shen, M., & Chen, L. (2010, November). Production Bottleneck Shiftiness Study. In *System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2010 International Conference on* (Vol. 2, pp. 213-216). IEEE.

Shuttleworth, M. (2008). Deductive Reasoning. *Exploreable.* https://explorable.com/deductive-reasoning (18 April 2015).

Slack, N., Chambers, S., & Johnston, R. (2010). *Operations management*. Pearson Education.

Smith, D & Smith, C. (2013). *Demand Driven Performance: Using Smart Metrics.* McGraw-Hill Professional; 1 Edition

Steele, D. C., Philipoom, P. R., Malhotra, M. K., & Fry, T. D. (2005). Comparisons between drum–buffer–rope and material requirements planning: a case study. *International Journal of Production Research*, *43*(15), 3181-3208.

Thomas, G. (2010). *How to do your case study: A guide for students and researchers*. Sage.

Vinodh, S., Gautham, S. G., & Ramiya R, A. (2011). Implementing lean sigma framework in an Indian automotive valves manufacturing organisation: a case study. *Production Planning & Control*, *22*(7), 708-722.

Appendix I – Enhanced Production Flow Analysis

Enhanced Production Flow Analysis, *EPFA*, is a methodology suggested by Prince and Kay (2003) to identify and create *VGs* within the current production system (Vinodh et al., 2011). *EPFA* is a further development and extension of Production Flow Analysis, *PFA*, presented by Burbidge (1991). *PFA* is according to Burbidge (1991) made up of the five sub-techniques, Company Flow Analysis, *CFA*, Factory Flow Analysis, *FFA*, Group Analysis, *GA*, Line Analysis, *LA* and Tooling Analysis ,*TA*. *FFA* and *GA* are the primary techniques used in order to identify groups. As in *PFA* is the central aspect in *EPFA* to create groups by assigning parts and products to machines but the methods differs on multiple points (Prince & Kay, 2003; Burbidge, 1991). The differences are:

- In *PFA*, a split of a module occurs when two completely different sets of machines are in the same module (Burbidge, 1991). This means that after the split, one of the two new modules must contain a duplicate or an alternative machine. However, Prince and Kay (2003, p. 313) mean that *"a module based around a machine with a classification of Special* [(S) therefore] *cannot occur"*. Using *EPFA*, this case therefore suggest the creation of a shared machine resource, which is then needed by two or more modules.

- In *PFA*, Burbidge (1991) creates special groups that incorporate exceptional machines who do not fit into the improved and simplified flow. These machines are usually non-attractive since they cause back-flow and cross-flow between groups. Prince and Kay (2003) extend the concept of these special groups towards more process-oriented *VGs*, which then contain machines that are not assigned to a certain *VG*. This occurs when a machine is not associated with a certain product or there are not enough machines available to form modules. Process-oriented *VGs* should according to Prince and Kay (2003) be eluded as much as possible since they go against the main purpose of creating *VGs*, that is changing from a process-focused management policy towards a product-focused one.

- Lastly, Price and Kay (2003) mean that the *SICGE* classification method presented by Burbidge (1991) does not provide enough flexibility when creating *VGs*, and therefore present an improved module algorithm making it possible to create focused sub-groups within a *VG*.

Underneath, the step-by-step methodology for *EPFA* presented by Prince and Kay (2003), is explained. Figure 33 illustrates the eleven different steps while complementing remarks for each step are discussed beneath.

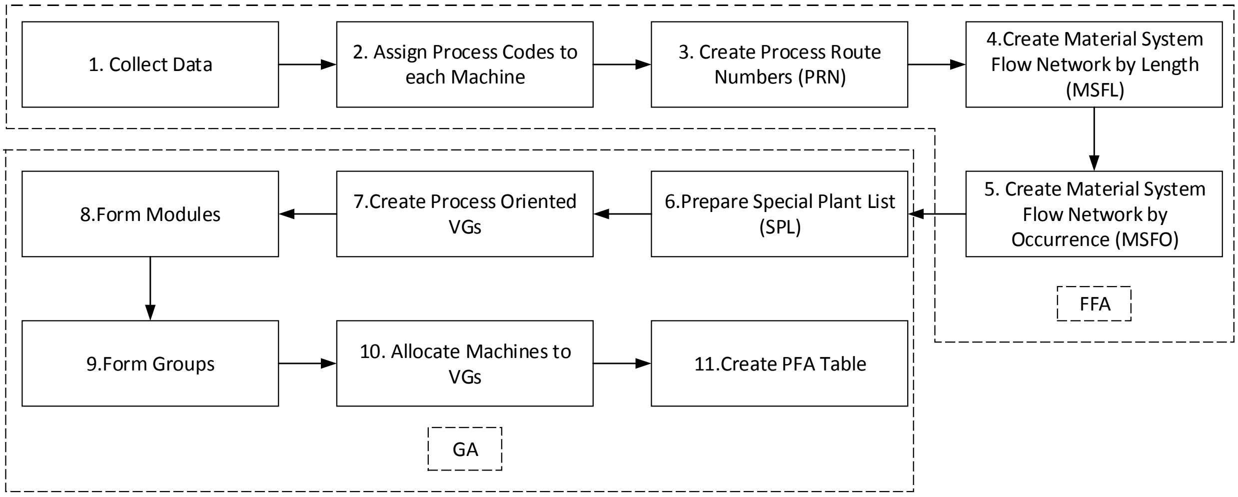

Figure 33: EPFA process (Prince & Kay, 2003).

[1] The data that is necessary to collect and format for creating *VGs* is not exhaustive (Prince & Kay, 2003) and should be available in most manufacturing firms. The data required includes machines or work centers, products, operations, routings, and descriptions of them all.

[2] This step introduces *FFA*, which task is to divide the factory into major groups and aims at creating a simple unidirectional flow system (Burbidge, 1991). Here, process codes are assigned to each machine while duplicate machines are given the same process code (Prince & Kay, 2003). This applies even if machines have some small differences in which operations and tasks they can execute.

[3] *PRNs* are created for each product by joining the process code for each process with the number it occurs in the routing (Prince & Kay, 2003; Burbidge, 1991).

[4] The goal of this step is to find the busiest processes and machines, and the largest product flows between machines (Prince & Kay, 2003). The *PRNs* are ranked in an ascending order of quantity in order to create a Material System Flow Network by length, *MSFL* (Prince & Kay, 2003). The *MSFL* can be shown as a From/To table or a visual network of nodes and arcs (Burbidge, 1991). The *MSFL* shows the total amount of products that flows between each pair of process codes in both directions (Burbidge, 1991; Prince & Kay, 2003).

[5] In this step is a Material System Flow Network by Occurrence, *MSFO*, created. A *MSFO* is identical to a *MSFL* except from that is counts the number of times each *PRN* occurs instead of the number of products that flows between each pair of processes (Prince & Kay, 2003; Burbidge, 1991). The purpose with the *MSFO* is to complement the *MSFL* and assure that focus lies on processes that risk becoming bottlenecks due to low quantities and a higher number of set-ups (Prince & Kay, 2003).

[6] The next step is to prepare the special plant list, *SPL*, which is used to identify modules containing products and machines that can be combined when forming *VGs* (Prince & Kay, 2003). This is a part of the sub-technique *GA* following *FFA* in *PFA*. The process of creating a *SPL* starts with classifying the machines into the *SICGE* classification explained in Burbidge (1991). The five different categories are as follow from Burbidge (1991):

- S - In the *Special category* there exist only one machine of each type, which implicates that it is difficult, if not impossible, to transfer its work to another machine type.
- I - The *Intermediate category* is the defined as the Special category, but there exists duplicates of that machine type.
- C - Machines that there exist several duplicates of and where operations easily can be transferred to other machine types are classified in the *Common category*.
- G - Machines in the *General category* are used for a high portion of all products or for many different types of products. These machines are often not able to include in any groups, since the operations they perform are generic, e.g. x-ray machines.
- E - The *Equipment category* consists of items that are used to support manual operations such as e.g. benches and manual tools.

When the plant list is sequenced according to *SICGE* and is ascending by *"the number of different parts with operations on a machine type"*, it can be used to select the key machines needed to form modules (Burbidge, 1991, p. 19).

[7] Process-oriented *VGs*, formerly described in 2.4.2, contains the machines that are not able to assign to a specific *VG* or due to too few number of machines (Prince & Kay, 2003).

[8] Modules need to be created since most companies have a large number of products or/and machines (Burbidge, 1991). Modules are created by using the enhanced module forming algorithm illustrated in Figure 34.

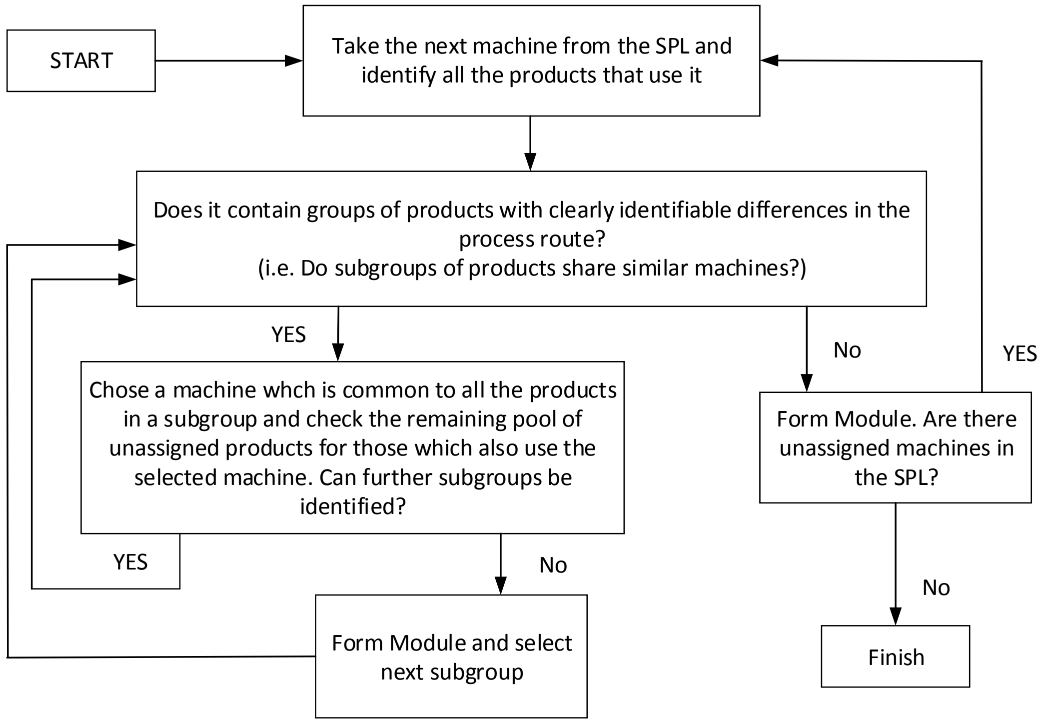

Figure 34: Enhanced module forming algorithm (Prince and Kay, 2003).

[9] In this step, the modules are to be combined to form the *VGs*. One should strive after combining the modules so that it is possible to form as many groups as possible with as few exceptions as possible (Prince & Kay, 2003). Burbidge (1991) put this as that the groups should have as little back- or crossflow as possible. An exception is according to Prince and Kay (2003, p. 314) defined as *"a product in one group that requires a machine that is part of another group except, where a group requires a shared machine".* Prince and Kay (2003) suggest that exceptions can be handled in two ways, by either combining modules or groups or modifying the functionality of a machine to shift its classification.

[10] After the group formation, this step verifies that no machine has been assigned multiple times and that the most suitable machine is delegate to each *VG* (Prince & Kay, 2003).

[11] In the final step of *EPFA* is the *PFA* table created to visualize the machines and products in each *VG* (Prince & Kay, 2003).