CHALMERS
UNIVERSITY OF TECHNOLOGY

Analyzing Bottlenecks and Capacity Scalability in Engine Manufacturing

Master's thesis in Production Engineering

GUSTAV HÄSSEL
JACOB REHN

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF SUPPLY AND OPERATIONS MANAGEMENT

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Analyzing Bottlenecks and Capacity Scalability in Engine Manufacturing

GUSTAV HÄSSEL
JACOB REHN

Department of Technology Management and Economics
Division of Supply and Operations Management
Chalmers University of Technology
Gothenburg, Sweden 2026

Analyzing Bottlenecks and Capacity scalability in Engine Manufacturing
GUSTAV HÄSSEL
JACOB REHN

Department of technology management and economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
Picture of the D6 diesel engine produced in Vara.

Gothenburg, Sweden 2026

Analyzing Bottlenecks and Capacity Scalability in Engine Manufacturing

GUSTAV HÄSSEL
JACOB REHN

Department of technology management and economics
Chalmers University of Technology

Abstract

This master thesis investigates the production capacity of the D4/D6 engine lines at Volvo Penta's manufacturing facility in Vara. The study leverages the combined strengths of Value Stream Mapping and the Theory of Constraints to identify systemic bottlenecks and evaluate optimal staffing configurations under varying production volumes. The research focuses on analyzing a gradual increase in customer demand to determine the scalability of the current assembly process. Through the development of a capacity model, the study identifies critical constraints that emerge as production scales. A key finding of the analysis is that the current production flow, in its existing configuration, possesses a maximum capacity of 31 engines per day shift, a significant increase from the current demand of 18 engines, without the immediate necessity for operational rebalancing. The analysis specifically identifies the point where the Longest Operation Time (LOT) creates a physical bottleneck. This shows that beyond a certain volume, adding more staff will no longer increase output without changes to the station layout. This serves as a decision support tool for Volvo Penta, highlighting when infrastructure investments become necessary to meet forecasted demand.

Keywords: Value stream mapping, Theory of constraints, Production scalability, Bottleneck propagation, Walking workers assembly line, Longest operation time, Line balancing

Acknowledgements

This master's thesis marks the conclusion of the Master's programme in Production Engineering at Chalmers University of Technology, carried out in collaboration with Volvo Penta. We would first like to thank our supervisors at Volvo Penta, Tom Jörgensen Björk and Erik Hemberg, for their guidance and for welcoming us to the Vara plant. Your support and practical insights have been invaluable to us throughout the project. We also want to thank our supervisor at Chalmers, Nils Tylén, for his consistent support and structured feedback, which helped us keep the study on the right track. Finally, a special thanks to the employees at Volvo Penta who took the time to contribute to this project. Your interest and positive energy made our time at the plant both motivating and enjoyable.

Gustav Hässel, Jacob Rehn, Gothenburg, May 2026

List of Acronyms

APP	Aggregate Production Planning
AT	Available Time
C/T	Cycle Time
ERP	Enterprise Resource Planning system
FIFO	First In First Out
L/T	Lead Time
LOT	Longest Operation Time
NVA	Non-Value Added time
Op	Operator
SAM	Sequence Based Activity and Time
T/T	Takt Time
TOC	Theory of Constraints
TPS	Toyota Production System
TPT	Total Paid Time
TW	Total Work Content
UT	Utilization
VA	Value Added time
VPS	Volvo Production System
VSM	Value Stream Mapping
WIP	Work in Progress
WWAL	Walking workers assembly line
FWAL	Fixed workers assembly line

Nomenclature

Below is the nomenclature of parameters, variables, and Greek symbols that have been used throughout this thesis.

Parameters

t_i	**Processing time** Individual processing time for operation i.
$N_{assemblers}$	**Number of assemblers** Total count of assemblers assigned to the production line.
$N_{operators}$	**Number of operators** Total number of operators active in the specific flow.
$L/T_{Inventory}$	**Inventory Lead Time** The non-value-added time products spend waiting in inventory.
$L/T_{Process}$	**Process Lead Time** The value-added time for products during actual processing.
L/T_{Total}	**Total Lead Time** The cumulative time for an engine to pass through the entire system.
$Utilization$	**Utilization** The used time compared to the total time.

Variables

$CycleTime$	**Cycle Time** The potential production pace per product based on work content and staffing.
LOT	**Longest Operation Time** The slowest single operation in a process; defines the structural bottleneck.
$TaktTime$	**Takt Time** The required production pace to meet customer demand based on available time.

Greek Symbols / Other

Σt_i	**Total Work Content** The sum of all individual processing times for a complete production cycle.

Contents

List of Figures

List of Tables

1

Introduction

Manufacturing companies operating in high-mix, variable demand environments face increasing challenges. Previous research highlights several challenges in high-mix, low-volume production environments, including variability in processing times, fluctuating demand, and low production volumes with high product variety (Gan et al., 2023). While demand may fluctuate, production systems are often designed around fixed resources and limited visibility of true capacity constraints. As a result companies risk having underutilization during low demand and bottlenecks emerging during high demand (Liker, 2021; Sun, 2024).

These challenges are commonly addressed through aggregate production planning, which focuses on balancing demand and capacity over a medium-term planning horizon (Nam & Logendran, 1992). Aggregate planning strategies typically involve decisions related to production rates, workforce size, and inventory levels in order to manage demand variability. Common approaches include maintaining a stable production rate and workforce (level strategy), adjusting capacity to follow demand fluctuations (chase strategy), or applying hybrid solutions that combine both approaches (Jamalnia et al., 2019). In practice however, the effectiveness of aggregate production planning depends on how accurately the production systems capacity and constraints are understood. Planning decisions are frequently based on assumed capacity levels, while true limiting factors remain partially hidden due to variation and inefficiencies in the production flow (Goldratt, 1990). When the underlying production flow is unbalanced, strategic decisions related to staffing and capacity adjustment risk being misaligned with the system's actual capabilities, leading to suboptimal performance (Mortada & Soulhi, 2023). This creates a gap between planned capacity and realized system performance.

In manufacturing systems, a key challenge is to accurately determine the actual production capacity and understand the factors that limit system performance. Capacity is often constrained by inefficiencies, imbalances, and variability in production processes, which makes it difficult to assess how the system will perform under changing demand conditions. In the context of lean production such inefficiencies are often associated with waste, defined as any activity that consumes resources but does not add any value from the customers perspective (Liker & Convis, 2011). However, waste and performance losses are often embedded in daily processes, making them difficult to identify without systematic analysis (Rother & Shook, 1999). To accurately assess production capacity, a clear understanding of how material and information flows actually operate is required, rather than how they are intended to function according to standards and work instructions (Womack & Jones, 2003). Visualization of the production flow is therefore a fundamental prerequisite, as it enables the identification of operational losses, bottlenecks, and capacity constraints (Librelato et al., 2014). According to the Theory of Constraints, every

system is limited by at least one system-critical bottleneck, which governs overall capacity and throughput (Rahman, 1998). Consequently, improvements that do not address the constraining resource are unlikely to significantly increase system capacity. Therefore, effective control and scalability of a production system require not only the identification of losses and bottlenecks, but also a quantification of the system's actual capacity.

While visualizing production is essential for identifying losses, evaluating the actual capacity boundaries requires an understanding of the system's constraints and flow dynamics. In modern manufacturing, hidden bottlenecks and system inefficiencies often dictate the maximum capacity of the entire system. Without a clear quantification of these constraints, management lacks the visibility needed to scale production and adjust staffing levels effectively in response to fluctuating demand (Mortada & Soulhi, 2023). This creates a critical gap between intended aggregate planning strategies and the operational reality on the shop floor. Therefore, identifying the true system bottlenecks and evaluating different staffing configurations becomes a strategic link between production demand and resource allocation. Analytical methods, such as line balancing, can be utilized in this context to evaluate the theoretical maximum capacity, and determine how task distribution impacts scalability. Together with flow visualization, this provides a theoretical foundation that allows management to establish robust staffing structures and make informed decisions on how to effectively scale and control production (Slack et al., 2019).

Overall, the literature suggests that a clear understanding of how production flows and system constraints determine actual capacity is essential for analyzing production system performance. This highlights the need to quantify actual production capacity and analyze how bottlenecks and staffing configuration influence system performance. A staffing configuration serves as a framework for how the workforce should be dimensioned based on varying demand requirements.

1.1 Volvo Penta

Volvo Penta AB is a subsidiary of Volvo group that focuses on powertrain for industry and marine applications. The company has two different factories located in Vara Sweden and Lexington USA. The subject of this thesis is the production line of the four- liter(D4) and the six- liter(D6) diesel engines at the factory in Vara. These engines are used for marine applications as inboard engines for medium size boats and has the highest production volumes of all engines in the Vara factory. The production facility in Vara for the D4/D6 line consists of a machining line, an assembly line, engine testing, a engine painting line and material handling.

Volvo Penta as part of Volvo Group has a strong lean initiative working actively with established lean methods to improve their processes. A result of this commitment is Volvo Production System (VPS) which is Volvo's own take on Toyota Production System (TPS). To evaluate how well different divisions of the Volvo Group have performed in their integration of VPS they are assessed based on a structured award system that ranges from VPS Platinum to VPS Bronze. Currently Volvo Penta in Vara has the VPS Silver award and is aiming to achieve VPS Gold.

Despite Volvo Penta's efforts in lean manufacturing, the operation faces challenges related to demand variability, changing production volumes, and an increasing need for scalability. At present, the production is characterized by relatively low utilization, however future volume increases are expected to place higher demands on efficient resource utilization and the ability to scale production capacity. Currently Volvo Penta is using temporary agency staffing a common strategy in chase oriented capacity adjustment to solve the issue of fluctuating demand in their assembly line. However, this approach relies on assumed capacity levels rather than a detailed understanding of the underlying production flow, including bottlenecks and workload imbalances. As a result, there is a risk that staffing adjustments are not aligned with the system's actual capacity and constraints.

In this context, there is a need to develop a clear understanding of the actual production flow of the D4/D6 production line. Of particular importance is identifying the system's bottlenecks, determining the maximum production capacity, and how staffing levels affect workforce utilization and bottleneck development. Such an analysis constitutes a necessary foundation for future decisions regarding resource allocation and continuous improvement initiatives, thereby motivating the focus and problem formulation of this thesis.

1.2 Problem description

Currently, the Volvo Penta plant in Vara faces uncertainty regarding the true production capacity of the D4/D6 production line, the location of system-critical bottlenecks, and how staffing levels influence workforce utilization. While the plant is currently experiencing low production volumes and excess capacity, future increases in demand require a clear understanding of how the production flow behaves under different conditions. A key challenge is that production planning and staffing decisions are often based on assumed capacity levels, rather than on a detailed understanding of the actual production flow and its constraints. This limits the ability to identify bottlenecks, evaluate capacity limits, and understand how changes in staffing and line balancing affect system performance.

Without a clear understanding of current production constraints, managers may face difficulties in managing human resources effectively. Inaccurate capacity assumptions may lead to unstable workforce planning, unnecessary temporary staffing, excessive layoffs during demand downturns, or sustained work overload during peak periods. These consequences show that the problem is not only technical, but also has societal and ethical relevance. A more transparent understanding of production capacity can therefore support more informed decisions that balance operational efficiency with workforce stability and sustainable resource utilization. To support such decision-making and enable scalable production, there is a need for an analysis of the production flow that captures the interaction between bottlenecks, capacity, and staffing configurations.

In addition to increasing production volumes, future changes in the product mix include the introduction of an additional engine model that will be assembled within the existing production system but will not pass through the machining processes. This implies that the load on the assembly processes is expected to increase disproportionately, placing greater emphasis on assembly capacity, bottleneck behavior, and staffing efficiency.

1.3 Purpose

The purpose of this study is to increase the understanding of how the D4/D6 production flow can meet higher demand by analyzing how capacity limitations, system-critical bottlenecks, and staffing configurations influence production scalability.

1.4 Research questions

- What are the system-critical bottlenecks in the D4/D6 production flow under current and increased demand conditions?
- What is the actual and maximum production capacity of the D4/D6 production flow based on identified constraints, and how it is affected by line balancing?
- How do different staffing configurations affect workforce utilization and bottleneck development in the production flow?

1.5 Limitations

Limitations with this project is the coverage of only D4- and D6-flow and not the other flows in the plant. A new machining line was under implementation during the study but was not yet operational at the time of data collection. Consequently, the available machining data was based on estimated performance values rather than observed production data. For this reason, the analysis primarily focuses on the assembly-related processes, while the machining results should be interpreted as indicative and exploratory rather than definitive. The capacity analysis in this study is primarily based on deterministic data through SAM analysises. Historical data regarding stochastic disturbances, such as machine breakdowns and variablity in operator performance, were not available for this analysis. This delimitation was chosen to prioritize a systemic evaluation of structural bottlenecks and capacity scalability. By focusing on "steady-state" capacity, the study isolates fixed architectural constraints from short-term operational fluctuations. Consequently, the results represent an idealized production flow. While these findings provide a baseline for capacity planning, it should be noted that in a real-world environment, unscheduled downtime and operator variance will impact net throughput. These results therefore serve as an upper-bound capacity limit, identifying the critical nodes where investments in robustness will yield the highest return.

2

Theoretical framework

This chapter presents the theoretical framework that underpins the analysis conducted in this study. The theories and methods discussed provided the conceptual foundation for understanding production flow, capacity limitations, and resource utilization in manufacturing systems. Particular emphasis is placed on Lean, Value stream mapping and Production planing, as these concepts are central to identifying system-critical bottlenecks, quantifying production capacity, and evaluating theoretical staffing structures in relation to production scalability.

2.1 Lean management

A production flow is defined by a sequence of activities designed to transform raw materials into a product that provides value to the customer. To achieve operational efficiency, it is essential to understand the dynamics of this flow, specifically distinguishing between value-adding and non-value-adding activities to maximize business profitability. One of the most prominent philosophies for achieving this optimization is Lean Management, which focuses on the systematic elimination of waste and utilizing resources as efficient as possible (Liker, 2021).

2.1.1 Value

Value is an important topic in this study. In order to make a profit, a company must create value for its customer but what exactly is value? In the context of lean production, value is fundamentally from the customers perspective. According to Womack and Jones (2003), value is the internal capability of a product or service to meet specific customer needs at a given time and price. Any activities that consume resources but does not contribute to meeting these needs is classified as waste. Muda is the Japanese word for waste and according to Ohno (1988), there are seven types of Muda that commonly occur in manufacturing processes: overproduction, waiting, transport, inappropriate processing, unnecessary inventory, unnecessary motion and defects. After Taiichi Ohno defined the seven Muda one additional has been added, underutilized peoples abilities. Which means that limiting employees authority and responsibility will generate a waste.

2.1.2 Muda

Within the framework of waste, it is important to be aware that all wastes are not unnecessary. There are two types of waste, necessary and unnecessary, necessary waste refers to activities that do not directly add any value to the product but are essential for the product to be delivered or manufactured with the correct quality. This can be transportation

to the customer or quality control in the production line. In contrast, unnecessary waste consists of activities that add no value and can be eliminated without negative consequences to the production. The main goal of lean production is to eliminate unnecessary waste while minimizing necessary waste through continues improvements.

To systematically identify these inefficiencies, Liker and Convis (2011) provides a framework consisting of eight distinct types of waste.

Overproduction

According to Liker and Convis (2011), overproduction is considered the most critical form of waste. It entails producing items for which there is no immediate customer demand, resulting in significant capital being tied up in inventory. More importantly, overproduction serves to obscure underlying inefficiencies and other types of waste within the system. This leads to an unnecessary production flow that is disconnected from actual market demand and is, therefore, entirely non-value adding.

Waiting

Waiting occurs when an operator remains idle while a machine is running, or when a process is stalled due to a lack of materials. Such delays are often a consequence of line imbalances or inadequate planning, which can serve as critical indicators for identifying operational bottlenecks (Jamalnia et al., 2019).

Unnecessary transport

Transporting materials, parts, or finished articles between storage facilities and workstations is a non-value adding activity and is, therefore, classified as waste. Each instance of material handling increases the risk of product damage and consumes time that could otherwise be allocated to value-adding processes. Consequently, Liker and Convis (2011) argues that the layout of a production flow should be as compact as possible to minimize the necessary transport area and streamline the movement of goods.

Over processing

Over processing occurs when more work is performed on a product than is actually required by the customer. Examples include following unnecessarily tight tolerances or implementing redundant administrative steps that do not enhance the final output. According to Liker and Convis (2011), this represents a waste of quality, an investment in resources and precision for which the customer is unwilling to pay, as it does not contribute to the product's perceived value.

Excess inventory

Excess inventory, whether in the form of raw materials, work-in-process (WIP), or finished goods, often serves to obscure underlying production issues such as line imbalances or unreliable supplier deliveries. Furthermore, inventory represents tied-up capital that could otherwise be allocated to value-adding activities. According to Liker and Convis (2011), inventory acts as a cushion that allows a company to avoid addressing the root causes

of its inefficiencies, effectively hiding problems that would otherwise demand immediate attention.

Unnecessary movement

In contrast to material transport, the waste of unnecessary motion specifically concerns the physical movements of operators. It represents a loss of both energy and time whenever an operator is required to walk to retrieve tools, reach for parts, or search for information. According to Liker and Convis (2011), an ergonomic and well-structured workstation often achieved through workplace organization methods is essential for minimizing these non-value adding movements and enhancing overall operational efficiency.

Defects

The production of defective parts necessitates either rework or scrapping, both of which are extremely costly as they consume resources, such as materials, machine time, and human effort without resulting in a scalable product. Liker and Convis (2011) argues that it is far more beneficial to halt the production line and address the problem immediately, rather than allowing a defective part to continue through the process. This proactive approach prevents the compounding of waste and ensures that quality is built into the process from the start.

Underutilized abilities

The eighth waste, under utilization of employee potential, was emphasized by Liker and Convis (2011) as a critical oversight in many organizations. They argued that a company loses significant time, creative ideas, and improvement opportunities when operators are excluded from the problem-solving process. By failing to engage with those who perform the work daily, a firm forfeits valuable intellectual capital and specialized knowledge.

2.1.3 Theory Of Constraints

Theory of constraints (TOC) is a fundamental philosophy in production engineering that will work in the background of all other methods in this thesis. The main concept of TOC is that every system must have at least one constraint often referred to as the ”bottleneck”, that determines the throughput of the system, if not true then that system would have an unlimited throughput (Rahman, 1998). The TOC philosophy also imposes a method on how to work with constraints to improve a production systems throughput in five steps (Rahman, 1998). First the constraints needs to be identified, the TOC philosophy does not impose any specific tool or method to do this however Librelato et al. (2014) purposes using VSM for this purpose a method also intended to be used in this thesis. The next step is to decide how to exploit the constraint meaning trying to make the constraint as effective as possible without any further investments. The third step is to subordinate everything else to help exploit the constraint. This step further solidifies on of the key principles of TOC that an investment in a non-constraint will not effect the throughput as the throughput is till controlled by the constraint. Therefore non-constraints must be adjusted to support the maximum effectiveness of the constraint. The fourth step is to elevate the systems constraints meaning to invest in increasing the throughput of the constraint as this will increase the throughput for the whole system. The fifth and

last step is to overcome inertia and is what makes TOC a continuous process, When one constraint disappear does not mean that the system is free from constraints it only indicates that now something else is the constraint and the process starts over.(Rahman, 1998).

2.1.4 Yamazumi Charts

To evaluate theoretical staffing structures and identify bottlenecks, this study utilizes Yamazumi charts which is a workload visualization tool that is commonly used for production leveling (Cannas et al., 2018). A Yamazumi chart, as shown in figure 2.1, is a bar chart where the height of each bar represents the total work content assigned to each work station within one production cycle. This allows for a direct comparison between workstations in order to reveal workload imbalances throughout the production line that may affect the production flow.
The work content is collected in the form of processing times for each workstation within the production flow. Takt time is included in the Yamazumi chart as a reference line, allowing the workload of each workstation to be evaluated against the required production pace. If the total workload at a workstation exceeds the takt time, it indicates that the station is overloaded and may act as a bottleneck, while stations with significantly lower workloads indicate underutilization. This visualization enables the identification of uneven workload distribution, which is a key source of inefficiency in production systems (Pakdil & Leonard, 2017).

Yamazumi charts provide a structured basis for redistributing tasks between operators in order to achieve a more balanced workflow. By reassigning tasks, the workload can be aligned more closely with the takt time, thereby improving flow efficiency and reducing idle time. This approach is closely related to the concept of line balancing and supports the principle of production leveling (Heijunka) within lean manufacturing (Ohno, 1988). For this reason, Yamazumi charts are a valuable tool when performing line balancing and evaluating alternative staffing configurations under different production conditions.

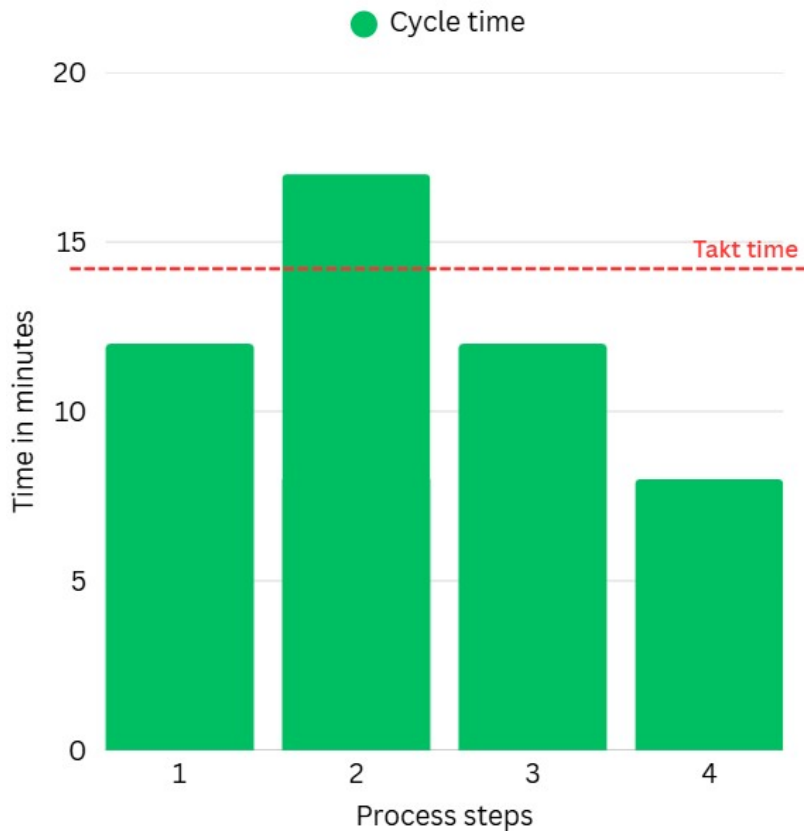

Figure 2.1: Example of a Yamazumi chart illustrating workload distribution across workstations (figure inspired by Kanban Zone, 2020)

2.2 Value Stream Mapping

Value Stream Mapping (VSM) was popularized in the 1990s , as a method for visualizing material and information flows throughout the value stream, from supplier to customer (Rother & Shook, 1999). The initial goal of VSM is to create a current-state-map of the process. This focuses on analyzing what Holweg et al. (2018) describes as the ”as-is” process. While the methodology traditionally includes both a current and a future state, it is also utilized as a standalone diagnostic tool to establish an empirical ”as-is” baseline (Librelato et al., 2014). Holweg et al. (2018) further argues that mapping the ”as-is” process is often perceived as prosaic when an understanding of how the process should operate already exist. However, the purpose of the ”as-is” map is not to capture how the process should operate but rather how it actually operates. This includes workarounds and other deviations from the standard that effects the process. Therefore, Holweg et al. (2018) emphasizes the importance of ”going to Gemba”, meaning going to the actual workplace when collecting data, in order to capture these workarounds, as they do not appear in formal documentation such as work-instructions. As stated by Librelato et al. (2014) VSM is a useful method for identifying and locating losses in a process which is at most interest for this thesis as one of the goals is to find the system-critical bottlenecks for the D4/D6 flow. Librelato et al. (2014) also criticizes VSM’s for its lack of a structured method for prioritizing for the improvement work of the future state map, Librelato et

al. (2014) solves this problem by combining the use VSM and TOC by mapping and identifying wastes using VSM and prioritizing and finding the root cause of these wastes using TOC. A similar workflow will be conducted in this thesis where VSM will be used to visualize the flow and identify wastes and TOC will be used to analyze which of the wastes is the system critical bottleneck. The bottleneck together with the time table developed in the VSM will be used to determine the maximum capacity of the production flow.

2.2.1 Construction of the VSM

This section describes the construction of the VSM used to visualize and analyze the production flow of the studied system following the methodology proposed by Rother and Shook (1999). The purpose of the VSM is to provide a structured overview of both material and information flows, enabling the identification of bottlenecks, capacity limitations, and improvement opportunities within the production system.

2.2.2 The customer

Rother and Shook (1999) states that the mapping process always starts at the customer. Customer demand data is to be collected for a yearly, monthly and daily basis and drawn in a data box connected to the customer. The daily demand is then used to determine the customer takt time using the available production time in the production. Takt time is at what pace the production needs to run to meet customer demand and is calculated using equation 2.2. The available production time is defined as the total time during which production resources are available to perform value-adding work, after subtracting planned down time such as breaks see equation 2.1, meetings and maintenance from the total scheduled working time (Rother & Shook, 1999).

$$\text{Available Production Time} = \text{Total Available Time} - \text{Planned Downtime} \tag{2.1}$$

$$\text{Takt Time} = \frac{\text{Available Production Time}}{\text{Customer Demand}} \tag{2.2}$$

2.2.3 Process steps and data

The next step is to identify and draw the process steps. In conjunction with this step, the appropriate level of detail for the VSM can be determined.Once the process is established process data is added to each process box. The standard process data in a VSM is cycle time, changeover time and up time (Rother & Shook, 1999).

Cycle time is defined as the time it takes from one unit being produced by the process until the next can be produced. In a fully staffed assembly line (one operator per station), the process cycle time is equal to the maximum takt or work time of the bottleneck station. In contrast, when running an understaffed or flexible line, where there are more stations than operators the cycle time calculation must adjust for operator cross-handling and walking time. One of the case studies features a WWAL where operators move along with the engine instead of staying at a single station. This approach enables the line to function with fewer operators than physical stations. To calculate the cycle time in a WWAL the

total work content i.e. the time it takes for one operator to complete all assembly stations in the line, is divided by the number of operators using equation 2.4. However this is only true if the cycle time is longer than the longest station work time otherwise the longest station will instead become the total cycle time in accordance to TOC (Rahman, 1998). For this reason the longest operation time (LOT) needs to be collected in order to verify if the calculated cycle time is valid. LOT is calculated using equation 2.3.

$$\text{LOT} = max(t_i) \tag{2.3}$$

$$\text{Cycle time} = max(\text{LOT}, \frac{\sum t_i}{N_{operators}}) \tag{2.4}$$

where:

- t_i = processing time for operation i
- $N_{operators}$ = number of operators in the line

2.2.4 Inventory and Buffer Levels

In addition to the process data, inventory and buffer levels between processes are identified and included in the VSM. According to the methodology described by Rother and Shook (1999), inventory levels are an important element of the VSM as they illustrate where material accumulates within the production flow and indicate potential imbalances between processes. These inventory points are represented in the VSM using the standard VSM inventory symbols and were quantified in number of units where possible. The purpose of including inventory and buffer levels in the map is to visualize where products are waiting between the processes and to provide insight into potential bottlenecks and flow interruptions. The inventory and buffer data is also used later in the process to calculate lead time. This information supports the analysis of production flow and helps identify areas where improvements in flow efficiency may be possible.

2.2.5 Material flow

After defining the process steps, process data, and inventory levels, the material flow between the processes is mapped. The material flow describes how products move through the production system from one process to the next. The first step in this process is to determine how raw material is supplied to the beginning of the system. Specifically, the analysis focuses on how often material is delivered, batch sizes of the deliveries, and the corresponding inventory levels. The next step is to investigate how material flows between processes within the system in order to determine whether the production flow follows a push or pull principle. A indicator of a pull flow is the presence of a Kanban system where production upstream is triggered by actual consumption downstream (Rother & Shook, 1999). This analysis provides insight into how production is controlled within the system and supports the understanding of material flow dynamics in the value stream.

2.2.6 Information flow

The next step in the VSM process is to map the flow of information within the system. The purpose is to gain insight into how information is currently communicated between processes and to identify potential points where information may be delayed or lost. The

collected information is then incorporated into the VSM to visualize how production planning and control signals propagate through the system.

2.2.7 Time line

The last step of creating the current state VSM is to create the time line used to calculate the total lead time for the products. According to the VSM methodology described by Rother and Shook (1999), the timeline distinguishes between value-adding processing time and non-value-adding waiting time within the system. This is visualized within the timeline by a zigzag formation with value-adding time being represented by the valleys in the time line and non-value-adding waiting time being represented by the peaks. The non-value-added waiting time is calculated by multiplying the inventory with the takt time using equation 2.5. By combining these elements the total lead time can be calculated and visualized in the VSM timeline. This representation provides an overview of how much time products spend being processed compared to the time they spend waiting within the system. To further enhance this representation the ratio between value adding and total lead time is calculated by dividing the sum of the processing time with the total lead time using equation 2.6. The timeline therefore supports the identification of inefficiencies in the production flow and highlights potential areas for improvement.

$$L/T_{inventory} = \text{Inventory} * \text{Takt time} \tag{2.5}$$

$$\text{Value-added ratio} = \frac{\sum L/T_{process}}{L/T_{total}} \tag{2.6}$$

where:

- $L/T_{Inventory}$ = The non-value-added lead time in the inventory
- $L/T_{Process}$ = The value-added lead time for products in the process
- L/T_{Total} = The total lead time of the system

2.3 Aggregate production planning strategies

Aggregate production planning (APP) is a medium-term planning approach that aims to balance demand and capacity over a defined planning horizon, often spanning 3 to 18 months. As suggested by the name planning decisions are made on an aggregated basis, focusing on overall production quantities, hiring and lay off rates, work force and inventory levels, back ordering and subcontracting volumes rather than individual products or customer orders (Jamalnia et al., 2019). The primary objective of APP is to determine how available resources should be utilized to meet forecasted demand at a minimum-cost, while considering constraints related to capacity, labor and inventory. Common decision variables used in APP is hiring and layoff rates, over and under time, inventory and backlog levels (Jamalnia et al., 2019).

Two fundamental strategies are often discussed in literature related to APP, Chase strategy and Level strategy. Level strategy relies on using inventory, backlog levels and overtime to handle fluctuations in demand while keeping the workforce and production levels at a constant. This approach provide stability in production and workforce planning allowing for retained workforce competence and production optimization for a specific

production rate, but may lead to excessive inventory and long lead times if demand deviates significantly from planned levels (Jamalnia et al., 2019). Chase strategy relies on chasing customer demand by scaling production rate and workforce according to foretasted demand. This strategy solves the issue of excessive inventory and long lead times but instead raises the problems of lost workforce competence and can be hard to implement due to constraints related to workforce availability and training requirements (Jamalnia et al., 2019).

Jamalnia et al. (2019) further suggest a mixed level and chase approach that combines elements of both strategies. In such an approach, a stable base capacity is maintained through a core workforce, while flexible resources are used to handle demand variability. For example, a core pool of skilled employees may be retained regardless of demand fluctuations, while temporary agency staffing is employed to manage short-term demand peaks.

2.3.1 Walking workers assembly line and fixed worker assembly line

Two strategies when it comes to chasing flow is Walking worker assembly line (WWAL) characterized by operators who follow a specific unit through multiple stages or the entire production sequence and Fixed worker assembly line (FWAL), where operators remain at assigned workstations while the product moves through the flow. The choice between WWAL and FWAL is central to optimizing system productivity under varying operational conditions. According to Catalano et al. (2025), several key parameters, including cycle time, system balance, and workforce characteristics, determine the suitability of each strategy.

FWAL is generally more productive when cycle times are short, as the time operators spend walking in a WWAL would become a too large percentage of their total work time. In these cases, the movement between stations creates a walking penalty that outweighs the benefits of a flexible line. WWAL becomes more advantageous as cycle times increases, as the time spent walking between stations decreases. System balancing also plays a critical role, where WWAL demonstrates superior performance in highly unbalanced flows. While FWAL remains sensitive to uneven task distribution across stations. The "chasing" way WWAL works allows for more fluid movement that effectively hides the impact of bottlenecks.

2.4 Manufacturing flexibility and scalability

Scalability within production defines as the systems capability to change its production volume to meet the fluctuation in demand in an cost effective manner.

Volume flexibility

Volume flexibility is defined as a systems ability to change its output levels over a specific time. It represents a vital capability for manufacturing organizations to manage fluctuating demand without losing operational efficiency or incurring excessive costs (Upton, 1994).

Design capacity

The design capacity refers to the theoretical maximum capacity of a process. The level of output a process is designed to produce under ideal conditions. It assumes a continuous flow without any interruptions, breakdowns, or losses. (Slack et al., 2019).

Effective capacity

Since no production line can operate at maximum design capacity in practice and it always includes losses, the concept of effective capacity is used to describe the output that is achievable after accounting for planned losses. These losses include scheduled maintenance, personnel breaks, and inherent constraints within production planning. (Slack et al., 2019).

Utilization

This metric measures the relationship between actual output and design capacity. It is mathematically expressed as equation 2.6.

$$\text{Utilization} = \frac{\text{Actual Output}}{\text{Design Capacity}} \tag{2.7}$$

A low utilization rate indicates that there is significant unexploited potential within the system, which creates the necessary space for scalability. By distinguishing between these levels, an organization can pinpoint exactly where operational losses occur and quantify the degree to which inefficiencies impact overall output. This serves as a theoretical foundation for identifying system-critical bottlenecks and determining the true capacity of the production flow. (Slack et al., 2019).

2.5 Losses in scaled production

As production systems scale, additional inefficiencies may emerge that reduce the achievable capacity of the production flow. According to Wild (1975), production performance is not only determined by the nominal cycle time of individual processes, but also by balancing losses, handling losses, and system losses. Combined, these losses affect the effective cycle time of operations and thereby influence the overall throughput and scalability of the production system.

2.5.1 Balancing losses

Balancing losses refer to inefficiencies caused by uneven distribution of work between workstations or operators. Uneven workload distribution results in certain operators completing their tasks earlier than others and therefore remaining idle while waiting for the next production cycle. According to Wild (1975), balancing losses generally increase as cycle times decrease, since shorter cycle times reduce the flexibility in distributing work evenly between stations. Consequently, reducing cycle times in order to increase throughput may not result in proportional capacity increases, as shorter cycle times may simultaneously increase balancing losses within the production system.

2.5.2 Handling losses

Handling losses refer to inefficiencies caused by supporting operations such as material handling, transportation, tool management, and internal logistics. According to Wild (1975), these activities constitute a larger proportion of the total work content as cycle times decrease. Consequently, handling losses may increasingly influence the effective throughput of the production system in a similar manner to balancing losses.

2.5.3 System losses

System losses refer to idle time and inefficiencies caused by variability and interruptions within the production flow, resulting in blocking or starvation between processes. According to Wild (1975), system losses are influenced by both the buffer capacity between workstations and the number of stations within the production system. Larger buffer capacities may decouple processes from one another, thereby reducing the impact of short-term disturbances and allowing downstream processes to continue operating independently of upstream fluctuations. In contrast, an increased number of interconnected stations increases the probability of disturbances arising and propagating throughout the production flow, thereby increasing the potential for system losses.

2.6 Analytical Framework

This study integrates concepts from Lean management, VSM, TOC, and Yamazumi charts, and manufacturing scalability in order to analyze production flow, identify bottlenecks, and evaluate system capacity and scalability.

Lean management provides the overall perspective on value creation and waste elimination, forming the foundation for analyzing production efficiency. VSM is used as the primary tool for visualizing material and information flows, enabling the identification of process inefficiencies and accumulation of work-in-progress. TOC is applied to identify system-critical bottlenecks, based on the principle that the throughput of a production system is determined by its constraint. This perspective is essential for both bottleneck identification and capacity analysis.

Line balancing principles, supported by Yamazumi enables a detailed evaluation of workload distribution across operators and workstations, making it possible to identify imbalances, overloaded stations, and underutilized capacity. This provides a basis for assessing theoretical staffing structures and how task allocation affects production flow.

From a planning perspective, APP, WWAL, and FWAL are used to interpret how the identified capacity limitations and staffing structures influence medium-term production decisions. In particular, the findings are related to different APP strategies, in order to evaluate how the production system can respond to demand variability while maintaining efficiency and workforce stability.

In addition, the concepts of balancing, handling, and system losses are used to understand the practical limitations of scaling the production flow. Recognizing these losses is important because increasing the workforce or reducing cycle times does not always

result in a perfectly linear increase in capacity. This perspective helps explain why the system's effective capacity may be lower than its theoretical design capacity, providing a more realistic foundation for evaluating production scalability.

Finally, the concept of manufacturing scalability is used to evaluate how the production system can adjust output in response to changing demand conditions. By analyzing how staffing adjustments influence bottlenecks and capacity, the study assesses the system's ability to scale production without introducing new constraints or inefficiencies.

Together, these concepts form a structured analytical framework where VSM provides the system overview, TOC identifies the system constraint, and line balancing evaluates how resources can be allocated to improve flow efficiency, and capacity analysis quantifies system performance. These operational insights are then interpreted through an APP perspective to assess planning implications and scalability. This integrated framework enables a comprehensive analysis of how bottlenecks, capacity, and staffing interact, directly addressing the research questions related to production scalability.

3
Method

This study adopted a deductive research approach where established theories and methods used within production planning and lean production were used to guide the analysis of an empirical case (Bell et al., 2022). A deductive approach is appropriate when research is based on existing theory and aims to examine theoretical concepts within an empirical context. An overview of the research flow is shown in figure 3.1 starting with deciding the problem and purpose of the study which then became the base of the literature study. The literature study focused on evaluating available methods to answer the research questions which determined what data that was gathered in the data collection phase of the study. As seen in figure 3.1 the outline of the data collection and validation phase is dotted. The reason for this is that this phase is further explained in figure 3.2 that shows the data collection and validation process in more detail. This study followed a mixed quantitative, qualitative approach where quantitative data was the main data source and qualitative data was used to verify quantitative data, a more detailed description of this process can be found in section 3.2. The collected data was then used to map the production flow using VSM as a standardized and structured analytical framework. The resulting map was then analyzed using established methods derived from the theoretical framework in order to identify system-critical bottlenecks, quantify actual production capacity, and analyze how staffing and line balancing influence system performance.

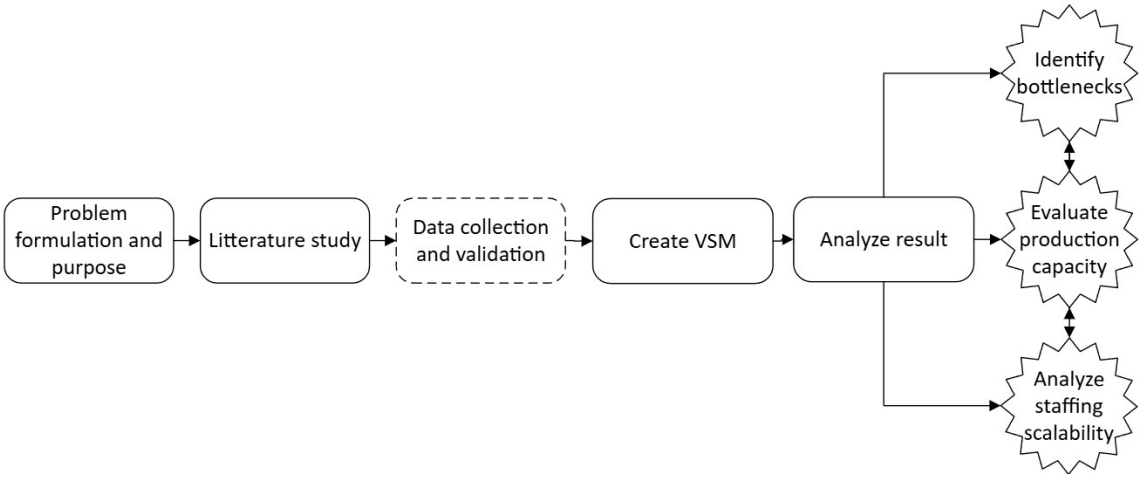

Figure 3.1: Method flow

3.1 Literature search

The gathering of literature was conducted by searching with keywords in different data bases such as Scopus and Google scholar. One specialized tool used in this study was Scopus AI, which assisted in constructing search strings and efficiently identifying relevant literature. Some keywords that where used are bottleneck, bottleneck detection, capacity, line balancing, lean management, value stream mapping, production losses, aggregated production planing, and theory of constrains. The literature selection was based on criteria such as citation count, publication year, and relevance to the research topic. Articles with a higher number of citations were prioritized to ensure credibility, while more recent publications were included to capture current research developments. Relevance was assessed based on how well the studies addressed key concepts related to bottlenecks, capacity, and production flow analysis.

3.2 Data collection

The success of this project is contingent upon rigorous data collection and the establishment of a validated dataset. The study used a mixed-methods approach that combines quantitative and qualitative data. The structure of the data collection method can be seen in figure 3.2 which describes the iterative data collection and validation procedure. The data collection process flow in figure 3.2 begins with the collection of quantitative data from CASAT in the form of Sequence Based Activity and Time (SAM) analyses. CASAT serves as the Vara plant's internal system for work instructions and time studies. Following data collection, a quality assessment was performed to determine the data quality. If the data where deemed sufficient, a validation process was initiated. Conversely, if the data where found to be insufficient, qualitative data gathering was conducted through various methods such as Gemba walk, interviews and expert panel. Which finally, was validated in consultation with subject matter experts to ensure accuracy. This multi-method approach enabled a triangulation loop, where quantitative datasets were cross-verified with qualitative insights through Gemba walks/interviews and expert panels to ensure a robust and consolidated foundation for the value stream mapping see figure 3.3.

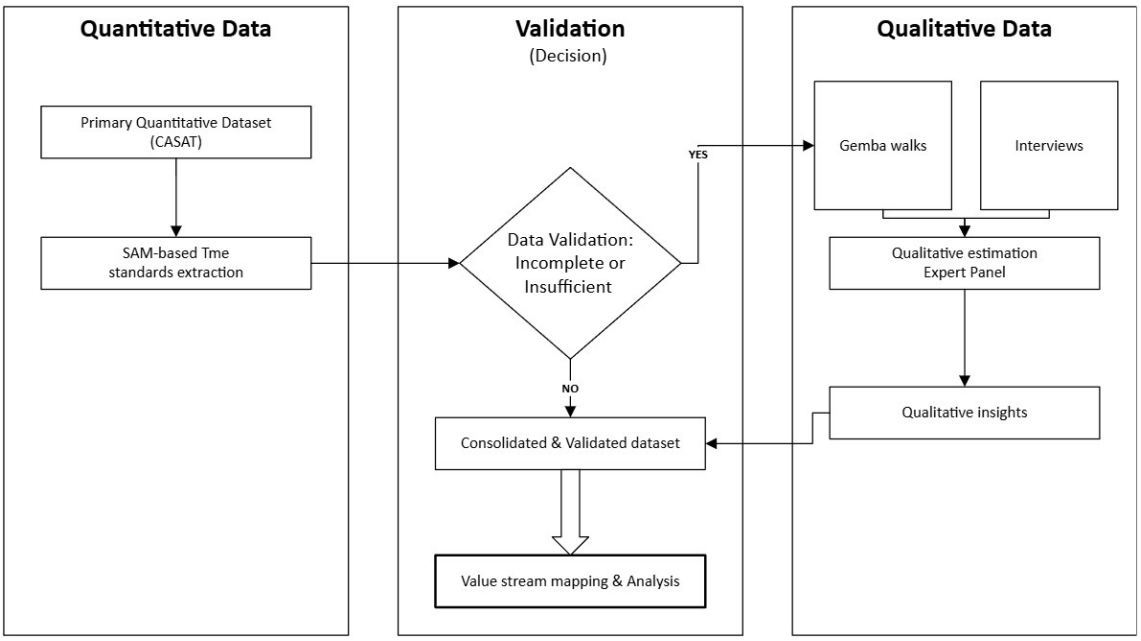

Figure 3.2: Detailed data collection and validation method

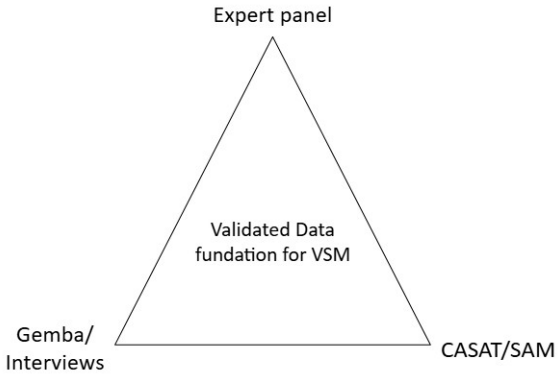

Figure 3.3: Triangulation

3.2.1 Quantitative data

The quantitative data focuses on gathering essential parameters including cycle times, process times, setup/changeover durations, and total available time, alongside comprehensive data on the number of operations, inventory levels, and material flow. Primary data was extracted from Volvo's internal system, CASAT, where time standards are derived from SAM analyses. SAM is a standardized tool used for time-studies to optimize work steps and create long-term time standards (MTM Föreningen Norden, n.d.). The tool is frequently used within Volvo group and is known international. In instances where system data was incomplete or insufficient, the project utilized qualitative data from Gemba walks/interviews and expert estimations. To ensure high validity, these estimations was developed in collaboration with experienced personnel through consensus-based validation.

3.2.2 Qualitative data

To complement the numerical data, qualitative insights was gathered to understand the underlying production flow. This was achieved trough interviews and Gemba walks, which is a concept derived from TPS that mean "going to the source" and not only relying on data collected by others (Liker & Convis, 2011). In this study Gemba walks was conducted through small internships of one day on each of the departments. This created a deeper understanding how the production flow actually operates and a clearer communication between experts and the researchers. This aligns with Liker (2021) who states that a full understanding of a production flow can only come from going to the source and observing the actual process. The qualitative insights derived from interviews and Gemba walks served to complement the quantitative dataset, thereby facilitating a comprehensive triangulation of the production flow. This integrated approach ensures that the analysis was grounded in both systemic data and the empirical reality of the shop floor.

3.2.2.1 Gemba walks

The Gemba walks were structured into four distinct segments. The first segment involved preparation and the establishment of goals for the observation. The second consisted of observing the flow to gain a firsthand understanding of the production process. In the third segment, findings were documented, and in the fourth, these findings were verified and compared against SAM analyses.

Two Gemba walks were conducted on each process step. The first was performed as above to establish a foundational understanding of the flow and the parameters of the SAM analyses. The second walk served to evaluate the analyses and SAM times, ensuring that every operation was accounted for and identifying any missing elements. This was done by printing out all SAM analyses for each station, from start to finish, and cross-referenced against the actual tasks performed by the operators in the line. This methodology significantly enhanced the reliability and validity of the SAM data trough real time observations and discussions with operators.

A important and helpful part in this study was that we the authors was situated at the Vara plant during the entirety of the study. This gave us the possibility to cross check the data when questions arose without needing to rely on a few scheduled study visits.

3.2.2.2 Interviews

To ensure a comprehensive understanding of the data, a combination of unstructured and semi-structured interviews was conducted. This multi-method approach was intended to enhance the credibility and triangulation of the study's results (Bell et al., 2022). Initially, unstructured interviews was carried out during Gemba walks, consisting of informal discussions with operators to capture real-time insights. Once a data overview was established via CASAT and these initial observations, semi-structured interviews was conducted. These aim of the session is to deepen the knowledge regarding how the plant utilizes data, as well as to verify its reliability and currency.

To ensure a comprehensive understanding of the production flow and to facilitate data triangulation, a series of interviews was conducted. The selection of respondents followed a purposive sampling strategy, where the choice of interviewee was determined by the specific nature of the data or validation required (Bell et al., 2022). When a high-level overview of the operational context was needed, interviews were held with management. Conversely, when deeper technical insights were required, operators or personnel working directly with the specific information flow were consulted. For instance, to validate the sequence of work tasks against the SAM data, or to ensure that every manual operation was accurately represented in the system at the specific workstations.

This rigorous approach ensures that the dataset is derived directly from the shop floor, confirming that the SAM time standards accurately reflect the empirical reality of the production process. An Overview of the interviewed respondents, the purpose for their selection and method used is presented in table 3.1. This qualitative phase constitutes a critical component of the iterative validation loop previously described in figure 3.2.

Table 3.1: Overview of interviewed respondents, the purpose for their selection and method used.

Role	Purpose	How
Production flow		
Manager, flow 1	Understanding operations, staffing, takt time, and lead time.	Interview
Manager, flow 2, 3, testing	Understanding staffing, takt time, and lead time.	Interview
Operator, flow 1	Deepening process understanding and validation of data.	Interview & Gemba
Operator, Engine Testing	Gaining insights into engine testing procedures and constraints.	Interview & Gemba
Operator, Processing	Deepening understanding of the machining and processing stages.	Interview & Gemba
Operator, flow 2	Deepening process understanding and validation of data.	Interview & Gemba
Global Process Manager	Understanding outbound logistics from the factory.	Interview
Manager, Distribution	Understanding inbound and outbound logistics at the factory level.	Interview
Information and Side Processes		
Manager, Digitization & IT	Mapping and understanding the digital information and data flow.	Interview
Production Technicians	Validating technical production data and standards for flow 1.	Interview
Production Planner	Understanding production scheduling and demand management.	Interview

3.2.2.3 Cross-functional expert panel

To ensure that the analysis reflects the complex reality of the D4/D6 production line, cross functional panels was established. According to Bell et al. (2022), engaging stakeholders in the research process is a form of respondent validation, which strengthens the credibility of the results. The panels was comprise of a diverse range of roles, including operators, production planners, and production engineers. The primary objective of this group was to reach a consensus regarding production flow, ensuring they accurately reflect the empirical reality of the shop floor. The composition of the expert panels was tailored to ensuring that the specific expertise of the participants aligned with the particular problem being addressed.

While the previously conducted Gemba walks and interviews focused on gathering diverse individual perspectives and specific operational details, the cross-functional panels served a different methodological purpose. These panels was organized after the interviews to facilitate a transition from individual observations to a collective agreement. By presenting the data and findings from CASAT and interviews to the panel can conflicting information be addressed and the different viewpoints could be translated into a single,

validated version of the reality.

This step was important and critical because the interviews provided the raw, bottom-up insights needed to challenge the official standards, while the expert panel provided the horizontal validation across departments. Conducting the panel sessions after the interviews ensures that the discussion was grounded in verified shop-floor reality which allowed the group to reach a final consensus on production times and flow dynamics that no single person/source could provide alone. This last step with horizontal validation was what transforms the collected data into reliable foundation for the VSM, which can be seen figure 3.2.

3.3 Analyzing the VSM

The analytical phase of this study was centered around the ”as-is” as a diagnostic baseline for evaluating system capacity. In contrast to a traditional VSM process that aims for a future state map, this approach focuses on the ”as-is” state to facilitate a quantitative stress-test of the existing production configurations.

After constructing the current VSM, the map was analyzed in order to identify system critical bottlenecks, determine the maximum production capacity and evaluate staffing structures for the production flow. The analysis followed a structured approach based on the theoretical concepts presented in the theoretical framework, mainly the TOC, line balancing and VSM. In this study, a bottleneck is defined as any process where the cycle time exceeds the takt time (Rother & Shook, 1999), or where accumulation of work-in-progress is observed in the VSM (Librelato et al., 2014). Capacity is defined as the maximum achievable throughput determined by the cycle time of the system bottleneck in relation to available production time (Slack et al., 2019). Scalability is evaluated as the system's ability to increase output through staffing adjustments without introducing new bottlenecks.

3.3.1 Bottleneck analysis

The first step of the bottleneck analysis was to evaluate the production flow by comparing the cycle time of each process with the calculated takt time for the whole system. Takt time represents the pace at which the production system needs to operate in order to meet customer demand. Therefore takt time is used as a benchmark when evaluating whether a process has sufficient capacity (Rother & Shook, 1999). The bottleneck analysis primarily focuses on the assembly processes.

The machining processes were analyzed separately, as they operate against stock rather than direct customer demand and therefore do not follow the same takt-driven production logic. In addition, the machining processes operate under different production conditions, including a two-shift system, whereas assembly is conducted in a single-shift setup. Furthermore, future changes in the product mix, including the introduction of an additional engine models that bypasses machining, imply that increased production volumes will primarily affect the assembly processes.

Processes where the cycle time exceeded the takt time were considered potential bottle-

necks, as these processes risk limiting the overall throughput of the system. In addition to the comparison between cycle time and takt time the accumulation of inventory or work-in-progress visualized in the VSM timeline was used to support the identification of bottlenecks. Such visual indicators are commonly used within value stream mapping to reveal imbalances in production flow and identify processes where material tends to accumulate due to capacity constraints (Librelato et al., 2014).

The combination of these indicators provided a structured basis for identifying processes that potentially constrained the throughput of the production system. To investigate how the bottleneck is affected by changes in production scale, the takt time was incrementally reduced in order to observe when and where the next bottleneck appeared in the production system. During this analysis, the staffing levels were adjusted in conjunction with the reduced takt time in order to evaluate how different staffing configurations influenced the system's capacity and bottleneck location.

3.3.2 Capacity analysis

Once the bottleneck process had been identified, the next step was to estimate the maximum production capacity of the system. Since the throughput of a production system is determined by its bottleneck, the system capacity was calculated based on the cycle time of the bottleneck in relation to the available production time (Slack et al., 2019).

To evaluate the maximum achievable capacity within the existing production system, staffing levels were incrementally adjusted. This was done by reallocating personnel in order to reduce the cycle time of the bottleneck process. Adjustments were continued until additional staffing no longer resulted in a reduction of the bottleneck cycle time, at which point the system was considered to have reached its theoretical capacity limit under the current production configuration.

This approach primarily applies to the assembly processes, where cycle times are influenced by manual work and therefore largely affected by staffing and task redistribution. In contrast the machining process is largely determined by machine processing and is therefore not largely affected by changes in staffing levels. The capacity in the machining was therefore instead adjusted using additional shifts.

However, to further examine the design capacity of the production system, an additional scenario was analyzed in which each process was assumed to be perfectly balanced. This was achieved by distributing the total work content evenly across the available workstations within each process. The resulting bottleneck cycle times from these scenarios were then used as the basis for calculating the maximum production capacity.

This analysis provides insight into how the production system scales with increased staffing and line balancing, and highlights the structural capacity limits of the current production configuration.

3.3.3 Staffing structure analysis

The staffing analysis was conducted to evaluate how changes in the number of operators influenced system performance. Staffing levels were incrementally adjusted, and the resulting effects on bottleneck behavior, capacity utilization, and overall throughput were analyzed. The purpose of this analysis was to examine how the production system responds to increased staffing under different demand conditions, and to identify thresholds where additional personnel no longer contribute to increased capacity.

This approach provides insight into how staffing influences system scalability and highlights the limitations of increasing capacity through additional labor. This is consistent with the Theory of Constraints, which states that system performance is governed by its bottleneck, implying that increases in resources outside the constraining process do not necessarily lead to improvements in overall system capacity (Rahman, 1998).

3.3.4 Operational Definitions of Flow Metrics

Different flow metrics and operational definitions can be see in table 3.2.

Table 3.2: Operational Definitions and Formulas.

Acr.	Full Name	Formula	Explanation
TW	Total Work Content	$\sum$(operation times)	Total active time to get one product through the process. According to SAM analyses
TW2	TW-Schedule	-	Active total work time according to schedule.
WIP	Work in Progress	Products in production flow	Unfinished products in the process.
TPT	Total paid Time	Paid time	Total time workers are paid for.
AT	Available Time	TAT - Planed downtime	The available production time.
UT	Utilization	Used time / AT	Percentage of used time.
T/T	Takt Time	AT / Demand	Required production pace per product.
L/T	Lead Time	T/T $\times$ WIP	Total time for one engine through the process.
C/T	Cycle Time	TW / Operators	Potential production pace per product.
LOT	Longest Operation Time	Max(operation time)	The longest single operation time.
VA	Value Added time	TW - (Buffer $\times$ Takt time)	Time where value is added to the engines.
NVA	Non-Value Added time	L/T - VA	Time spent waiting in buffers, etc.
Op	Operator	-	Person working in line

3.4 Assumptions

The analysis conducted in this study is based on several simplifying assumptions. Firstly, process times are assumed to be deterministic and constant, as they are collected from

SAM analyses. Consequently, short-term variability in processing times caused by operator differences or minor operational disturbances was not analyzed separately, but instead accounted for through a utilization allowance applied to reflect operational losses and sustainable working conditions.

Secondly, the analysis assumes a consistent product mix, focusing on the most frequent engine variants based on a months production data. This simplifies the modeling of the production flow and makes the visualization tools less complicated making them easier to understand but may not fully capture variations introduced by less frequent engine variants. This assumption was developed in collaboration with technicians and operators, whose input indicates that these engine variants are representative of all variations.

Thirdly, the production system was analyzed under assumed stable operating conditions in order to evaluate the underlying structural capacity limitations and bottleneck behavior of the production flow. Consequently, major disturbances such as machine breakdowns and material shortages were not included in the analysis, as they represent deviations from normal operating conditions and were therefore considered outside the scope of this study. Additionally there was a lack of data related to material shortages and breakdowns.

As the machining line was not yet operational at the time of data collection, the machining data is based on estimated performance. Furthermore, since the future product mix is not fully reliant on the machining process, the machining-related results should be interpreted as indicative rather than definitive.

These assumptions enable a structured and comparable analysis of the production system, but should be considered when interpreting the results.

3.5 Validity and Reliability

To ensure the quality and trustworthiness of the study, validity and reliability concepts will be applied. According to Bell et al. (2022), these concepts serves as the primary criteria for evaluating the quality of business research.

Reliability refers to the consistency and stability of the research process and the extent to which similar results could be obtained if the study were repeated using the same methods (Bell et al., 2022). In this study, the reliability is supported through the use of standardized data sources and methods, including SAM-based time standards and structured value stream mapping. In addition, the use of a cross-functional expert panel reduces the risk of individual bias by establishing consensus on key parameters such as process times and bottleneck identification.

Internal validity concerns the degree to which a study´s findings accurately reflects the reality of the situation (Bell et al., 2022). In this study, internal validity is secured through data triangulation (Patel & Davidson, 2019), where quantitative production data from the CASAT system is combined with qualitative data from interviews and direct observations conducted during Gemba walks. To ensure internal validity during Gemba walks, the potential impact of the Hawthorne effect (Roethlisberger & Dickson, 1939) will be addressed by maintaining a long-term presence at the plant and clear communication of

the aim and purpose of the study. By comparing and validating findings across these complementary data sources, the study reduces the risk of relying on assumed or incomplete representations of the production flow and ensures that the analysis reflects the real "as-is" condition of the system.

Furthermore by focusing the analysis on high-volume engine variants, the study ensures that the identified bottlenecks are representative of the plant's primary operational load. While the empirical analysis is limited to a single production flow, the deductive and theory-driven research design enables analytical insights that are applicable to manufacturing systems facing similar challenges related to demand variability, capacity constraints, and scalability (Yin, 2018).

The external validity of this study is limited by its focus on a single case, namely the D4/D6 production flow at Volvo Penta in Vara. As the analysis is based on a specific production system with its own layout, processes, and product characteristics, the qualitative results, such as identified bottlenecks and capacity levels, cannot be directly generalized to other production environments. However, the analytical approach applied in this study, combining value stream mapping, bottleneck analysis based on the Theory of Constraints, and line balancing, is based on established and widely used methods within production engineering (Rahman, 1998; Rother & Shook, 1999; Slack et al., 2019). These methods are not case-specific and can therefore be applied to similar manufacturing contexts, particularly in variable-demand environments (Voss et al., 2002). Consequently, while the specific results are dependent on the production context, the findings aim to contribute to the theoretical understanding of how bottlenecks, capacity, and staffing interact, and may provide transferable insights for similar production systems. Rather than representing all production systems, the results are analytically generalized to theory (Yin, 2018).

3.6 Ethical consideration

This study will be conducted with employee privacy in mind, where focus will be on process rather than individual effort. All company-specific data will be handled within applicable confidentiality to secure that no sensitive data can harm the company. Furthermore, work environment aspects will be taken into account by analyzing if line balancing can reduce uneven workload, which could contribute to a more sustainable working situation for operators.

3.7 Use of Artificial Intelligence in the thesis

Artificial Intelligence (AI) was utilized as a supportive tool throughout the research process. AI tools were employed for linguistic refinement and grammatical editing to enhance text clarity and readability; however, this assistance was strictly confined to surface-level editing and did not influence the core content, analysis, or conclusions of the study.
Additionally, AI-assisted discovery tools, including Scopus AI, were leveraged during the literature review phase to optimize search queries and identify relevant scientific literature. All literature identified through these means was independently evaluated and vetted by the authors to ensure academic rigor and relevance.

4
Current production

This chapter presents the current production layout and current state VSM. The goal with this chapter is to provide an understanding of current production strategy and layout in the Vara plant and establish a foundation for analyzing how staffing and layout affects the bottleneck and capacity.

4.1 Production layout and process steps

The production layout is structured as illustrated in figure 4.1. The figure illustrates five of six key process steps, assembly flow 1, paint shop, assembly flow 2, engine testing, and assembly flow 3. The preceding machining process, although physically located adjacent to flow 1, operates under different conditions due to production planning and capacity. As mentioned before in section 2.2.3 the Vara plant uses walking-workers as their primary production strategy. This is used in flow 1, flow 2 and flow 3 which are the steps where most of the operators are stationed.

In these assembly flows, the walking-workers strategy dictates that operators follow a specific engine throughout the entire sequence of stations within a given flow. This methodology separates the number of active operators from the total number of workstations, allowing the production line to maintain a high degree of volume flexibility. By adjusting the operator headcount within the operator pool, the plant can scale production capacity in response to fluctuating demand without requiring physical reconfigurations of the line layout. However, this flexibility also requires careful monitoring of internal buffer levels and workstation balance to ensure that the LOT does not become a restrictive constraint as the workforce is scaled.

4.1.1 Process steps

Production in Vara initiates with raw casted cylinder blocks and heads, which undergo machining in the first process step. Following machining, components are held in intermittent storage before being transported to assembly flow 1. A defining characteristic of the machining stage is that it produces towards future demand and stock rather than being triggered by specific customer orders. This shortens the lead time for the end customer, as the customer-specific lead time is initiated only when an operator prints out a work order card at the first station in flow 1.

Following machining, assembly flow 1 consists of a fully manual assembly line where the cylinder head and block are joined and all internal components are installed. The engines subsequently undergoes leak testing before being transported to the paint shop via an autonomous forklift. The current configuration of flow 1 consists of 24 stations, currently manned by seven operators.

The paint shop process step consists off a manual engine lift of from forklift, a manual masking station, two automated painting robot cells (one for the left and one for the right side), a drying tunnel and a manual unmasking station. Each engine takes two laps in the painting and drying stations, one for primer paint and one for final paint. Most of the process time in this process step is used by the robots and drying tunnel not the operators. In current production, one operator is stationed in the paint shop.

The next process step is flow 2, an assembly line where external components, hoses and cables are assembled on the engines before being sent to engine testing. This process step is located directly next to the paint shop, see figure 4.1 and has seven stations which are manned by five operators.

In the engine testing process each engine is filled with oil and coolant and connected to a test bench, which executes an automated testing program. The engines are then drained of liquids and put through a cooling tunnel before continuing to assembly flow 3 which is the final assembly of the Vara plant. Engine testing does not follow a walking-workers configuration, the testing of the engines is parallel in testing boxes. In the current production setup, three boxes are used, each manned by one operator, although seven boxes are available for testing engines. See figure 4.1 for engine testing location

In assembly flow 3 the engines are finalized to the customers specification and packaged before being sent to final storage to wait for delivery. Flow 3 has eight stations and is manned by two operators. In table 4.1 below is a summary of how many stations there is in each process step.

Table 4.1: Summary: number of stations in each process step

	Flow 1	**Paint shop**	**Flow 2**	**Engine testing**	**Flow 3**
Number of stations	24	3	7	7	8

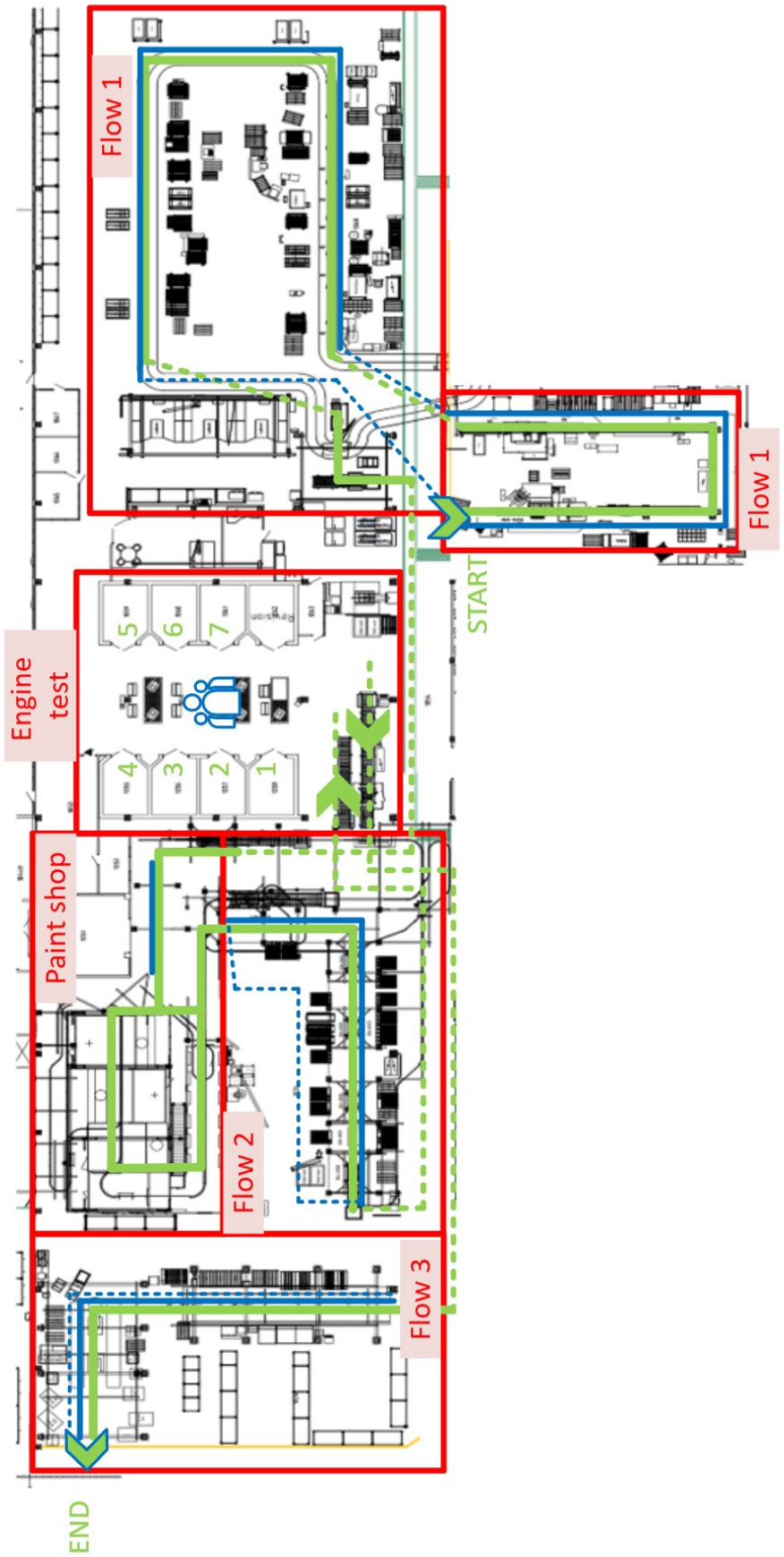

Figure 4.1: Plant layout with process steps. Green lines are the engines path, blue lines are staffing path and dashed lines are transportation

4.2 Current State VSM

This chapter will break down and present the information in the current state VSM that was created in this study in order to answer the research questions. The current state value stream map of the D4/D6 flow is presented in figure 4.2 and a higher resolution figure can be found in the appendix A.1.

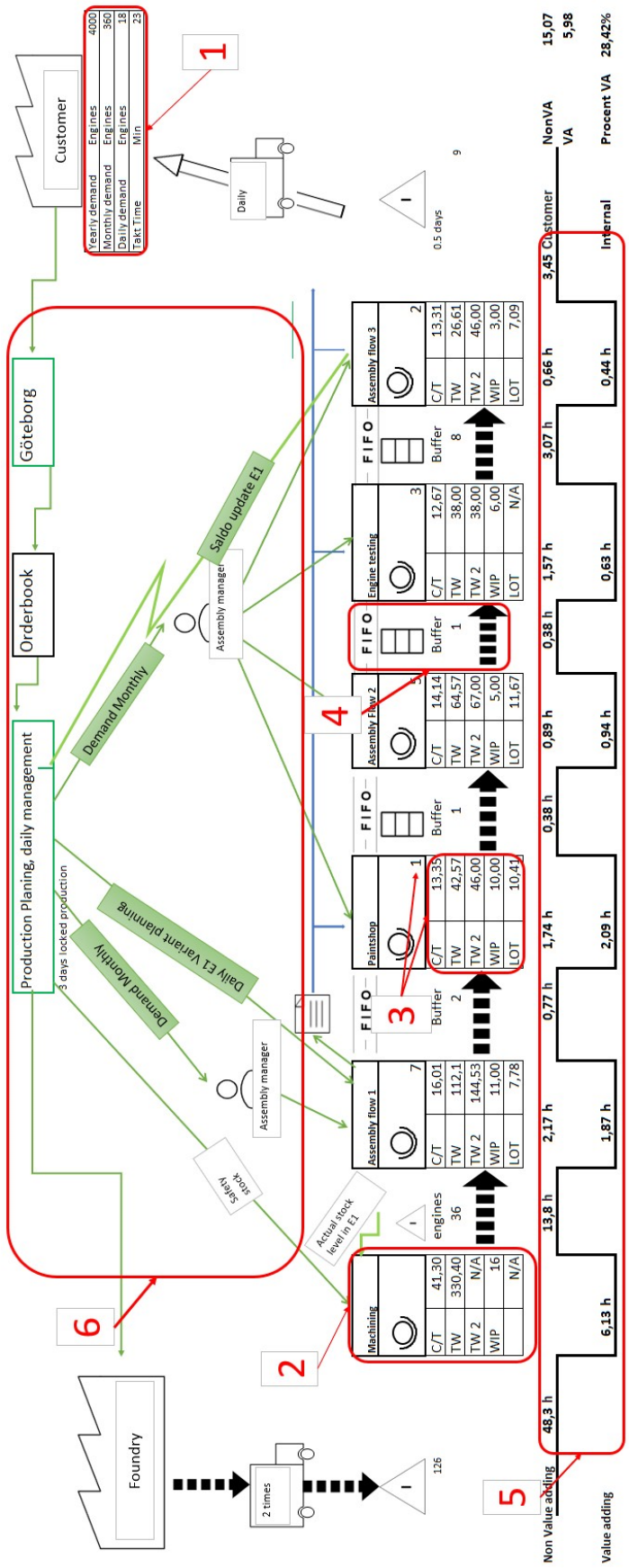

Figure 4.2: Current state map of the D4/D6 production flow with numbering

Table 4.2: Description of numbered elements in the VSM.

Number	Description
1	Customer Data box
2	Process step
3	Process data
4	Material flow and buffers
5	Timeline
6	Information flow

4.2.1 Customer demand and takt time

The current customer demand is 360 engines monthly which translates to 18 engines a day. During the study it was noted that the available time differed between the process steps which theoretically leads to different takt times. How the available time was calculated and why it differed is described in section 5.2.1. The overall takt time that is displayed in box number one in figure 4.2 is calculated on the shortest available time which was 398 minutes which lead to the takt time of 22 minutes using equation 2.2.

4.2.2 Process steps

The process steps, machining, flow 1, paint shop, flow 2, engine testing, and flow 3 previously described in section 4.1 can be seen in figure 4.2 as box two together with the five other similar boxes in the VSM. They show the process steps order and holds information about the specific process data.

4.2.3 Process data

For each process step, process data was collected and displayed as seen in box three in figure 4.2. The same data is summarized in table 4.3 below. The data indicates variation in workload between processes, indicating differences in utilization across the production system. Some processes operate closer to the takt time, while others show significantly lower cycle times. Machining stands out with cycle time of 41,3 minutes which is significantly higher than the calculated takt time of 22 minutes. But what needs to be taken in consideration is that the machining process operates at two shifts leading to about double the available time making the machining takt time longer than the other processes. The machining process also has a high level of automation which makes it possible to continue production during shorter breaks.

Table 4.3: Overview of the process data that can be found in the VSM

Process data	Machining	Flow 1	Paint shop	Flow 2	Engine testing	Flow 3
Staff [Pers]	N/A	7	1	5	3	2
C/T [min]	41,3	16,01	13,35	14,14	12,67	13,31
TW [min]	330,40	112,10	42,57	64,57	38,00	26,61
TW2 [min]	N/A	144,5	46	67	38	46
WIP [pcs]	16	11	10	5	6	3
LOT [min]	41,3	7,78	10,41	11,67	N/A	7,09

4.2.4 Material flow and buffers

After defining the process steps and data, the material flow between processes was mapped to provide an overview of how products move trough the production system as seen i box four in figure 4.2. There are buffers separating each process, but their size and type differ between machining and assembly steps. The buffer after the machining process is a warehouse and significantly larger than the ones between the assembly processes. The reason for this is that various buffers serve different purposes. The buffer after the machining process is used to handle production disturbances and allows the machining process to run two shifts which is necessary because of the significantly longer cycle times in the machining process see table 4.3. The buffers between the assembly processes are instead of the First In First Out (FIFO) type and have the main purpose of eliminating waiting time due to transportation. Therefore these buffers are much smaller often just one engine. What was noticed when analyzing the process data was the existence of internal buffers in the different processes. These buffers where also FIFO buffers and had the same purpose as the buffers between the assembly processes. To visualize these buffers, WIP was added to the process data, the difference between the amount of operators and WIP indicates the size of the internal buffer.

4.2.5 Value added time

As seen in box five in figure 4.2, a timeline was created in the bottom of the VSM. The timeline presents the lead time of an engine going through the factory and the relation between value-adding processing time (VA) and non-value-adding time (NVA) within the system. This is visualized in a zigzag formation where the valleys represent the VA and the peaks represent the NVA. To handle the internal buffers within the specific processes, some valleys contain two values; in these instances, the upper number represents the internal NVA within the process step (e.g., internal FIFO wait time), while the bottom number represents the actual VA. The analysis distinguishes between an internal perspective and a customer perspective due to the difference in production logic. During the initial machining stage, engines are produced to stock rather than to a specific order, as evidenced by the large warehouse buffer following the machining process as discussed in section 4.2.4. It is first upon entering assembly flow 1 that the engine is assigned a specific "owner" in the order book and produced according to the product specification that is printed at the first station of flow 1. The results from the timeline is presented in table 4.4.

Table 4.4: Results of timeline in VSM

Perspective	L/T [h]	VA [h]	NVA [h]	VA %	NVA %
Internal	82,50	11,25	71,25	13,50	86,50
Customer	20,99	5,39	15,60	25,91	74,09

4.2.6 Information flow

The information flow of the production system is illustrated in box 6 in figure 4.2. Customer orders are received by the sales department in Gothenburg and registered in a digital order book. The production planning department in Vara then reviews the order book and determines the monthly demand, which is communicated to the assembly managers. Based on this information, the assembly managers establish the takt schedule for their respective assembly processes.

The production planning department also determines the sequence in which customer orders are to be produced and enters this information into E1, Volvo Penta's enterprise resource planning system (ERP). At the first assembly station, an engine card is printed from E1 containing information regarding engine variant and customer assignment. The engine card is then physically attached to the engine and follows it throughout the remaining production flow.

In contrast to assembly, the machining process is not directly governed by customer orders. Instead, machining operates according to a push-based principle, where production is planned against a predetermined safety stock level based on forecasted monthly demand. Production in machining therefore continues until the target inventory level in the warehouse has been reached.

4.3 Takeaway from current production

This chapter presents the foundation of the D4/D6 production line through the development of a current VSM. The mapping reveals a production environment characterized by high volume flexibility, supported by a walking-workers strategy and significant internal buffers.

Key takeaway, at current demand of 18 engines per day, the system operates without a active bottleneck, as all process cycle time remain below the required takt time. Secondly, the analysis shows a significant difference in lead time between the internal perspective 82,5 hours and the customer perspective 20,99 hours, primarily due to the strategic inventory held after the machining process.

The validated data and flow visualizations presented in this chapter serve as a basis for the subsequent analysis. In chapter 5, these findings are utilized to test the systems constrains, evaluate where bottlenecks occur when demand rises and shifts, and how the identified capacity limits impact the overall scalability of the D4/D6 production line.

5
Results

This chapter presents the results obtained from the analysis of the current state value stream map. The results are structured in accordance to the analytical methods presented in section 3.3 and focus on identifying system-critical bottlenecks, determining the maximum production capacity, and evaluating theoretical staffing configurations for improved scalability in order to answer our research questions.

5.1 Bottlenecks

The identification of bottlenecks was based on the comparison between cycle time and takt time supported by observations of work-in-progress accumulation in the VSM.

With a current takt time of 22 minutes and the existing staffing structure, the results indicate that no process exceeds the takt time, as shown in table 4.3. This suggests that there is no active bottleneck in the production system under current operating conditions. Again the machining process is a bit misguiding in the table as it runs with two shifts and have a higher available time giving it a higher takt time of 47 minutes. This indicates that the process currently runs below its theoretical capacity and is constrained by production demand rather than process capacity. This was further supported by Gemba walks as no excessive accumulation of WIP was noticed at any time. The inventory levels visualized in the VSM are instead primarily determined buffer levels, as discussed in section 4.2.4. These are intended to accommodate transportation and handling between processes rather than being a result of downstream constraints. However, this situation changes as demand increases, which reveals underlying capacity limitations within the system.

5.1.1 Bottleneck development in the assembly under increased demand

As demand rises the takt time decreases which leads to new bottlenecks emerging in the system. Based on the data displayed in table 4.3 the current process can produce at a takt time of 16 minutes before assembly flow 1 becomes a bottleneck. In the following analysis the takt time was incrementally reduced corresponding to an increase in demand by one engine at a time, while staffing levels at each process were adjusted accordingly. This approach allows for an evaluation of how bottlenecks shift within the system as production volume increases. Figure 5.1 presents Yamazumi charts at the critical takt times where the cycle time of specific processes exceeds the takt time, thereby identifying the active bottleneck in each scenario.

Scenario 1
In scenario 1, one more operator is introduced in flow 1 then the cycle time drops to 14,01. The next critical takt time is 14,08 where flow 2 becomes the bottleneck with a cycle time of 14,14 minutes as seen in figure 5.1.

Scenario 2
In scenario 2, one more operator is introduced to flow 2 which reduces the cycle time to 11,40 minutes, which is not possible due to the LOT of 11,67 minutes in flow 2. The LOT becomes an internal bottleneck, in this case limiting the overall throughput of flow 2 and limits the cycle time to 11,67 minutes as seen in scenario 2 in figure 5.1. With the next critical takt time of 13,52 min the bottleneck returns to flow 1 with a 14,01 minute cycle time.

Scenario 3
In scenario 3 the staff in flow 1 is increased once again which reduces the cycle time to 12,46 minutes. The new critical takt time is 13 minutes and the bottleneck moves to the paint shop and flow 3 with a cycle time of 13,35 minutes as seen in scenario 3 in figure 5.1.

Scenario 4
In scenario 4 the staff in both the paint shop and flow 3 was increased. One more operator in flow 3 brings down the cycle time to 8,87 minutes. Introducing one more operator in the paint shop reduces the calculated cycle time to 6,68 minutes which is below the LOT of the paint shop resulting in the actual cycle time for the paint shop becomes 10,41 minutes. The next critical takt time is 12,51 minutes which moves the bottleneck to the engine testing which has a cycle time of 12,67 minutes as seen in figure 5.1.

Scenario 5
In scenario 5 the staff in the engine testing process is increased with one person which decreases the cycle time to 9,50 minutes. The next critical takt time is 12,07 minutes and the bottleneck once again moves back to flow 1 with a cycle time of 12,46 demonstrated in scenario 5 in figure 5.1.

Scenario 6
In the last scenario 6 one more operator is introduced in flow 1 which reduces the cycle time to 11,21 minutes. The next critical takt time is 11,65 minutes which moves the bottleneck to flow 2 as demonstrated in scenario 6 in figure 5.1. Because flow 2 has reached its LOT of 11,67 minutes introducing one more operator will not affect the cycle time of the process as it will still be constrained by the slowest operation in the process in accordance with the TOC. A conclusion was therefore made that flow 2 is the primary bottleneck of the assembly flow that limits the scalability of the whole production line.

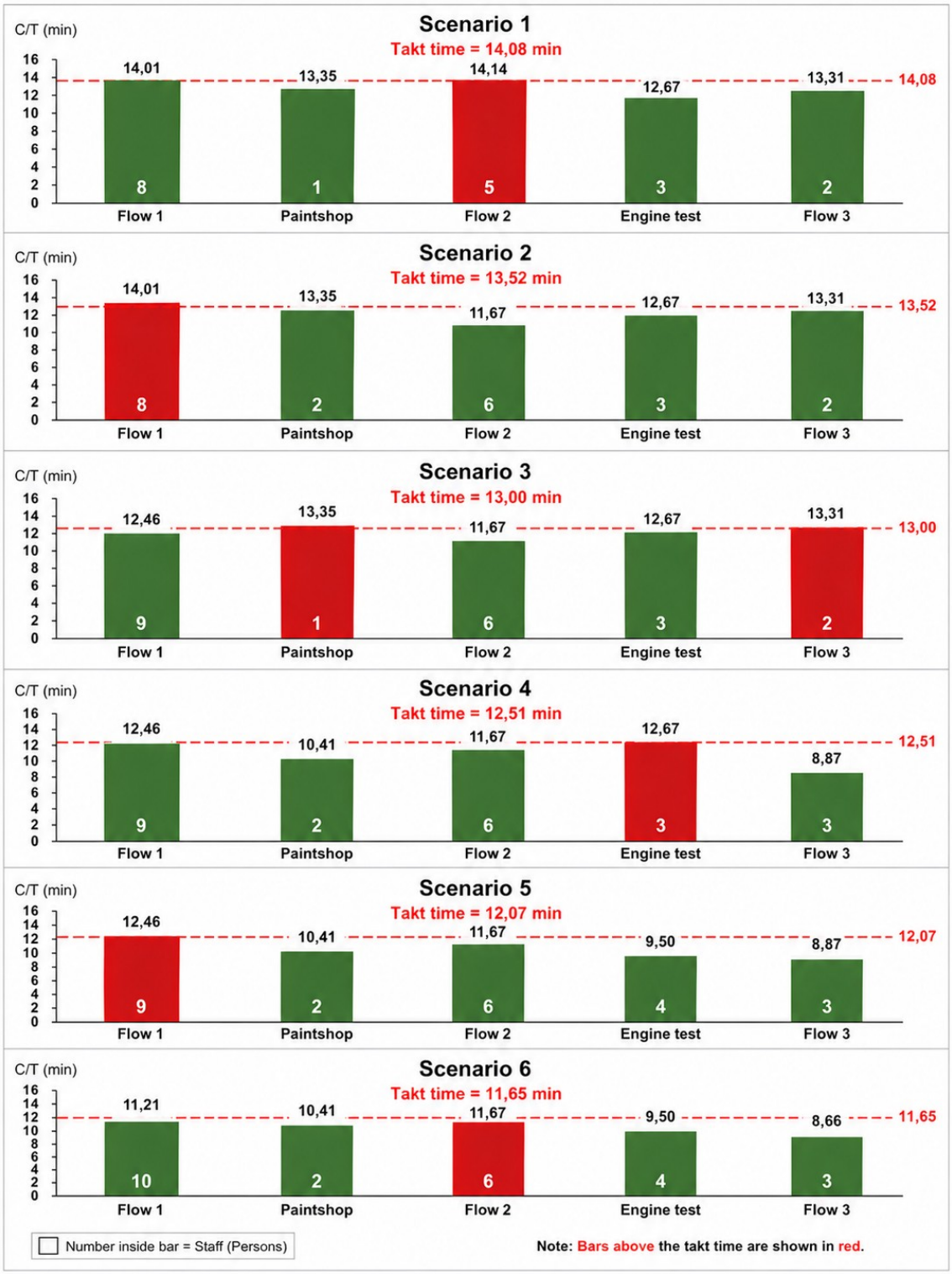

Figure 5.1: Yamazumi charts showing bottleneck propagation

5.1.2 Accounting for system losses

As discussed in Section 2.5.3, system losses largely depend on buffer capacity (Wild, 1975). In a WWAL configuration, empty stations between operators can function as internal buffers by absorbing short-term disturbances and reducing the risk of blocking and starvation. In the current state, each process step contains a number of empty stations, as presented in table 5.1. Since the internal buffer capacity is above two stations in most process steps, a general system loss assumption of 5% was applied in the bottleneck calculations. This assumption is based on Wild (1975) system loss curve, where system losses remain approximately around this level when buffer capacity is higher. However, as additional operators are introduced, this internal buffer capacity is reduced. This indicates that increasing the workforce may simultaneously increase the sensitivity of the flow to system losses.

The increase in system losses is not expected to be linear. As illustrated by (Wild, 1975), the difference in system losses is relatively small when buffer capacity is already high, whereas the effect becomes more pronounced at lower buffer levels. In the context of this study, this suggests that reducing the internal buffer capacity from two stations to one may substantially increase the sensitivity of the flow to disturbances. In the bottleneck analysis, adding one more operator to flow 2 reduces the internal buffer capacity from two to one. As a sensitivity adjustment, this reduction in buffer capacity is therefore assumed to increase system losses from 5% to 10%. If the cycle time is adjusted from the assumed 5% loss level to the potential 10% loss level, the effective cycle time for flow 2 increases from 11.67 minutes to 12.23 minutes, as shown in Equation 5.1.

$$11,67 * \frac{1,10}{1,05} = 12,23 \text{ min} = \text{New C/T} \tag{5.1}$$

This adjustment would make flow 2 the bottleneck already in Scenario 5 instead of Scenario 6. However, the overall conclusion remains unchanged, as flow 2 still represents the limiting process step. The inclusion of system losses therefore shows that the scalability of the WWAL configuration becomes less linear as the process steps become more fully staffed and internal buffer capacity is reduced.

Table 5.1: Summary: Internal buffers

	Flow 1	**Paint shop**	**Flow 2**	**Engine testing**	**Flow 3**
Buffer capacity	17	4	2	4	6

5.1.3 Summary bottleneck development with machining

As stated in section 5.1, the primary bottleneck within the assembly system is identified as flow 2, corresponding to a takt time of 11,65 minutes or 12,23 minutes if potential system losses are accounted for. This conclusion is based on the assumption that the new machining line is capable of supplying engines at this required rate. Under current conditions, the machining process has a cycle time of 41.3 minutes, as shown in table 4.3, which would make machining the system bottleneck if considered in isolation, even when operating under a two-shift system. However, unlike assembly, the machining cycle time is largely determined by machine processing time and is therefore not significantly

affected by changes in staffing levels. Instead, capacity in machining can be increased by extending operating hours, for example by introducing a third shift. This provides an alternative means of scaling capacity compared to assembly, where capacity is more directly influenced by staffing. Additionally future changes in the product mix will result in a lower relative load on the machining as not all engines will be processed in this stage.

5.2 Capacity

The capacity of the D4/D6 production line depends on several different parameters, first the available production time, second the staffing and third the bottleneck previously described in section 5.1. These three areas are what determines the production capacity of this production line.

5.2.1 Available time

Available time is the actual time where production can operate. Available time is calculated as total paid time when all planed downtime which can be breaks, meetings and cleaning is subtracted. Different departments have different available times based on differences in planed breaks, cleaning and production scheduling. Even though the departments has different available time the plant has a uniform takt time chosen for all departments to ensure a steady flow see table 5.2.

Table 5.2: Available production time and takt time per process step.

Process step	**Available time [min]**	**Takt time [min]**
Machining	850,6	47
Flow 1	398	22
Paint shop	427	22
Flow 2	427	22
Engine testing	427	22
Flow 3	427	22

At the plant, an agreement with the union specifies that Volvo should operate at a speed of 114% relative to a normal operator's pace. This pace is applied to the SAM analysis and a new process time is established for each specific operation. An example of a calculated process time can be seen in equation 5.2.

$$\text{SAM calculation 10 min} \implies \frac{10}{1,14} = 8,78 \text{ min} = \text{Volvo process time} \tag{5.2}$$

When designing a manufacturing line, a sustainability factor is often used to adjust utilization and protect the health of the operators and long-term performance of the operators. At Volvo Penta Vara, a target utilization of 85% is pursued, which defines the maximum capacity for sustainable production. This factor also acts as a buffer against balancing losses, handling losses, and system losses that emerge within the production system (Wild, 1975). Since indirect work, variability, and uneven workload distribution reduce the effective productive time available, the theoretical cycle time must remain below the takt time in order to sustain stable production flow over time. This factor affects how the

cycle time should be viewed, if the takt time is set to 22 minutes the cycle time should be set to 85% of takt time to make allowances for the sustainability factor. See equation 5.3 for an example calculation.

$$\text{Takt time 22 min} \implies 22 * 0,85 = 18,7 \text{ Cycle time} \tag{5.3}$$

5.2.2 Maximum capacity & staffing

The staffing of an assembly line varies significantly depending on the operational methodology and the organizational structure of the workforce. As previously detailed in section 2.2.3, this study focuses on a WWAL configuration, as this is the current operational model for assembly flows 1, 2, and 3.

In a WWAL, operators "chase" their specific motor through the entire assembly process. This setup allows the number of stations to be lower than the total number of operators. It also provides the flexibility to scale production volume simply by adding or removing operators from the flow. At the Vara plant, this method was specifically implemented to maximize flexibility.

As previously noted, the assembly and machining processes were evaluated separately because they operate under different conditions. The VSM and bottleneck analysis indicate that the line is currently not operating at full capacity. Consequently, there is an opportunity, if needed to increase engine output by adding more operators to the flow. The following paragraphs detail the maximum capacity thresholds for each process steps and the staffing structures required to support them. There are two sections of different calculation that shows the process, one with no balancing of the line and one where balancing has happened. The available time used as the basis for all calculations is detailed for each process step in table 5.2.

5.2.2.1 Current balance

The following paragraphs present the results of maintaining the current line configuration while only adding operators. It will present the capacity for each process step as well as the operators needed to complete all operations.

Machining

Currently the machining process runs at two shifts. The estimated performance of the new machining line working two shifts is 23 engines per 24 hours. Increasing to 3 shifts gives an estimated performance of 32 engines per 32 hours.

Flow 1

Flow 1 has a special setup, there is a total of 24 operating stations and 3 in line buffer stations where seven operators are currently working. Two of the stations are quality control where one person is stationed, all the other stations are in a WWAL configuration. The longest operation time is 8 minutes which includes Volvo's 114% speed and the TW trough the flow is 112,1 minutes. When calculating the maximum capacity without

balancing the operational strategy will be the same, with stationed operators on quality control and WWAL on all other stations.

Calculations with no balancing, equation 5.4 shows that the line is able to produce 42 engines on a regular day shift if they just add operators. The number of operators needed is 15 with two of them based in the quality control and the calculations can be seen in equation 5.5.

$$\text{F1 Capacity without balancing} = \frac{\text{AT} * 0,85}{\text{LOT}} = \frac{338,3}{8} = 42,29 \implies 42 \text{ Engines} \quad (5.4)$$

$$\text{Operators needed} = \text{Qual.Contr Ops} + \frac{L/T}{LOT} = 2 + \frac{102,57}{8} = 14,82 \implies 15 \text{ Ops} \quad (5.5)$$

Paint shop

The paint shop is made up of three manual stations, two painting robot cells and a drying tunnel. Contrary to the other processes in the assembly the majority of the process time in the paint shop is made up of machine time and is therefore not scalable with staffing which makes the capacity calculations different from the other processes in the assembly. The most time consuming operation of the paint-shop is the drying of the paint which takes 41 minutes per engine. However, it is essential to distinguish between the total processing time in the oven and the effective cycle time of the process. Because the drying oven is designed as a tunnel that accommodates multiple engines simultaneously, it operates as a continuous flow process.

The cycle-time is therefore determined by the frequency at which engines exit the tunnel rather than the total duration of the drying cycle for a single unit. Effectively the total drying time is divided by the amount of engines in the drying oven to calculate the cycle-time, however each engine does two laps in the painting process, first a base layer paint and then final paint, which results in only half of the engines that leave the oven is ready to go to the next process step. This all together results in a LOT of 10,41 minutes for the painting process which is a structural bottleneck that cannot be reduced by increasing the number operators. Using the available time for the paint shop in table 2.1 the capacity of the paint shop was calculated to be 34 engines a day with two operators. The calculations can be found in appendix C.2

Flow 2

Assembly flow 2 has a structure similar to flow 1, although with fewer workstations and operators. The current configuration consists of seven stations manned by five operators. In this setup, the first six stations operate according to a walking workers configuration, while the final station is dedicated to quality control with a stationary operator. The process is characterized by a LOT of 12 minutes and TW of 64,57 minutes.

As specified in section 5.2.1, the Available Time for flow 2 is 427 minutes, which is 29 minutes more than flow 1. Based on these values, the system reaches a theoretical capacity of 31 engines by increasing the workforce to six operators. In this expanded staffing

scenario, the operational methodology remains consistent with the current method, the additional resource is allocated to the line, increasing that pool to five operators while the final quality control station remains with one stationary operator. Calculations for flow 2 can be seen in appendix C.3.

Engine testing

Engine testing is a process step where balancing is not feasible, as the work content is already inherently balanced. The testing procedure is identical for each engine, with the only variables being the specific volumes of oil and coolant required for D4 and D6. Consequently, the calculations for both current and rebalanced scenarios remain the same. In the current production setup, three testing boxes are utilized, each manned by one operator. However, there are seven usable boxes available in total. By utilizing all seven boxes with seven dedicated operators, the engine testing stage is capable of producing 66 engines per day shift. Calculations for engine testing can be seen in appendix C.4

Flow 3

Assembly flow 3, which handles the final customer specification and packaging, operates similarly to flow 2 but on a smaller scale. The current configuration consists of four stations manned by two operators. In this setup, the operators work entirely in a WWAL configuration, following the engine through all finalization steps. The process is characterized by a LOT of 7,09 minute an a TW of 26,61 minutes.

As specified in section 5.2.1, the available time for the flow is 427 minutes. Based on these values, the system reaches a theoretical maximum capacity of 51 engines by increasing the workforce to four operators. Increasing the workforce more will not increase the capacity any longer as the process is constrained by the LOT at this staffing level and balancing is needed to increase capacity further. Calculations for flow 3 can be seen in appendix C.5

5.2.2.2 Capacity balanced flow

The following section evaluates the capacity of the FWAL and details the flow balancing along with the required operator allocation for each process step.

Machining

As the machining capacity is primarily determined by machine processing time, it is not significantly affected by the balancing of manual tasks. Since the machining line was not yet operational at the time of data collection, no empirical data was available regarding workload distribution between machines. Consequently, the machining capacity is assumed to remain at 32 units for the purpose of this analysis.

Flow 1

When capacity calculations with balancing was calculated the line theoretically switched to FWAL where every operator was stationed on a specific station. This way of working minimizes operator transportation between stations while also making workload imbalances and bottlenecks more visible, as each station operates against the defined takt time

with one dedicated operator. The results of the calculation can be see in equation 5.6 and C.2 which shows a capacity of 72 engines per day shift. The number of stations used will be the same as the number of operators and in this case it is chosen to use the same amount of stations as the real flow 1.

$$\text{F1 new LOT} = \frac{\text{TW}}{\text{Number of stations}} = \frac{112,1}{24} = 4,67 = \text{ LOT2} \tag{5.6}$$

$$\text{F1 Capacity after balancing} = \frac{\text{AT} * 0,85}{\text{LOT2}} = \frac{338,3}{4,67} = 72,43 \implies \text{72 Engines} \tag{5.7}$$

Paint shop

As previously mentioned the LOT in the paint shop is a structural bottleneck and is not effected by the number of operators and therefore balancing is not relevant in the paint shop and the capacity remains at 34 engines.

Flow 2

Similar to flow 1, the balancing of flow 2 was modeled by transitioning from walking-worker to a FWAL, where each operator is assigned to a specific workstation. This transition aims to minimize operator movement and enhance process visibility. However, a significant constraint for flow 2 is the limited number of available workstations while flow 1 comprises 24 stations, flow 2 is restricted to seven. Given the current staffing levels and this physical station constraint, the potential for capacity expansion in flow 2 is inherently more limited than in flow 1. The total capacity in flow 2 is 39 engines with seven operators on a regular day shift. See appendix C.3 for calculations.

Engine testing

As mentioned before is engine testing already balanced and will therefor have the capacity of 66 engines when operating seven testing boxes each manned by one operator. Calculations for engine testing can be seen in appendix C.4

Flow 3

Similar to previous assembly flows, the balancing of flow 3 was modeled after a FWAL, By dividing the total work by the amount of stations in the flow the theoretical shortest possible assembly cycle time is calculated. Similar to flow two, flow three has a structural constraint of only consisting of 8 stations, however, the total work content to be distributed across these stations is significantly lower, at only 26.61 minutes compared to 64.57 minutes in flow 2 which can be seen in table 4.3. This lower workload relative to the number of stations allows for a more efficient task distribution, resulting in flow 3 achieving the highest theoretical capacity in the system, reaching 109 engines per day shift after balancing. Calculations for flow 3 can be seen in appendix C.5

5.2.3 Summary capacity

Table 5.3 summarizes the current production capacity of the line, as well as the theoretical potential both with and without balancing. The red markings in the table identify the process steps with the lowest capacity, representing the primary system constraints.

Important to remember that all capacity and staffing values were based on working speed of 114% and a utilization of 85% of the available time. This to ensure a sustainable production over time.

Table 5.3: Staffing and capacity analysis for different process steps. Red text is bottlenecks.

Process step	Machining	Flow 1	Paint shop	Flow 2	Engine test	Flow 3
Current staffing	2 shifts	7	1	5	3	2
Capacity with current staffing	23	21	27	22	28	27
Add Staff - max capacity	3 shifts	15	2	6	7	4
Capacity	32	43	34	31	66	51
Add Staff - max capacity, balancing	3 shifts	24	2	7	7	8
Capacity with balancing	32	72	34	39	66	109

In the current manufacturing state, the total system capacity is constrained by flow 1 with a capacity of 21 engines per day, which is slightly higher than the current target of producing 18 engines. All other process steps possess higher capacities, which inherently results in operational losses, primarily in the form of waiting time.

If additional personnel were added to the current configuration without balancing, the total capacity would reach 31 engines per day shift, a limit determined by the constraints in flow 2. While flow 2 utilizes a similar WWAL setup to flow 1, it lacks the same ease of scalability. The limited number of workstations in flow 2, relative to flow 1, causes it to become the system-critical bottleneck. This progression is documented in table B.1, which illustrates the migration of the bottleneck as demand increases and the takt time is reduced.

If the production flow where balanced while maintaining the current number of stations, the machining would limit the capacity at 32 engines with the current product mix. However with the future product mix the paint shop would emerge as the primary bottleneck with a maximum capacity of 34 engines. This constraint is dictated by the required duration in the drying tunnel rather than manual assembly operations. As indicated by the bottleneck development analysis in table 5.3, the paint shop would operate near its maximum capacity and act as a secondary constraint once flow 2 reaches its limit. Consequently, these findings suggest that flow 2 is the only process step where a balancing of tasks would currently yield a significant improvement in potential system throughput without great investments. This is because the drying process in the paint shop is machine time, and would only improve if new investments or change in AT would take place.

5.2.3.1 Including system losses

As discussed in section 5.1.2, system losses are expected to have a greater impact at higher production volumes, since the increase in workforce reduces the internal buffer capacity in the WWAL configuration. This effect is particularly relevant for the maximum capacity scenarios with line balancing, where most assembly stations are assumed to be fully

staffed and the number of empty stations between operators is therefore reduced.

In table 5.4, the capacity calculations are based on a utilization rate of 85%, which accounts for operational losses and sustainable working conditions. However, this utilization rate only reflects the initial system loss assumption of approximately 5%, as discussed in section 5.1.2. In the maximum staffing scenario without balancing, the increased system loss primarily affects flow 2, where the reduced internal buffer capacity is estimated to increase system losses to approximately 10%. This reduces the capacity of flow 2 to 29 engines per day shift, as shown in Appendix C.9.

In the balanced capacity scenario, all assembly stations in flow 1, Flow 2, and flow 3 are assumed to be fully staffed. This largely eliminates the internal buffer capacity in these flows, making them more sensitive to blocking, starvation, and short-term disturbances. Consequently, system losses are estimated to increase to approximately 25% in all three assembly flows. The resulting capacity values are presented in Table 5.4.

The results show that including increased system losses has a significant effect on the capacity analysis. Flow 2 remains the limiting bottleneck even after balancing, since the increased system losses substantially reduce the theoretical capacity improvement achieved through additional staffing. This indicates that the balanced capacity scenario presented in table 5.4 should be interpreted as a theoretical upper limit, while Table 5.4 provides a more conservative estimate of achievable capacity when reduced buffer capacity and increased system losses are taken into account. The calculations used to derive the values presented in table 5.4 can be found in Appendix C.

Table 5.4: Staffing and capacity analysis with system loss. Red text is bottlenecks.

Process step	Machining	Flow 1	Paint shop	Flow 2	Engine test	Flow 3
Add Staff - max capacity, balancing 25% system loss	3 shifts	24	2	7	7	8
Capacity with balancing	32	57	34	31	66	86

5.3 Effective staffing scalability

This sections aims to answer research question three and presents an analysis of staffing compared to demand level and the utilization of resources. In figure 5.2, 5.4, 5.6 is analysis of how staffing could be structured compared to different demand levels. The desired utilization as mentioned before was 85% and when the operators was burdened over 85% one operator was added to the flow to reduce the utilization. The daily volume scale (x-axis) starts at 14 engines per day, a figure derived from a safety margin applied to the previous historical low of 16 engines. The scale extends to 34 engines per day, representing the system's maximum theoretical capacity following line balancing.

5.3.1 Flow 1

As seen in the figure 5.2 the utilization was at the lowest when the production is at 14 engines per day shift and at the desired utilization 85% when the production is at 23

respective 33 engines. But as mentioned before in section 5.2.3 flow 2 is the bottleneck of the production without balancing and the maximum capacity in the line was 31 engines per day shift. At this volume flow 1 has a utilization of 80%.

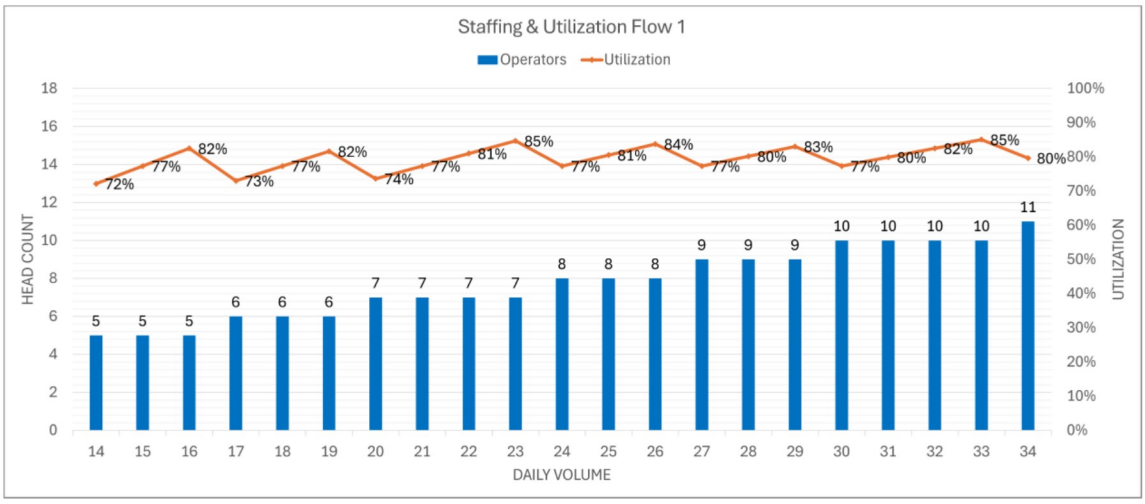

Figure 5.2: Utilization and staffing for flow 1

5.3.2 Paint shop

The paint shop reaches its lowest utilization in the analyzed range at production volumes of 14 and 28 engines per day shift, with a utilization of 44%. The highest utilization, 84%, is achieved at a volume of 27 engines per day shift. As shown in figure 5.3, there is a substantial drop in utilization between 27 and 28 engines. This drop occurs because the paint shop has relatively few operators; adding one extra operator to enable production of 28 engines nearly doubles the process capacity of the manual stations, while the required output increases by only one engine.

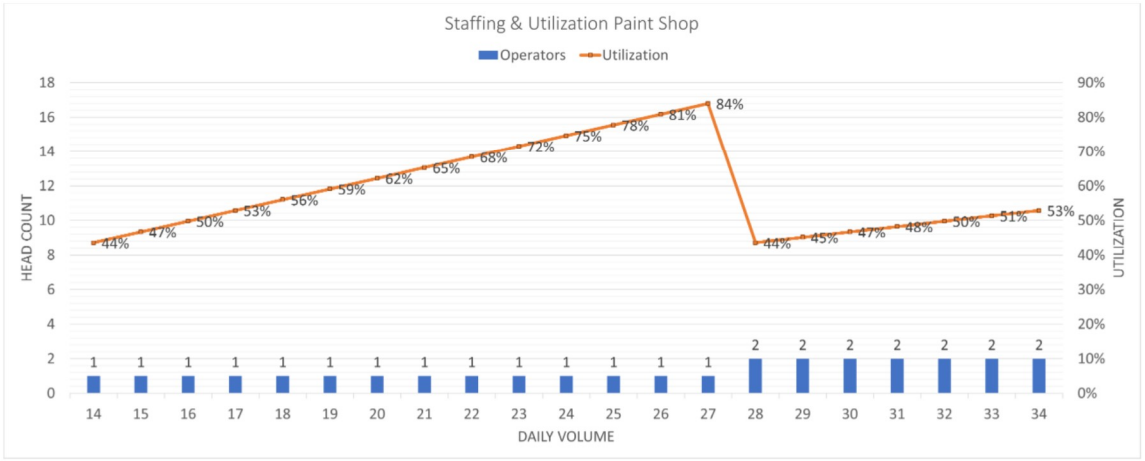

Figure 5.3: Utilization and staffing for paint shop

5.3.3 Flow 2

At flow 2 the lowest utilization within the studied interval is observed at a daily volume of 17 engines see figure 5.4, where resource efficiency drops to 64% due to the introduction of a fourth operator into the flow. Similar to flow 1, flow 2 achieves its most efficient resource utilization (83% and 85%) just before the staffing level must be increased, which occurs at volumes of 22 and 28 engines per shift. As previously established in section 5.2.3, flow 2 constitutes the primary bottleneck of the system in its current configuration without balancing. Consequently, the scalability of this flow was maximized at 31 engines per day, where utilization reaches 78% with a staffing level of six operators. At this stage, further volume increases are constrained by the physical design of the stations and the LOT, rather than a lack of personnel capacity.

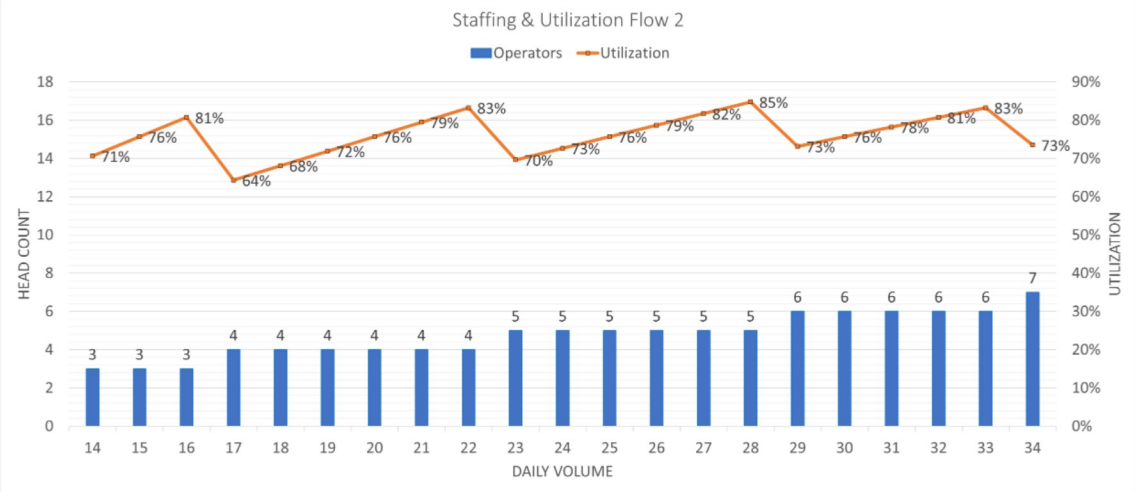

Figure 5.4: Utilization and staffing for flow 2

5.3.4 Engine testing

As seen in figure 5.5 the utilization was at its lowest when production is at 14 engines per day shift and reaches the desired utilization of 85% when the production is at 28 engines. But as mentioned before in section 5.2.3, flow 2 is the bottleneck of the production without balancing and the maximum capacity in the line was 31 engines per day shift. At this volume, engine testing has a utilization of approximately 69% when utilizing four operators and four boxes.

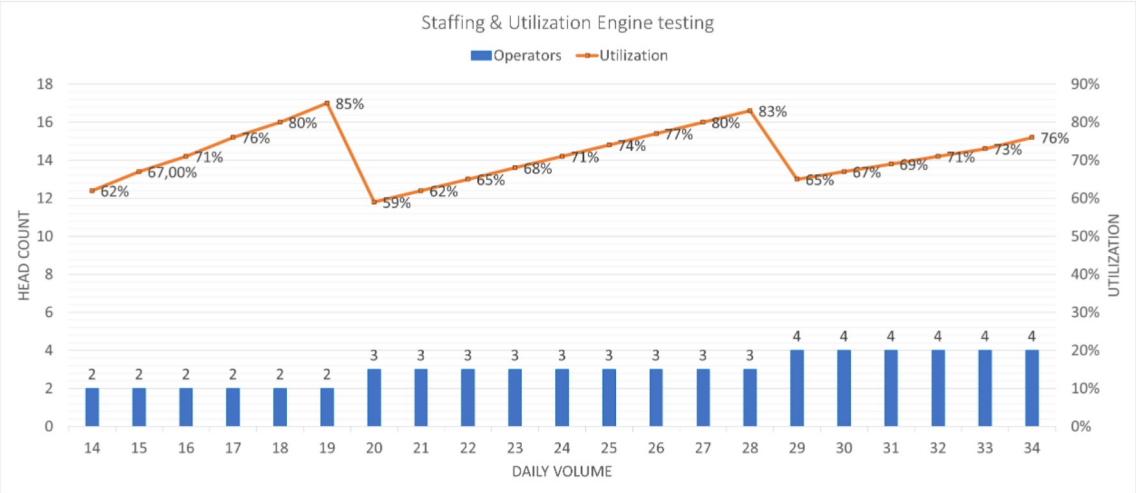

Figure 5.5: Utilization and staffing for Engine testing

5.3.5 Flow 3

Figure 5.6 presents the staffing and utilization analysis for flow 3 which demonstrates a high degree of flexibility with minimal staffing. Utilization rises from a low of 43% at 14 engines per day to 83% at 27 engines, at which point the flow was manned by two operators. The requirement for a third operator at a volume of 28 engines led to a significant reduction in utilization to 58%. This fluctuation was a result of each staffing adjustment representing a substantial percentage of the total workforce in such a small-scale flow. Even at a volume of 34 engines, flow 3 maintained a utilization of only 70% with three operators, indicating that it possesses a high theoretical capacity.

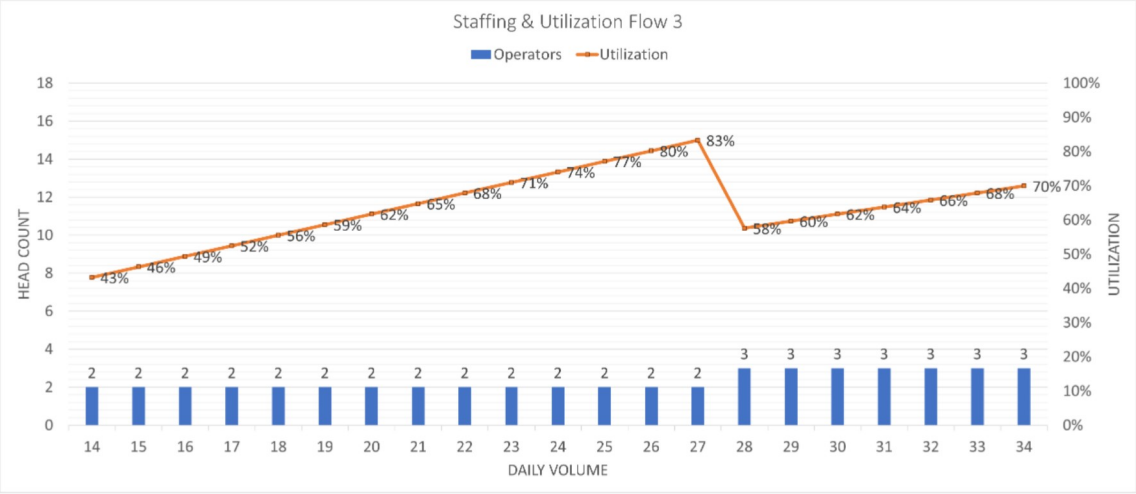

Figure 5.6: Utilization and staffing for flow 3

5.3.6 Summary staffing vs demand

The production system exhibits a distinct "sawtooth" pattern in utilization as volumes increase. This occurs because operators cannot be added fractionally, when demand exceeds the capacity of the current workforce, an additional full-time operator must be

introduced, resulting in temporary overcapacity at that production level. As illustrated in figure 5.2, the spread in utilization decreases as daily volume increases. This is because the relative capacity increase contributed by adding one operator diminishes as the total number of operators in the line grows. Consequently, scaling the production line becomes more efficient at higher volumes. This effect is further emphasized when comparing different processes: those with fewer operators, such as flow 3 and the paint shop, exhibit more pronounced drops in utilization at threshold points than flow 1 and flow 2.

6
Discussion

This chapter discusses the findings in the results in relation to the theoretical framework and research questions. It analyzes bottleneck propagation, workforce scalability, and the operational trade-offs between flexibility and stability. Furthermore, the methodology is critically evaluated, followed by a presentation of the study's practical contributions, generalization, and areas for future research.

6.1 Bottleneck propagation

The dynamic movement of the bottleneck under increasing demand highlights a distinct transition from resource-based constraints to structural system bottlenecks. Initially, the system exhibits a high level of volume flexibility, as capacity is directly linked to the number of operators in the WWAL configuration. Within this flexible range, emerging bottlenecks can be mitigated simply by scaling the workforce. However, once the required takt time drops below the LOT of flow 2, this flexibility is lost. The constraint shifts from workforce availability to fixed processing time. According to TOC , introducing more operators at this stage will not improve throughput but rather result in increased system losses in the form of queuing. Because the constraint in flow 2 is based on manual operations, this specific structural bottleneck can be addressed using line balancing principles.

In contrast, the paint shop presents a fundamentally different type of structural constraint. While its LOT does not trigger the primary bottleneck in the initial analysis, it is dictated by machine time (specifically the drying process) rather than manual labor. Consequently, line balancing would have no effect on this process. If flow 2 is successfully balanced, the paint shop will inevitably emerge as the system's next limiting factor. Potential solutions to increase paint shop capacity include physically extending the drying tunnel to accommodate more engines, or introducing buffers before and after the process to decouple it from the main flow. However, both alternatives present a strategic trade-off, as they would inherently increase WIP and total lead time within the production line.

6.2 Line balancing

The results from table 5.3 show that line balancing would increase the capacity drastically for all assembly processes. As flow 2 was concluded to be the system bottleneck that limits scaling with additional staff, balancing of this process should be prioritized. Furthermore, the results show that balancing the other assembly processes becomes less relevant in the near future, as their capacity before balancing is still greater than a perfectly balanced flow 2 and other capacity constraints, like the paint shop and the machining process.

This shows that the difference in capacity between the processes is not just a result of the processes being unbalanced internally, but also unbalanced in relation to each other. This indicates that reaching the maximum capacity of each assembly process is impossible without significant productivity improvements in the other processes.

An aspect that was not assessed in this study is balancing by moving process steps between different processes. In the current configuration, this is impossible, as the paint shop and engine testing processes require certain components to be assembled to function. The line balancing analysis conducted in this study was purely time-based, which is a simplification of reality. When moving process steps between stations, additional material handling or unnecessary movement can be introduced, which is not included in the analysis and should be taken into consideration when interpreting the results.

6.3 Capacity and Workforce

The findings of this study demonstrate that the D4/D6 production line possesses a higher capacity than the current demand of 18 engines per day. To accommodate future demand increases, Volvo Penta can employ three primary strategies, adjusting staffing levels, increasing available production time, and rebalancing the line (Jamalnia et al., 2019; Liker, 2021; Womack & Jones, 2003).

6.3.1 Constraints with scaling workforce

Adjusting workforce levels is a direct and efficient method for increasing capacity, particularly in assembly flows 1, 2, and 3. In these three process steps, which utilize a WWAL configuration, personnel can be added without disrupting the current setup or station times. This flexibility in the WWAL allows for scaling, but only until structural limits are reached, as demonstrated by the transition to a structural bottleneck in flow 2. However, any strategic decision to increase workforce must address two critical factors.

The first consideration is the decrease in marginal utilization of additional labor. As detailed in section 5.3.6, the relative capacity gain per added operator decreases as the total workforce on the line grows. In small flow such as flow 3, the introduction of a single operator causes a significant drop in overall utilization. Conversely, in larger processes like flow 1, the impact on workforce utilization is less pronounced. This creates a non-linear pattern in resource efficiency, which complicates workforce planning, particularly at lower production volumes and within smaller process steps.

The second critical factor to address when discussing adding staff is the LOT. As seen first i flow 2 in scenario 6 in section 5.1 the LOT becomes a bottleneck when one extra operator was added. When the headcount results in a calculated cycle time lower than the LOT, adding operators will no longer increase throughput. At this point, the process shifts from being constrained by workforce to being structurally constrained by LOT.

6.3.2 Flexibility vs Stability

The utilization of WWAL as a primary assembly strategy presents a strategic trade-off between volume flexibility and operational stability. On the positive side, this method

allows the system to scale production volumes by simply adjusting the workforce size, provide more diverse task for operators, and maintains a decoupling between the number of physical stations and the number of operators. However, from a lean perspective, this strategy introduces significant challenges that can hinder long-term efficiency.

One conflict exists between the chase strategy and the lean principle of Heijunka (production leveling). While the chase strategy aims to go with customer demand levels, lean thinking often emphasize stability in the process to eliminate waste (Liker, 2021). By constantly shifting the workforce size, the plant risk introducing instability into the standardized work, which according to Liker (2021) is the very foundation of continuous improvement.

The chase strategy has one operational hurdle which is the long training times for new operators especially when the flows are long and complex, as flow 1. It becomes complex for first line leader to scale up and down in workforce when the required competence takes months to learn and the learning period can possibly be a bottleneck for the line. In section 5.3, the results shows when production requires adding or subtracting an operator but in the real world is it not that easy to integrate an operator possessing the required technical expertise. Effective implementation of a chase strategy requires workforce adaptability, without it, the strategy's ability is threatened by the operational bottlenecks from in lengthy training durations.

When subtracting an operator their is not just less payout and better utilization in the flow, the company loses knowledge that the operator gathered during their period at the company. The company needs to evaluate and do trade-off between keeping operators and knowledge or train new operators when demand varies. A strategy that takes these challenges and losses in consideration is therefore needed. Consequently, the strategy must account for both current and projected demand within the framework of workforce planning.(Jamalnia et al., 2019).

Within this strategy, operational decisions cannot be guided solely by demand fluctuations and technical production constraints. Social and ethical dimensions must be fully integrated, ensuring that Volvo Penta is perceived as a secure employer that provides an excellent working environment. Consequently, cycle times, ergonomic factors, and the overall physical and mental demands placed on operators must be maintained at a long-term sustainable level.

6.4 Social, ethical, and environmental considerations

The results of this study is not only relevant from a technical capacity perspective, but also have societal, ethical, and environmental implications. From a societal perspective, improved understanding of production capacity can support more stable and informed workforce planning. inaccurate capacity assumptions may lead to either unnecessary layoffs during lower demand periods or excessive workload during demand peaks. By identifying capacity limitations and bottleneck development more clearly, the study can support decisions that balance production flexibility with workforce stability.

From an ethical perspective, the analysis highlights the importance of not treating staffing

only as a numerical capacity variable. Staffing decisions affect operators' workload, training requirements, competence development, and perceived work stability. If capacity is overestimated, operators may be exposed to sustained workload pressure during demand peaks. If capacity is underestimated, the organization may result in premature workforce reductions. Therefore, decisions based on the capacity analysis should also consider the work environment, long-term competence retention, and the consequences for employees affected by staffing changes.

The environmental aspects of the study are more indirect since this study does not include an in depth analysis of the environmental impact of the production flow. However an improved capacity understanding can still contribute to more sustainable resource utilization. By identifying when capacity can be increased through better staffing structures or line balancing, unnecessary investments, overcapacity, and inefficient use of production resources may be avoided. At the same time, the study should not be interpreted as an environmental evaluation of the products themselves, since the analysis is limited to the production flow, capacity, and scalability within the studied system.

6.5 Difference between machining and assembly

Something observed during this study is the fundamental difference between the machining and assembly processes within the production flow. The machining process is highly automated, meaning that its cycle time is primarily determined by machine processing time rather than staffing levels (Slack et al., 2019). In contrast, assembly is predominantly manual and therefore strongly influenced by staffing and task distribution. Consequently, machining capacity is mainly scaled through increased operating time, such as additional shifts, while assembly capacity can be adjusted more incrementally through staffing changes. In accordance with APP principles, this makes machining more suitable for a level strategy, while assembly is better suited for a chase strategy (Jamalnia et al., 2019).

The storage between machining and assembly decouples the machining process from the production pace of the remaining flow (Rother & Shook, 1999). This is necessary due to the significantly longer cycle times in machining, enabling machining to compensate for the output consumed during assembly operations through multiple shifts. However, this also limits the ability to expand assembly capacity through additional shifts, as the machining process may then emerge as the system bottleneck (Rahman, 1998). The decoupling further reflects a difference in production control principles, where machining follows a push-based approach producing against stock, while assembly is more closely aligned with a pull-based system governed by takt time and customer demand (Rother & Shook, 1999).

Within this strategy, operational decisions cannot be guided solely by demand fluctuations and technical production constraints. Social and ethical dimensions must also be considered to ensure long-term workforce sustainability and a healthy working environment. Consequently, cycle times, ergonomic conditions, and the overall physical and mental workload placed on operators should be maintained at a sustainable level over time

6.6 Method discussion

This section will discuss the data collecting method, the chosen lean principles, the suitability for research questions and the limitations with this methods.

6.6.1 Data collecting method

While SAM data established a foundational understanding of the production flow, the Gemba walks and interviews were essential for capturing the whole picture. These observations revealed critical operational details, such as the strategic use of buffers to minimize operator idle time, that raw data could not deliver. The expert panels then served as a decisive validation tool, where analyzed times and procedures were refined through consensus with technicians. This rigorous triangulation ensured a stable foundation for the VSM, providing the credible results needed to address the research questions.

6.6.2 Analytical approach

The integration of VSM and TOC was a strategic choice to bridge the gap between system visualization and actionable prioritization. VSM provides the broad overview and the "as-is" state of the production and information flow and TOC was chosen as a filter that identifies which specific wastes actually limited the system´s throughput (Librelato et al., 2014).

This approach was chosen over Discrete event simulation (DES) due to its operational accessibility and transparency. Unlike DES models, which often become one-shot tools that are difficult to maintain without external expertise (Mourtzis, 2020), VSM provides a transparent and visually intuitive representation of the production flow, facilitating communication and process understanding throughout the organization (Rother & Shook, 1999). This approach aligned with the expectations set during our first meeting with Volvo Penta Vara, where they expressed the need for a method that could be maintained internally by their own team. DES had previously been tested at the plant, but limited internal resources and simulation expertise made the models difficult to maintain and therefore less useful in practice for operational management.

However, the chosen analytical approach also has limitations. The deterministic nature of the model does not fully capture dynamic variation within the system. In practice, capacity may be affected by operator-to-operator differences, product mix variation, material disturbances, short stops, and other sources of variability that influence blocking, starvation, and system losses. An alternative approach such as discrete event simulation could have captured these dynamic effects more explicitly which would likely provide a more detailed estimate of achievable capacity, particularly at higher staffing levels where internal buffer capacity is reduced. However, such a method would require more detailed and validated input data regarding variation, disturbance frequencies, and process-time distributions which was not available at the time of the study. Given the purpose and scope of this study, the selected VSM, TOC, and line-balancing approach was considered appropriate for creating a transparent capacity model and identifying the main structural constraints, while the lack of dynamic variation modelling remains an important limitation.

6.6.3 Suitability for research questions

The method with VSM and TOC are allowing for a scenario-based analysis where the takt time was incrementally reducing takt time to see where bottlenecks occur and when the capacity for the production line reaches its maximum (Librelato et al., 2014). The incrementally reduction has also allowed for a visualization of how the bottleneck moves between process steps and when it moves from resource-based to a structural bottleneck.

While the analysis uses constant times, the method provided a baseline for understanding how staffing levels affect workforce utilization. The visualization of staffing levels and their utilization creates a vital theoretical understanding when operators should be added or subtracted and how they affect utilization in process step, but as previously mentioned in section 6.3.2 these choices must be carefully considered. This approach provided a data-driven foundation for evaluating how the D4/D6 line can scale effectively, moving beyond a nominal capacity assumptions to a grounded understanding of the system´s actual performance limits.

6.6.4 Linearity of losses

In the capacity analysis, a sustainability factor of 85% was used to account for operational losses within the production system. As this factor remained constant throughout the analysis, the resulting scalability exhibited a relatively linear relationship between additional staffing and reduced cycle times. This assumption was considered reasonable due to the comparatively long cycle times of the studied production system in relation to typical series production systems. According to the data presented by Wild (1975), balancing losses and handling losses remain relatively stable at higher cycle times, reducing their relative impact on scalability within the analyzed production range.

However, system losses caused by variability, blocking, and starvation between interconnected processes were not explicitly modeled in the analysis and may therefore influence the achievable capacity in practice as production volumes increase. According to Wild (1975), system losses are influenced by both the number of stations and the size of buffers within the production system. In the present analysis, the number of stations remained constant and were therefore assumed to have a limited influence on changes in system losses throughout the analysis.

However, the empty stations between operators in the chasing assembly lines effectively act as buffers capable of absorbing short-term disturbances within the production flow. As additional assemblers are introduced into the assembly lines, these buffers decrease or may disappear entirely, potentially increasing the impact of system losses. This may become particularly relevant at higher staffing levels where internal flexibility within the production flow decreases. An example observed in this study is flow 2 approaching full staffing at higher production volumes, where fewer internal gaps remain between operators and stations. Although system losses were not explicitly measured, reduced internal buffers may increase the sensitivity of the flow to disturbances and short-term variability, potentially reducing the practically achievable capacity compared to the theoretical model.

Although system losses were not empirically measured in this study, they were included as a sensitivity adjustment in the capacity analysis. This showed that reduced internal buffer

capacity may substantially reduce the practical capacity improvement achieved through line balancing. Therefore, the balanced capacity scenario should be interpreted as a theoretical upper limit, while the adjusted system-loss scenario provides a more conservative estimate of achievable capacity. This highlights that the scalability of the WWAL configuration becomes less linear as staffing levels increase and internal buffer capacity is reduced.

Consequently, the calculated capacity presented in this study may overestimate the achievable capacity at higher staffing levels. In such cases, further analysis using Discrete Event Simulation may be appropriate, as this method is better suited for capturing dynamic interactions, disturbances, and system-level losses within complex production flows (Mourtzis, 2020). Despite this limitation, the model remains useful for identifying structural bottlenecks and evaluating relative capacity changes between staffing configurations.

6.6.5 Limitations

A limitation of the chosen product mix is that low-volume engines were excluded, which may obscure constraints caused by high product complexity. These low-volume variants potentially possess longer operation times that could negatively impact the overall capacity and flow of the system. The decision regarding which engines to include in the analysis was made based one month production data, together with following discussions with operators and technicians, leading to the conclusion that focusing on the most frequent engine variants was the most viable approach. According to the operators, almost all engine variants share similar production times, furthermore, analyzing the entire range of variants would have been beyond the feasible scope and time frame of this study.

In the VSM and TOC, the study was assuming stable conditions regarding external deliveries of materials to the flow. This could possibly affect the production capacity but the chosen method is not able to handle these changes in deliveries (Rother & Shook, 1999).

6.7 Future research

While this study provides a data-driven analysis of bottlenecks, capacity limitations, and staffing scalability within the D4/D6 production flow, several areas remain outside the scope of the study. During the analysis, a number of additional factors and relationships were identified that may influence long-term production performance and scalability. The following sections therefore present areas where further research could contribute to a deeper understanding of production flow behavior, workforce flexibility, and system constraints in scalable manufacturing systems.

6.7.1 Increasing available production time

An alternative to hiring personnel is to maximize the utilization of existing resources by increasing the available time (Womack & Jones, 2003). Available time determines the total duration a process remains operational, thus, increasing this window enables the entire production line to increase output using the current resource base. By analyzing the disparity between total paid time and available time, capacity can be recovered without the need for capital-intensive investments in new machinery or the addition of extra shifts. Potential approaches include optimizing of scheduled breaks, cleaning routines,

and meeting structures. For the paint shop, where automated cycle times constitute a bottleneck before manual labor limits are reached, increasing available time is the primary method to scale output without substantial investment.

6.7.2 Cross-training

One area for future research is the impact of cross-training operators between different assembly processes. Since the results indicate that staffing flexibility plays a major role in managing bottlenecks and scaling production capacity, increased operator flexibility may improve the system's ability to adapt to changing demand conditions (Hopp et al., 2004). Future studies could therefore investigate how cross-training affects bottleneck propagation, workload balancing, and production scalability in practice. Additionally, cross-training may also influence the training time required for operators, creating a trade-off between broader operator competence and the time needed to develop such competence. Future research could therefore provide further insight into the balance between workforce flexibility, training time, and operational efficiency.

6.7.3 Material handling and sub flows

The production flow visualized in the VSM shown in figure 4.2 was simplified in order to highlight the key processes within the production system. In practice, the main flow is supported by several sub flows, such as material handling and pre-assembly operations. Including all supporting processes in the value stream map would likely have increased the complexity of the visualization and reduced its effectiveness as an analytical tool for identifying system-critical bottlenecks and flow imbalances (Rother & Shook, 1999).

During the Gemba walks, interviews, and analysis of the collected data, no indications were found that these supporting processes currently act as system-critical bottlenecks. However, these sub flows remain essential for maintaining the functionality of the main production flow. As production volumes increase, the load on supporting operations is also expected to increase, which may introduce new constraints outside the primary assembly processes. Future research could therefore expand the analysis to include supporting flows and internal logistics in order to evaluate how they influence long-term production scalability and overall system performance.

6.7.4 Machining data and product mix

As mentioned in section 1.5, the lack of data from the new machining facility limited the scope of this study. Consequently, a consideration for future research is to further analyze the machining facility once it is operational and more data becomes available. This would provide a more holistic view of the entire production flow and enhance the overall relevance of the study. Additionally, the planned new product mix will have an impact on capacity requirements. Since it is still uncertain how significant this impact will be and how it affects the balance between the machining and downstream processes, this presents a suitable area for future research.

6.7.5 Process-time improvements

This study have relied on staffing adjustments and available time to investigate as the scaling factors for capacity. However, the processing time for individual operations, which were based on SAM analyses, were assumed to be fixed parameters. A suitable area for future research would be to investigate actual process-time improvements. This becomes increasingly relevant as the capacity analysis pinpoints the paint shop to be the limiting factor for further scaling. As the paint shop is in the majority automated, reducing the process time may be a effective way to decrease the cycle time of the process without increasing WIP. Additionally for the manual work stations applying continuous improvement methods, such as ergonomic optimizations, better tooling, or the automation of specific manual tasks, the actual work content and the LOT could be reduced.

6.8 Practical Contributions

This study has provided an empirical analyze and results that will be useful for Volvo Penta Vara. The results has provided the factory with a foundation that visualizes the factory´s capacity, where bottlenecks occur with different demand and how they can work with workforce scaling. By integrating SAM analyses with VSM, TOC, line balancing, and utilization metrics, the organization can gain a validated understanding of the actual production flow and its inherent constraints.

In addition, the discussion highlights critical areas for future implementation at Volvo Penta, specifically the development of a structured workforce adjustment strategy and the formalization of data collection regarding material shortages and machine breakdowns.

Furthermore, the results supports the factory´s journey towards achieving VPS gold by utilizing lean tools to visualize imbalances and losses within the production flow. Which previously where embedded within daily operations.

6.9 Generalization

While the specific results of this study, such as the identification of flow 2 as the critical system bottleneck, are unique to the D4/D6 engine production at Volvo Penta, the underlying methodologies and insights possess a high degree of external relevance. Through analytical generalization (Voss et al., 2002), the approach utilized in this study can be effectively transferred to other manufacturing environments. The study successfully demonstrates how established tools such as Value Stream Mapping and the Theory of Constraints can be integrated to quantify scalability in environments characterized by fluctuating demand.

6.9.1 Comparison with alternative methods

The selection of an empirical analysis as an alternative to complex simulation models, such as DES, is a central component of the study's external relevance. Previous research indicates that DES models are frequently reduced to isolated one-shot projects that lack long-term integration into daily operational routines (Mourtzis, 2020). Furthermore, such simulation models are often highly resource-intensive to maintain and update, requiring

both specialized software licenses and extensive datasets (Robinson, 2014).

In contrast to these systems, the methodology presented in this study prioritizes operational accessibility. By utilizing VSM visualization, which facilitates transparent communication across various levels within the organization, from shop-floor operators to production management (Rother & Shook, 1999). For other organizations operating in high-variability environments, this implies that the framework offers a pragmatic and accessible approach for identifying system constraints without the need for specialized expertise or the extensive resource requirements associated with advanced simulation software

6.9.2 Transferability to similar manufacturing contexts

The analytical insights regarding the interaction between system constraints, resource utilization, and the system's maximum capacity are directly applicable to other manual or semi-automated assembly flows (Mortada & Soulhi, 2023). The study's conclusions regarding how a WWAL strategy provides volume flexibility, yet is limited by the diminishing marginal utility of additional labor during upscaling, offer valuable theoretical guidance for any operation employing similar workforce strategies. Ultimately, this research contributes to a broader understanding of how production scalability can be navigated when human resources and structural constraints interact.

7

Conclusion

This thesis aimed to analyze the D4/D6 engine production flow to identify system-critical bottlenecks, quantify actual and maximum production capacity, and evaluate how workforce and line balancing influence production scalability. By integrating VSM, TOC, and line balancing principles, the study provides an understanding of the production system's underlying constraints and scalability limitations.

Under the current demand of 18 engines per day, the production system operates with excess capacity and no active bottlenecks. However, as demand increases, the bottleneck propagates dynamically through the system. Initially, assembly flow 1 acts as the primary bottleneck, which can temporarily be mitigated through increased staffing. At higher demand levels, the system-critical constraint shifts to flow 2, where the LOT of 11.67 minutes creates a strict structural bottleneck. In addition, the paint shop acts as a secondary structural constraint due to its fixed automated drying cycle.

The analysis shows that the current production configuration has a maximum capacity of 31 engines per day without balancing, or 29 engines per day when increased system losses in Flow 2 are considered. With theoretical line balancing and a transition from a WWAL strategy to FWAL, the capacity of the assembly flows could be significantly increased. However, when increased system losses caused by reduced internal buffer capacity are included, the practical capacity improvement becomes lower than the theoretical balancing results suggest. The balanced capacity scenario should therefore be interpreted as a theoretical upper limit, while the adjusted system-loss scenario provides a more conservative estimate of achievable capacity of 29 engines. The total system capacity would still be limited by the structural constraints in Flow 2 and the paint shop.

The current WWAL strategy provides flexibility by enabling production to scale incrementally through staffing adjustments. However, the analysis demonstrates that workforce scalability is non-linear, as the marginal effect of adding personnel decreases with increasing staffing levels. Once the calculated cycle time reaches the LOT of a process, additional staffing no longer increases throughput. Consequently, Volvo Penta requires a structured workforce adjustment strategy.

Ultimately, the study demonstrates that while flexible staffing supports short-term volume adaptability, long-term scalability cannot be achieved through staffing adjustments alone. As production volumes increase, the system transitions from being constrained by workforce availability to being limited by structural bottlenecks. Future scalability efforts must therefore focus on the system's fixed constraints, particularly the LOT in flow 2 and the capacity limitations of the automated paint shop.

To further understand how to effectively scale production in the long run, future research can look at the bigger picture. This includes evaluating the gains from cross-training

staff and seeing how supporting flows affect the line when demand increases. Additionally, integrating real machining data and analyzing how a new production mix will affect flexibility could be highly useful for maintaining high efficiency.

References

Bell, E., Bryman, A., & Harley, B. (2022). *Business research methods* (6th ed.). Oxford University Press.

Cannas, V. G., Pero, M., Pozzi, R., & Rossi, T. (2018). Complexity reduction and kaizen events to balance manual assembly lines: An application in the field. *International Journal of Production Research*, *56*(11), 3914–3931. https://doi.org/10.1080/00207543.2018.1427898

Catalano, F., Zennaro, I., Berti, N., & Persona, A. (2025). Comparing fixed and walking worker strategies: Design implications of individual worker efficiency on assembly line performance. *International Journal of Production Research*, *63*(23), 9089–9111. https://doi.org/10.1080/00207543.2025.2533520

Gan, Z. L., Musa, S. N., & Yap, H. J. (2023). A review of the high-mix, low-volume manufacturing industry. *Applied Sciences*, *13*(3), 1687. https://doi.org/10.3390/app13031687

Goldratt, E. M. (1990). *The haystack syndrome: Sifting information out of the data ocean*. North River Press.

Holweg, M., Davies, J., de Meyer, A., Lawson, B., & Schmenner, R. W. (2018). *Process theory: The principles of operations management*. Oxford University Press.

Hopp, W. J., Tekin, E., & Van Oyen, M. P. (2004). Benefits of skill chaining in serial production lines with cross-trained workers. *Management Science*, *50*(1), 83–98.

Jamalnia, A., Yang, J.-B., Xu, D.-L., Feili, A., & Jamali, G. (2019). Evaluating the performance of aggregate production planning strategies under uncertainty in soft drink industry. *Journal of Manufacturing Systems*, *50*, 146–162. https://doi.org/10.1016/j.jmsy.2018.12.009

Kanban Zone. (2020). *Yamazumi chart: Why use it*. Retrieved March 26, 2026, from https://kanbanzone.com/2020/yamazumi-chart-why-use-it/

Librelato, T. P., Lacerda, D. P., Rodrigues, L. H., & Veit, D. R. (2014). A process improvement approach based on the value stream mapping and the theory of constraints thinking process. *Business Process Management Journal*, *20*(6), 922–949.

Liker, J. K. (2021). *The toyota way: 14 management principles from the world's greatest manufacturer* (2nd) [Accessed via AccessEngineering Library]. McGraw Hill. https://www.accessengineeringlibrary.com/content/book/9781260468519

Liker, J. K., & Convis, G. L. (2011). *The toyota way to lean leadership: Achieving and sustaining excellence through leadership development*. McGraw-Hill Education.

Mortada, A., & Soulhi, A. (2023). Improvement of assembly line efficiency by using lean manufacturing tools and line balancing techniques. *Advances in Science and Technology Research Journal*, *17*(4), 89–109. https://doi.org/10.12913/22998624/169257

Mourtzis, D. (2020). Simulation in the design and operation of manufacturing systems: State of the art and new trends. *International Journal of Production Research*, *58*(7), 1927–1949. https://doi.org/10.1080/00207543.2019.1636321

MTM Föreningen Norden. (n.d.). *Sam-analys*. MTM Föreningen Norden. Retrieved March 26, 2026, from https://mtmnorden.com/vara-kurser/sam-analys/

Nam, S.-J., & Logendran, R. (1992). Aggregate production planning—a survey of models and methodologies. *European Journal of Operational Research*, *61*(3), 255–272. https://doi.org/10.1016/0377-2217(92)90356-E

Ohno, T. (1988). *Toyota production system: Beyond large-scale production* [Original work published 1978]. Productivity Press.

Pakdil, F., & Leonard, K. M. (2017). Implementing and sustaining lean processes: The dilemma of societal culture effects. *International Journal of Production Research*, *55*(3), 700–717. https://doi.org/10.1080/00207543.2016.1200231

Patel, R., & Davidson, B. (2019). *Forskningsmetodikens grunder: Att planera, genomföra och rapportera en undersökning* (5th ed.). Studentlitteratur.

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.

Robinson, S. (2014). *Simulation: The practice of model development and use* (2nd ed.). Palgrave Macmillan.

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker: An account of a research program conducted by the western electric company, hawthorne works, chicago*. Harvard University Press.

Rother, M., & Shook, J. (1999). *Learning to see: Value stream mapping to add value and eliminate muda*. Lean Enterprise Institute.

Slack, N., Brandon-Jones, A., & Burgess, N. (2019). *Operations management* (9th ed.). Pearson Education.

Sun, T. (2024). Excess capacity and demand-driven business cycles. *The Review of Economic Studies*, *92*(4), 2730–2764. https://doi.org/10.1093/restud/rdae072

Upton, D. M. (1994). The management of manufacturing flexibility. *California Management Review*, *36*(2), 72–89.

Voss, C., Tsikriktsis, N., & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management*, *22*(2), 195–219.

Wild, R. (1975). On the selection of mass production systems. *International Journal of Production Research*, *13*(5), 443–461.

Womack, J. P., & Jones, D. T. (2003). *Lean thinking: Banish waste and create wealth in your corporation* (Revised and updated edition). Free Press.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

A

Appendix A

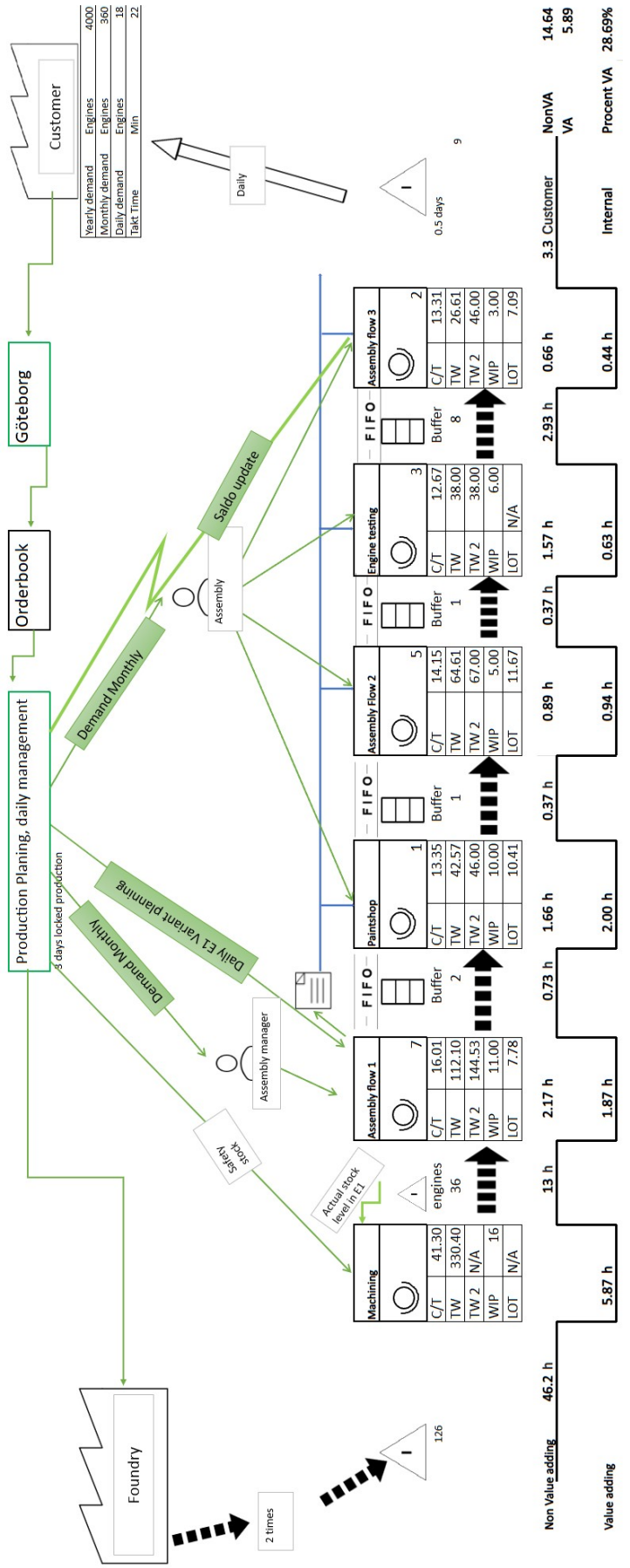

Figure A.1: Current state value stream map

B

Appendix B

Table B.1: Bottleneck development

Process data	Flow 1	Paintshop	Flow 2	Engine test	Flow 3
Scenario 1					
Staff [Pers]	8	1	5	3	2
C/T [min]	14,01	13,35	14,14	12,67	13,31
LOT [min]	7,78	10,41	11,67	N/A	7,09
Scenario 2					
Staff [Pers]	8	2	6	3	2
C/T [min]	14,01	13,35	11,67	12,67	13,31
LOT [min]	7,78	10,41	11,67	N/A	7,09
Scenario 3					
Staff [Pers]	9	1	6	3	2
C/T [min]	12,46	13,35	11,67	12,67	13,31
LOT [min]	7,78	10,41	11,67	N/A	7,09
Scenario 4					
Staff [Pers]	9	2	6	3	3
C/T [min]	12,46	10,41	11,67	12,67	8,87
LOT [min]	7,78	10,41	11,67	N/A	7,09
Scenario 5					
Staff [Pers]	9	2	6	4	3
C/T [min]	12,46	10,41	11,67	9,50	8,87
LOT [min]	7,78	10,41	11,67	N/A	7,09
Scenario 6					
Staff [Pers]	10	2	6	4	3
C/T [min]	11,21	10,41	11,67	9,50	8,66
LOT [min]	7,78	10,41	11,67	N/A	7,09

C

Appendix C

Calculations for capacity and staffing for each process step. All values are in minutes that not are specified are in minutes.

C.1 Flow 1

Rebalancing Flow 1 with 25% System loss

$$\text{Adjusted utilization}_{25\%} = 0.85 \cdot \frac{0.75}{0.95} = 0.671 \approx 0.67 \tag{C.1}$$

$$\text{F1 Capacity after balancing} = \frac{\text{AT} \cdot 0,67}{\text{LOT2}} = \frac{268,65}{4,67} = 57,53 \implies 57 \text{ Engines} \tag{C.2}$$

C.2 Paint shop

Special thing with paint shop was the machine time that is included in total work time, which implies that the calculations only needs to use time used by operator.

$$\text{Time used by operator} = \text{TW - Machine time} = 42,57 - 29,2 = 13,37 \tag{C.3}$$

No rebalancing

$$\text{Paint shop Capacity} = \frac{\text{AT} * 85\%}{\text{LOT}} = \frac{362,95}{10,41} = 34,87 \implies 34 \text{ Engines} \tag{C.4}$$

$$\text{Operators needed} = \frac{\text{Time used by operator}}{LOT} = \frac{13,37}{10,41} = 1,28 \implies 2 \text{ Ops} \tag{C.5}$$

Rebalancing

In paint shop the LOT will not change when rebalancing due to it being machine time.

$$\text{Paint shop Capacity rebalancing} = \frac{\text{AT} * 85\%}{\text{LOT}} = \frac{362,95}{10,41} = 34,87 \implies 34 \text{ Engines} \tag{C.6}$$

C.3 Flow 2

No rebalancing

$$\text{F2 Capacity} = \frac{\text{AT} * 85\%}{\text{LOT}} = \frac{362,95}{11,67} = 31,1 \implies 31 \text{ Engines} \tag{C.7}$$

With 10% system loss

$$\text{Adjusted utilization}_{10\%} = 0.85 \cdot \frac{0.90}{0.95} = 0.805 \approx 0.81 \tag{C.8}$$

$$\text{F2 Capacity} = \frac{\text{AT} \cdot 81\%}{\text{LOT}} = \frac{344.25}{11.67} = 29.5 \implies 29 \text{ Engines} \tag{C.9}$$

$$\text{F2 Operators needed} = \text{Qual.Contr Ops} + \frac{L/T}{LOT} = 1 + \frac{57}{11,67} = 5,85 \implies 6 \text{ Ops} \tag{C.10}$$

Rebalancing

$$\text{F2 new LOT} = \frac{\text{LOT}}{\text{Number of stations}} = \frac{64,57}{7} = 9,22 = \text{ LOT2} \tag{C.11}$$

$$\text{F2 Capacity rebalancing} = \frac{\text{AT} * 85\%}{\text{LOT2}} = \frac{362,95}{9,22} = 39,35 \implies 39 \text{ Engines} \tag{C.12}$$

With 25% system loss

$$\text{Adjusted utilization}_{25\%} = 0.85 \cdot \frac{0.75}{0.95} = 0.671 \approx 0.67 \tag{C.13}$$

$$\text{F2 Capacity rebalancing} = \frac{\text{AT} \cdot 67\%}{\text{LOT2}} = \frac{286.875}{9.22} = 31.11 \implies 31 \text{ Engines} \tag{C.14}$$

C.4 Engine testing

Engine testing represents a unique phase in the production process, as all operations are identical and performed in parallel within dedicated testing cells. This parallel configuration precludes traditional line balancing. Consequently, the capacity calculations remain constant regardless of whether the flow is rebalanced or not. In this setup, the number of operators is intrinsically linked to the number of available testing cells, as each unit requires a discrete, parallel resource.

No rebalancing/Rebalanced

$$\text{Engine test Capacity} = \frac{\text{AT} * 85\%5}{\text{LOT}} = \frac{362,95}{12,67} = 66,86 \implies 66 \text{ Engines} \tag{C.15}$$

$$\text{Engine test Operators} = \text{Number of testing boxes} = 7 \tag{C.16}$$

C.5 Flow 3

No rebalancing

$$\text{F3 Capacity} = \frac{\text{AT} * 85\%}{\text{LOT}} = \frac{362,95}{7,09} = 51,19 \implies 51 \text{ Engines} \tag{C.17}$$

$$\text{F3 Operators needed} = \frac{L/T}{LOT} = 1 + \frac{26,61}{7,09} = 3,75 \implies 4 \text{ Ops} \tag{C.18}$$

Rebalancing

$$\text{F3 new LOT} = \frac{\text{LOT}}{\text{Number of stations}} = \frac{26,61}{8} = 3,33 = \text{ LOT2} \tag{C.19}$$

$$\text{F3 Capacity rebalancing} = \frac{\text{AT} * 85\%}{\text{LOT2}} = \frac{362,95}{3,33} = 109,12 \implies \text{109 Engines} \tag{C.20}$$

With 25% system loss

$$\text{Adjusted utilization}_{25\%} 0.85 \cdot \frac{0.75}{0.95} = 0.671 \approx 0.67 \tag{C.21}$$

$$\text{F3 Capacity rebalancing} = \frac{\text{AT} \cdot 67\%}{\text{LOT2}} = \frac{286.875}{3.33} = 86.15 \implies \text{86 Engines} \tag{C.22}$$

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF SUPPLY AND OPERATIONS MANAGEMENT
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY