Performance engineering of high-mix low-volume production systems

TNO Public ⟩ TNO 2025 R10304
14 March 2025

ICT, Strategy & Policy
www.tno.nl
+31 88 866 50 00
info@tno.nl

TNO 2025 R10304 – 14 March 2025

Performance engineering of high-mix low-volume production systems

Author(s)	Bram van der Sanden Jacques Verriet
Classification report	TNO Public
Title	TNO Public
Report text	TNO Public
Number of pages	30 (excl. front and back cover)
Number of appendices	0
Programme name	AIMS
Project name	AIMS 2024
Project number	060.59493/01.01

Our partners

Acknowledgement
The research is carried out as part of the AIMS program under the responsibility of TNO-ESI with Canon Production Printing as the carrying industrial partner. AIMS is funded by Holland High Tech | TKI HSTM via the PPS allowance scheme for public-private partnerships.

Abstract

High-mix low-volume (HMLV) production systems are crucial for enabling mass customization to meet diverse and rapidly changing customer demands. These systems require optimization across various time scales to handle fluctuations in job mix and ensure robustness against unforeseen changes, such as rush orders. Traditional scheduling techniques and commercial tools often fall short in addressing the complexities of HMLV manufacturing, including material changes, setup times, and job flexibility.

This report identifies a gap in systematically reasoning about HMLV manufacturing and proposes new methodologies to analyze and optimize these systems. We address two primary research questions: (1) How to specify HMLV production systems? and (2) How to effectively allocate jobs to available equipment considering constraints and uncertainties?

Our findings highlight the importance of characterizing HMLV production systems using a framework that incorporates flexibility across different time scales and operational levels. We introduce a performance engineering methodology that leverages simulation models to analyze loosely connected production systems and propose future enhancements, including domain model generation and optimization for job allocation. This approach aims to improve the efficiency and adaptability of HMLV production systems, ensuring they can meet the demands of mass customization.

Contents

1 Introduction

High-mix low-volume (HMLV) production systems are an important enabler for mass-customization of products for different types of customers. Customer demands are changing more rapidly over time and can differ quite significantly from each other. This poses challenges in the production process, that needs to be optimized on various time scales considering the varying job mix. For example, at a yearly time scale to deal with seasonality, and at a monthly or daily time scale to deal with fluctuations in customer orders. Planning and scheduling decisions need to be made continuously, and these should make the production system robust to unforeseen changes (e.g., rush orders). The production system also needs to have a high degree of flexibility to deal with the large variety of intermediate- and end-products; i.e., resources like machines and operators that have various capabilities and can deal with various materials.

It is challenging to use available commercial tooling and traditional scheduling techniques to capture and reason about the complicated nature of HMLV manufacturing [1] [2]. Typical challenging elements include accounting for the changes in materials being handled, the setup times associated with changing between materials being processed, dealing with rush jobs that must be squeezed in existing schedules, and accurately capturing the flexibility that jobs can be processed using different machines or in different possible orders.

We observe that there is a gap to reason on HMLV manufacturing in a systematic way. Analyzing and optimizing performance of HMLV production systems requires new methodologies that explicitly consider the flexibility on different time scales and can predict the impact of planning and scheduling decisions. In this report, we address this research gap, by providing an answer to the following research questions:

- **Research question 1:** How to specify high-mix low-volume production systems?
- **Research question 2:** How to effectively allocate jobs to available equipment considering job and equipment constraints and uncertainties?

In the remainder of the report, we introduce the background on the domain of HMLV production systems in Chapter 2. Here, we also make the comparison between HMLV and low-mix high-volume (LMHV) and describe the application domains of HMLV. Chapter 3 describes the characteristics of HMLV production systems, and what metrics are currently available to analyze their performance. We also describe the different types of decisions that need to be made in planning and scheduling, and how to analyze HMLV production systems. Chapter 4 describes the first version of our performance engineering methodology for HMLV production systems, building upon earlier work we conducted around modeling and analysis of production lines. Chapter 5 concludes the report, reflecting on the research questions.

2 High-mix low-volume production systems

In this chapter, we present a background on HMLV production systems. We describe the HMLV domain in Section 2.1 and in Section 2.2 why HMLV is gaining importance. In Section 2.3 we describe how it contrasts to traditional low-mix high-volume production that focuses on producing large amounts of the same product at a low cost. Section 2.4 describes some application domains where high-mix low-volume is playing an important role.

2.1 High-mix low-volume

In modern manufacturing, where customization and flexibility have become increasingly important, **High-Mix Low-Volume (HMLV) systems** have emerged as a strategic approach to meet diverse customer demands efficiently. HMLV systems specialize in producing a wide variety of products (high-mix) in smaller quantities (low-volume), in contrast to traditional mass production systems that prioritize large quantities of identical products. HMLV production environments focus on producing one-of-a-kind products or small batches of the same product [3].

HMLV systems can be considered considering the Product-Process Matrix, as defined by Hayes and Wheelwright [4]. The matrix illustrates how certain types of production processes are better suited for specific product-volume mixes. They show the continuum from highly customized products produced in low volumes to standardized products produced in high volume. If the volume and level of flexibility are not in line, then either there is more flexibility than needed or less flexibility than needed, leading to a higher cost.

Table 2.1: Product Process Matrix by [4]. It shows how higher product volumes often go together with more streamlined flow process structures. It also shows that off-diagonal configurations are (typically) not economically feasible.

Product structure Process structure	Low Volume Unique products	Low Volume Multiple products	High Volume Standardized products	Very High Volume Commodity products
Jumbled Flow	Job shop		*More process flexibility than required, so higher cost*	
Disconnected Line Flow		Batch		
Connected Line Flow	*Less process flexibility than required, so higher cost*		Assembly line	
Continuous Flow				Continuous

2.2 Mass customization as trend driving HMLV

High-mix low-volume closely links to the trend of **mass customization**, where the individual needs of the customer are put first. **Make-to-order (MTO)** manufacturing [5] is closely related to mass customization. With MTO, products are produced on demand, under a service license agreement (SLA) that describes the due date for the order as well as the design and specifications of the products to be manufactured. One step closer to the customer is **engineer-to-order (ETO)** [6], that also includes the design process. In ETO, a customer has specific needs or a required concept, and from there a design is created together with the manufacturing company. This can be in a tight collaboration, or by shifting design tools to the customer, exposing the capabilities of the manufacturing process. For example, when designing a custom photo album or a front plate for electronic equipment. Both for MTO, and ETO, typically the product is produced in a low volume. This contrasts with make-to-stock, where items are mass-produced to match inventory with forecasted demand.

2.3 Comparing HMLV and LMHV

The decision of whether to go for HMLV or low-mix high-volume (LMHV) production might be influenced by the expected market approach in the application domain. For example, the production of cheap products is typically done using LMHV production, whereas the personalization is typically performed using HMLV production.

HMLV production has several strengths compared to LMHV production [7]:

- It allows for increased customizability of processes to meet specific customer demands, enhancing customer satisfaction and enabling tailored solutions.
- The agile nature of HMLV production ensures that manufacturers can swiftly adapt to changes and disruptions in supply chains, maintaining continuity and responsiveness.
- HMLV facilitates the decentralization of production processes, promoting supply chain independence and enabling adaptation to country-specific requirements.
- HMLV production also optimizes space efficiency within production sites, as it requires less room for large-scale production runs.
- HMLV supports late-stage customization, allowing for a diverse range of products to be customized closer to the point of delivery, thereby reducing inventory costs and increasing market responsiveness.

There are also various weaknesses compared to LMHV production:

- The costs per product are typically higher than in LMHV, as there are more frequent changeovers and there is a reduced economy of scale.
- The processing times for products are typically longer than in LMHV, caused by the repeated switching between the different jobs.
- Maintaining a consistent quality across a diverse product range is more difficult than producing the same product in large quantities.
- Managing the supply chain, including inventory management and procurement, is much more challenging, as there is a wide variety of parts and materials.

2.4 Application domains

HMLV manufacturing is found in various domains. Some examples where HMLV manufacturing can be found a lot include:

- **Printing domain:** with customized packaging, labels, and promotional materials with varying designs, sizes, and finishes to meet specific client requirements. Examples of products are calendars, business cards, notebooks, booklets, labels, stickers, printed shirts, mugs, and backpacks. The products can be fully unique, or produced in small batches, for example a small batch of merchandise items.

- **Personalized healthcare:** including personalized medications and medical devices that are tailored to the patient. Examples of medical devices include orthoses, protheses or ear plugs. Each product is typically unique, made for a specific patient.
- **Pharmaceutical industry:** using small-scale production lines that manufacture and package small batches of medicine or medical products. For example, the sterile coating of rubber products used in healthcare, assembly of syringes, packing products in blisters, and adding tamper evidence solutions like tear-off edges to medical packaging.
- **Metal fabrication shops:** producing customized make-to-order metal products, using flexible welding, cutting, forming, and machining equipment.
- **Tool shops:** producing tools, dies, and molds. For example, custom-made molds can be produced for plastic injection molding, where each mold is designed to produce unique parts with specific shapes and features. Tool shops may also produce bespoke cutting tools, such as custom drill bits or milling cutters, designed to perform specific tasks or work with materials in niche industrial applications.
- **Additive manufacturing:** also known as 3D printing, to fabricate physical 3D objects based on virtual 3D models. Typically, limited quantities are produced, for example spare parts, products that cannot be produced by traditional CNC machines, or personalized products.

3 Characterizing HMLV production systems

In this chapter, we describe the typical characteristics of HMLV production systems. These characteristics form an important ingredient to answer the first research question on how to describe HMLV systems. Section 3.1 describes the characteristics of HMLV production systems. The well-known framework of the 4M's of manufacturing can be used as starting point to structure the characterization of HMLV production systems, explained in Section 3.2. Section 3.3 describes the different types of flexibility that are key to produce a high variety of product in small quantities. Section 3.4 explains the types of decisions that need to be made when operating an HMLV production system. Section 3.5 explains how performance analysis can be performed, considering the aspect of distinguishing element of flexibility in HMLV production systems, and finally Section 3.6 explains how to give insight in the analysis results.

3.1 HMLV production system characteristics

HMLV production systems have specific characteristics to deal with unpredictable demand and unpredictable changes in the product mix. Typically, HMLV production systems have multi-functional machines and multi-skilled workers that are working on multiple jobs at the same time and shifting their production methods frequently. Setup times and cycle times of any operation may vary per job and may depend on the job sequence. These variations affect performance, i.e., the produced output of the production system per time unit. To optimize performance, jobs are often batched on the machines [8], and sometimes also share the same material (e.g., the same metal or paper sheet), referred to as ganging [9]. These strategies help to optimize performance of a set of jobs, for example of a whole shift. Note that for individual products the cycle time might degrade, due to the buffering during the production process. Another important factor when optimizing performance is the order in which jobs are sequenced.

Consequently, we identify the following characteristics making HMLV production systems complex:

- Multiple jobs will be processed at the same time, competing for shared resources like operators and equipment, and each job might have their own routing through the production system.
- The different operations inside a job may have precedence relations. Different jobs may contain different operations and precedence relations between its operations. The operations are not necessarily sequential, meaning that multiple operations of the same job could run in parallel.
- Setup times and cycle times of any operation may vary per job. Setup times for operations may also depend on the job sequence.
- Due to low-volume orders, production is based on shorter runs, with setup times playing a key role.

3.2 Characterizing using the 4M's

At a high level, a HMLV production system can be characterized using the 4 M's of manufacturing, just like any other manufacturing system. For each "M" characteristic, there are specific performance aspects to consider. Later additions introduced 4 additional M's (Measurement, Mother Nature, Money, and Management System), but these are less important from a manufacturing process perspective.

Table 3.1: The 4 M's of manufacturing[1]

"M" characteristic	Description	Performance aspects
1. Manpower	Labor of people involved in delivering products and services for production	Efficiency of the operators
2. Machine	Equipment, facilities, tools employed for production	Usage and maintenance scheduling, avoiding unnecessary downtime
3. Method	Process, shipping, schedule, procedure	Products flow efficiently through the production line, finished products quickly exit the line
4. Material	Raw materials, consumables, components used to satisfy production	Raw materials are available, and parts and materials are close to the workstation, while not overcrowding the operator.

Depending on the specific question(s), a HMLV production system model will describe each of the four M's (manpower, machine, method, material) in less or more detail. Typical aspects that are modelled for each "M" characteristic for HMLV production systems are shown in Table 3.2.

Table 3.2: "M" characteristic modelling aspects

"M" characteristic	Modelling aspects
Manpower	Operators: skills, working schedules (shifts, working days, working times ...)
Machine	Resource types, resource instances, capacity
Method	Jobs, operations, operation dependencies, duration of operations, required resource (type(s)), sequence-dependent setup times
Material	Raw materials, intermediate products, consumables

3.3 Characterizing flexibility

A key characteristic of high-mix low-volume systems is having flexibility to deal with changes in demand and product mix over time. An often-used definition of manufacturing flexibility is given by Upton [10]: "the ability of a manufacturing system to change or react to changing circumstances with little penalty in time, effort, cost or performance". Note that there will always be a penalty, as having flexibility to produce a variety of products costs in terms of an increased delivery time, increased stock, or reduced production capacity. Flexibility types, called dimensions, like process or machine flexibility, can be defined in terms of four attributes [11]:

1. **Range:** the extension of the differences under a given dimension. The range increases if a wider set of options or alternatives are possible.

[1] https://leancommunity.org/4m-method-meaning/

2. **Resolution:** how close are alternatives within the range of a given dimension.
3. **Mobility:** how easy it is to switch between alternatives under a given dimension. Mobility is high if the transition penalties between different alternatives are low. For example, whether it is easy to switch from performing one operation to another one.
4. **Uniformity:** how the system performance varies while moving within the range. A high uniformity means that the performance (e.g., incurred cost or required resources) is similar for different alternatives.

Figure 3.1 shows an illustration of the four flexibility attributes: range, resolution, mobility, and uniformity.

Attribute	Low attribute value	High attribute value
Range	[ ]	[ ]
Resolution	[x x x x x]	[x x x x x x x x x x x]
Mobility	10 10 [x x x] 25	25 40 [x x x] 80
Uniformity	[x x x] +/- ++ -	[x x x] ++ ++ ++

Figure 3.1: Illustration of the four flexibility attributes. An "x" is an alternative. Arrows with a number indicate a switch between two alternatives with an associated transition penalty. Performance of an alternative is scaled from – to ++, where -- represents low performance and ++ represents high performance.

Manufacturing flexibility encompasses a broad set of flexibility dimensions, that can be linked to the 4 M's introduced in the previous section. For example, labor flexibility links to manpower, whether operators are trained to handle a wide range of tasks, and whether the company can easily deploy additional staff and vary working hours.

The flexibility dimensions can be considered at different time scales and at different operational levels, as shown in Figure 3.2. There are long-term effects, for example having expansion flexibility to grow the company, or short-term effects, for example which machine is used to produce a product item. The different forms of flexibility can also be considered at different operational levels in the production environment. From a high company-level to the low product-level. The figure captures the flexibility dimensions that are most often considered in the context of manufacturing flexibility [12].

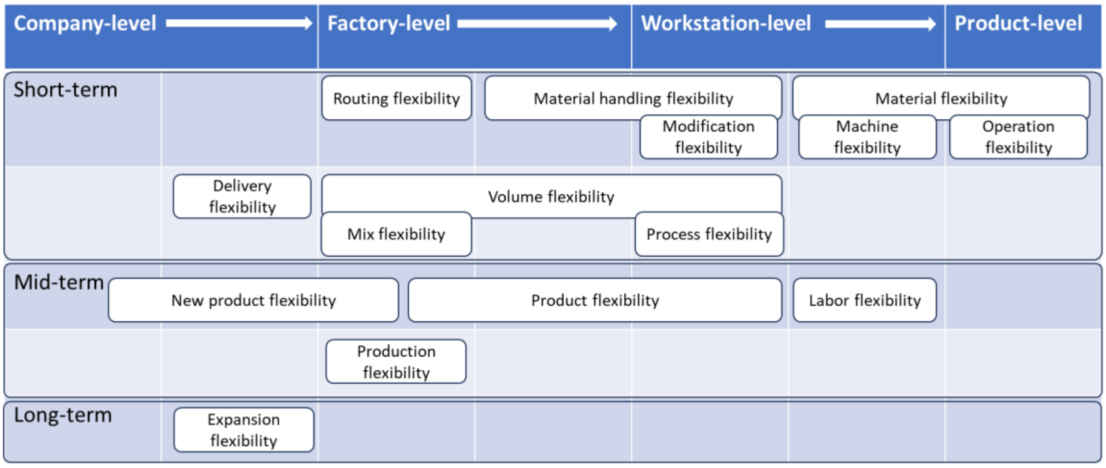

Figure 3.2: Flexibility elements on the short-, mid-, and long-term along the time axis, and at the company-, factory-, workstation-, or product-level on the operational level axis. Adapted from [12].

3.4 Characterizing the levels of decision making

Decision making in manufacturing systems can be considered at different time scales. Planning is often divided into three levels: strategic, tactical, and operational (see e.g., Chapter 14 in [13]). Strategic planning is long-term, focusing on developing strategies that form the basis for tactical planning as well as operational planning. Typical questions involve deciding on the products to offer, addressing market demands and determining the required factory equipment as well as potentially outsourcing operations to other companies. Tactical planning is mid-term, anticipating expected demand in a season or year, as well as what is required to meet this demand. Operational planning is short-term and is about scheduling the equipment and operators to ensure that incoming jobs are delivered on time.

On each level, there are various decisions that need to be made, based on the available information. An overview with typical questions is shown in Figure 3.3.

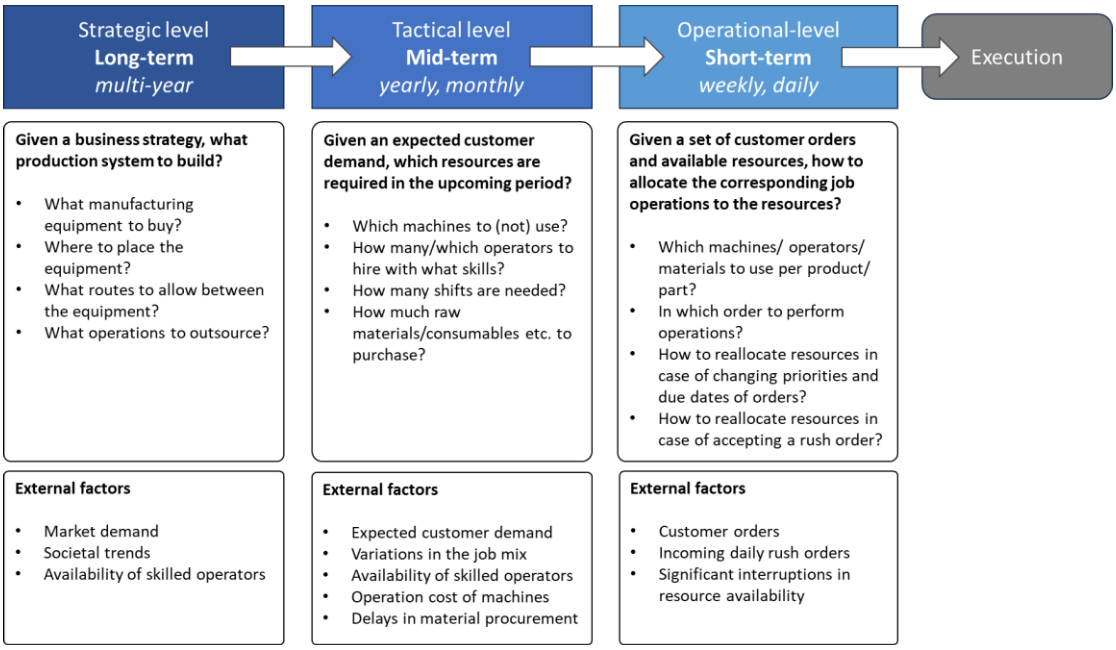

Figure 3.3 Decision support on the three levels of production planning.

3.5 Characterizing performance

Performance of production systems in general is often characterized by overall equipment effectiveness (OEE), discussed in Section 3.5.1. In Section 3.5.2, we consider the link between OEE and flexibility, as flexibility negatively impacts OEE.

3.5.1 Overall Equipment Effectiveness

To measure the performance of HMLV systems, it is important to have Key Performance Indicators (KPI) that quantify the performance. OEE is a well-known KPI, often used for manufacturing systems in general. It relates the actual system performance to the (theoretical) maximum performance, considering the aspects of availability, performance, and quality:

- **Availability:** actual production time as percentage of the scheduled production time.
- **Performance:** actual produced output per time unit as percentage of the expected output per time unit.
- **Quality:** non-defective output as percentage of the total output.

The OEE of a system is computed as the product of the availability rate AR, the performance efficiency PE, and the quality rate QR: $OEE = AR \times PE \times QR$ [14]. These ratios are computed as follows:

- $AR = \frac{OT-DT-ST}{OT}$, where OT represents the system's planned operation time, DT the system's delay time, and ST the system's setup time.
- $PE = \frac{PA \times CT}{OT}$, where OT represents the system's planned operation time, PA the produced number of produced products, and CT the ideal cycle time per product.
- $QR = \frac{PA-DA}{PA}$, where PA represents the produced number of products and DA the number of defectives products produced.

An OEE of 100% means that only good parts are produced (100% quality), at maximum speed (100% performance), and without interruption (100% availability). If the OEE is not 100%, then there are losses in one or multiple of the aspects. Table 3.3 lists the loss categories associated to the OEE aspects, with typical examples of losses.

Table 3.3: OEE aspects and related losses.

OEE	Loss categories	Examples
Availability	Unplanned stops	Equipment failure, breakdown, unplanned maintenance, lack of materials
	Planned stops	Setup and adjustments, cleaning, planned maintenance
Performance	Small stops	Idling and minor stops, periodic cleaning
	Slow cycles	Operator inexperience, worn-out equipment, startup and shutdown at lower production speed
Quality	Production rejects	Process defects, incorrect settings
	Startup rejects	Reduced yield, scrapping parts

OEE measurement methods focus on a single machine or production line instead of the whole production system. They also do not differentiate between production effectiveness per product in case the system produces multiple types of products. This makes it difficult to use OEE to quantify and reason on productivity of a production system involving multiple resources and multiple product types being produced.

3.5.2 Considering flexibility in OEE

As HMLV systems' characteristics are different from traditional mass production systems', dedicated methods are needed to reason about these systems. Static metrics like OEE that are used to analyze mass production systems are typically not suited to deal with the variability, important role of setup times, and short jobs found in HMLV systems. OEE does not consider flexibility of the production system. Dealing with a high product mix in low volumes often even has a negative effect on OEE:

- **Lower availability:** as more set-ups are needed to produce different types of products.
- **Lower performance:** as runs are short, and the run-in, run-out behavior plays an important role.
- **Lower quality:** as the increased impact of start-up defects leads to more frequent quality issues.

A well-known principle is that "what you measure is what you get", as for example is articulated by historian H. Thomas Johnson: "*Perhaps what you measure is what you get. More likely, what you measure is all you'll get. What you don't (or can't) measure is lost".* Therefore, it is important to have the right characterization of HMLV systems, that also considers the positive aspects of flexibility.

Van De Ginste [15] has introduced OEE-Flex as a KPI that is based on the aspects of flexibility. OEE-Flex considers the effectiveness of equipment, linked to the aspects of mobility and uniformity, and the capability to adapt, linked to the range aspect. The paper however does not give a detailed description of how to compute OEE-Flex. Figure 3.4 shows the relationship between the losses considered by OEE and OEE-Flex.

To the best of our knowledge, there is no other related work on metrics with a dedicated focus on HMLV systems.

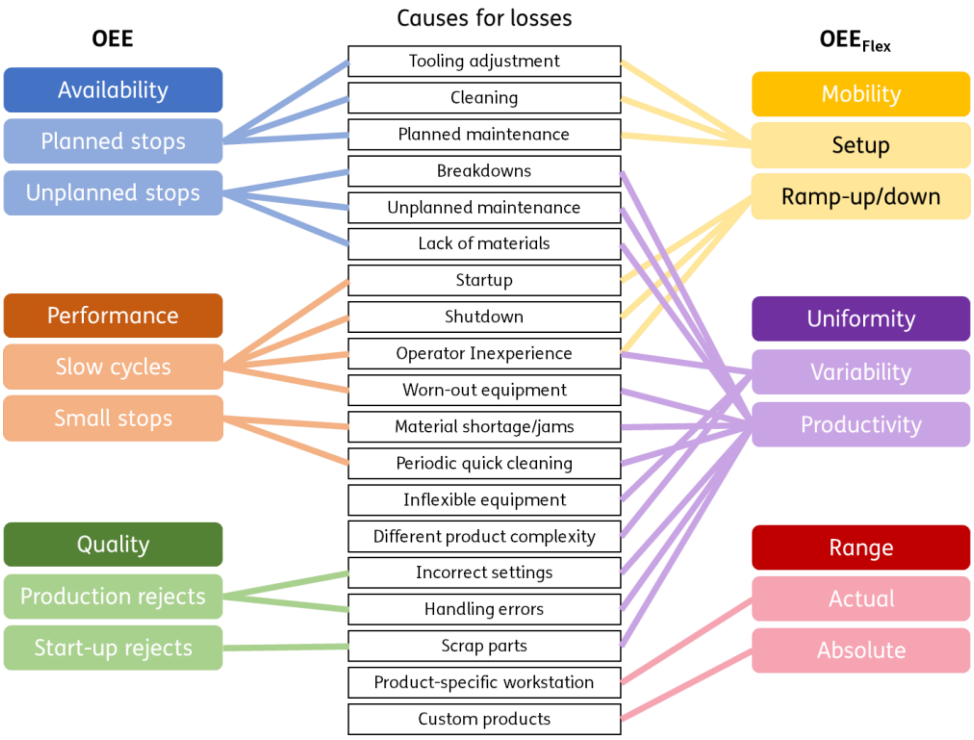

Figure 3.4: Relationship between the aspects covered by OEE and the flexibility aspects covered by OEE-Flex. Figure adapted from [15].

3.5.3 Metrics covering multiple machines

As discussed in the previous section, OEE is originally intended to analyze a single machine running long batches of the same product. Over time, various extensions have been proposed to focus on a factory scope, covering multiple machines or even an entire production facility, instead of considering only a single machine. OFE and OTE are extensions proposed by academia but seem to be not broadly adopted in industry. OPE and OAE are terms that are claimed to be used more in industry but are not well-defined. Concluding, there does not seem to be consensus on how to analyze productivity using commonly accepted metrics in contrast to OEE which is broadly used and has a well-defined meaning.

Metric	
Overall Factory Effectiveness (OFE) [16]	OFE is used as term about integrating decisions and actions across subsystems, ensuring that aspects like planned downtime and setup time are synchronized. There is no consensus about the actual metrics in literature for OFE [17].
Overall Throughput Effectiveness (OTE) [18]	Evaluate throughput effectiveness at a factory level, composed of series, parallel, assembly, and expansion subsystems. OTE is defined as a generalization of OEE, focusing on the factory instead of a single machine.

Metric	
Overall Production Effectiveness (OPE) Overall Asset Effectiveness (OAE)	Both terms are used to identify and measure losses with the overall production process. The terms are not used a lot in literature, but one report indicates that they are used extensively in industry [17]. The specific elements or losses measured differ from between industries.

3.6 Performance observability

After characterizing the performance of a HMLV production system, it is important to make the performance insightful. Insights can be given for each of the four "M" characteristics, described in Section 3.2. Table 3.4 shows typical analysis aspects for each of the four "M" characteristics.

Table 3.4: "M" characteristic analysis aspects

"M" characteristic	Typical analysis aspects / metrics?
Manpower	Operators: not working, idle, working
Machine	Resource state (e.g., off, on, idle, waiting, working), workload
Method	Job schedule: start/finish time job, due date, early/lateness
Material	Product location and state over time

Gantt charts are often used as visualization, for example focusing on the jobs being processed over time, or the use of the resources:

- **Gantt chart with jobs** with time on the X-axis and jobs on the Y-axis, where blocks in the Gantt chart are showing waiting time, or operations waiting or being executed. Such a view is useful to analyze the total execution time of the job, and where and how waiting time or execution time is being spent.
- **Gantt chart with resources** with time on the X-axis and resources (e.g., equipment and operators) on the Y-axis, where blocks are operations, and could be colored based on other properties like the corresponding job. Such a view is useful to analyze which resources are the bottleneck, and the utilization of the resources in general.

4 Performance engineering for efficient decision making

In Chapter 3, we explained how HMLV systems and their performance can be characterized. In this chapter, we describe the first version of our performance engineering methodology for modeling and analysis of HMLV production systems. Our focus is to make the right scheduling decisions to effectively allocate jobs onto available equipment, considering job and equipment constraints and uncertainties.

In Section 4.1, we present the previous work on modeling and analysis of tightly coupled production lines. Section 4.2 presents our vision on how we want to extend the methodology towards modeling and analysis of loosely coupled production lines. One of the key ingredients is a simulation model, that can deal with the stochastic timing behavior and can simulate different allocations of jobs onto available equipment. The structure of this model is described in Section 4.3. An example model then given in Section 4.4 with the performance analysis results described in Section 4.5.

4.1 Previous work on modeling and analysis of production lines

The work presented in this chapter extends upon earlier work that focused on describing and analyzing performance of individual tightly coupled production lines [19] [20], not yet considering the flexibility in allocating jobs to different production lines, and only considering fixed execution times when executing operations. Figure 4.1 gives an overview of the approach, where design assistance is provided to analyze and optimize the planning of a production line. The human operator captures the materials being handled, available equipment, and the jobs that need to be processed in the domain model. The operator also specifies the allocation which equipment is used for each job. From this domain model, a constraint graph is generated. Such a graph captures the relative timing constraints between start and/or end points of operations that need to be performed. This graph can be analyzed to compute a scheduled, visualized as Gantt chart, showing the timing behavior of the production line. Based on the Gantt chart, the human operator can identify bottlenecks and determine how the planning could be improved. Then, a next design loop can be started by adapting the model with the improvements and running the analysis again. The design assistance will then help to determine whether the modification indeed leads to an improvement in the production plan.

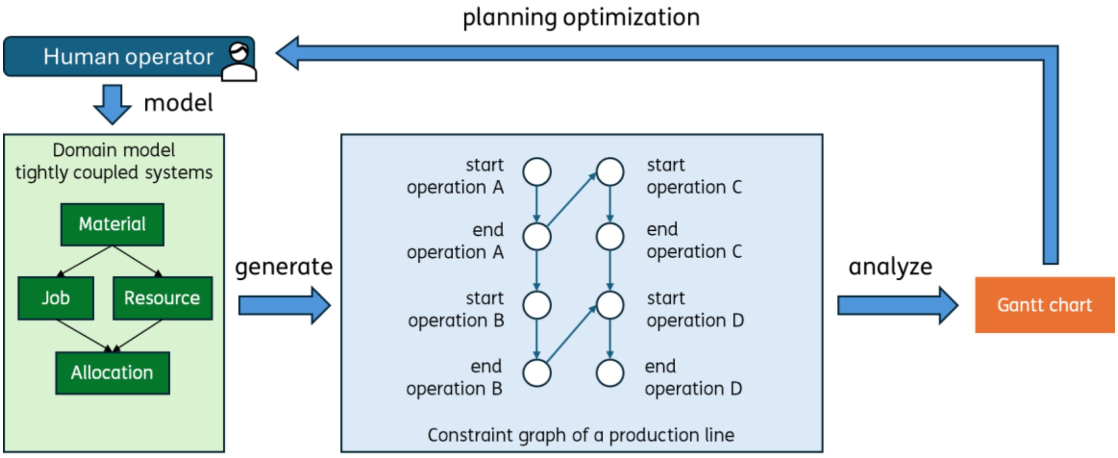

Figure 4.1: Performance analysis of individual production lines using the methodology described in [19].

The domain model used to model production lines links closely to the four M's introduced in Section 3.2, but is structured in a slightly different way:

Table 4.1: Mapping of the elements in [19].

Domain modeling element [19]	"M" characteristic
Material model	Material, Method (available operations)
Job model	Method
Equipment model	Machine, Method (operations, sequence-dependent setup times), Manpower
Allocation model	Does not directly relate to a characteristic, as it is a mapping of elements.

As can be observed from the table, this domain model makes an explicit distinction between equipment and jobs, whereas in the 4 M's, they are both part of the Method characteristic. The reason for this is that the model is used to analyze different allocations and sequencings of work to be done. Here, the allocation refers to which operations are executed by what equipment or operators and sequencing refers to the order in which operations are executed by the equipment or operators. Machine and Manpower are combined in this domain model, as for both the only aspect of importance is which jobs the machines and operators can execute, discarding distinguishing aspects like skills or work schedules that apply only to operators.

4.2 Extension to loosely coupled production lines

In this report we extend upon the previous work, by focusing not only on tightly coupled production lines, but rather on manufacturing systems with multiple of such tightly coupled production lines. These tightly coupled production lines are loosely coupled. This loose coupling might be realized by buffers containing intermediate products, and by transportation by human operators or forklifts moving products from one production line to another.

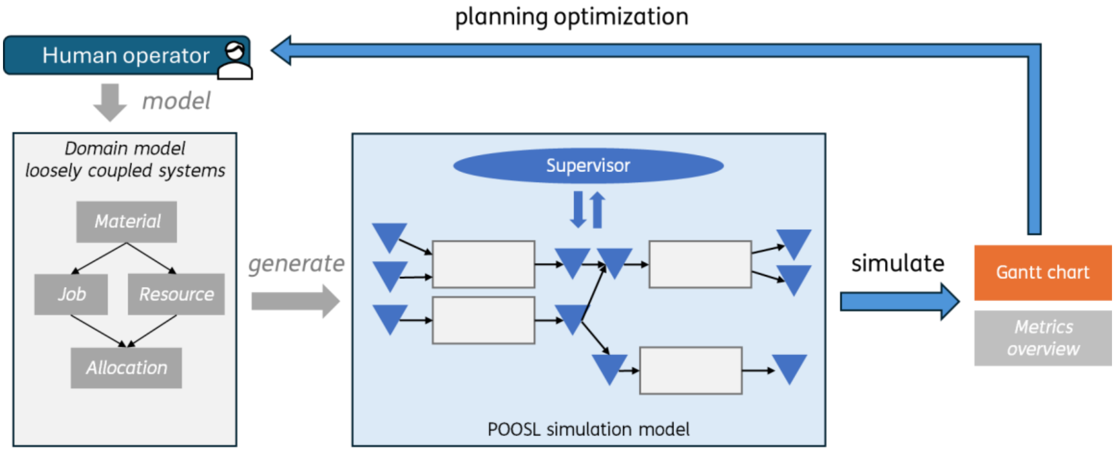

Figure 4.2: Vision of the extended performance engineering methodology to model and analyze loosely coupled production lines.

Figure 4.2 shows the vision for the extended methodology. To describe loosely connected production lines, the domain model should be extended. For example, to describe buffers, operators, and timing variability. This extension has not yet been developed. In the current version, we have manually created a simulation model that could later be generated from a domain model. A supervisor in the simulation model can execute scenarios where jobs have been allocated to the equipment. We envision an optimizer as part of the supervisor that generates efficient allocations and can dynamically adapt the allocation when changes are needed. For example, when rush orders arrive, or there is a machine breakdown. This optimizer is not yet present in the current version of the solution. With simulation, a schedule can be computed that can be visualized as Gantt chart. Next to the Gantt chart, we would like to add a dashboard showing computed metrics. For example, showing the utilization of the resources, and the lead times of jobs.

We have built a proof-of-concept simulator using Eclipse POOSL [21][2]. The focus of the current methodology is to reason on decision-making at the operational level (see Section 3.4), how to allocate jobs and operations onto the available equipment. We have put emphasis on how multiple production lines can be connected and operated by a supervisor. The constraint graphs used for individual production lines are not yet integrated in the simulator. These could be integrated by invoking the constraint graph analysis from the simulation.

In Section 4.3, we explain the structure and the simulation model. In Section 4.4, we show an example in Eclipse POOSL. Section 4.5 explains the performance analysis that can be done using Eclipse POOSL and Eclipse TRACE4CPS [22][3].

4.3 Simulation model structure

Before explaining the simulation model structure, we first explain the structure of the underlying production system. We assume a structure which matches the equipment hierarchy of the IEC 62264-1:2013 standard [23], as shown in Figure 4.3. To explain the concepts, we use an imaginary company that prints customized items like photo albums, posters, business cards, and mugs on demand. The company, “Enterprise” in Figure 4.3, has a few sites in different countries. On a site, there are typically different areas. In our example,

[2] https://www.poosl.org/
[3] https://eclipse.dev/trace4cps/

there is one area where mugs are printed, whereas in another area all paper products are printed.

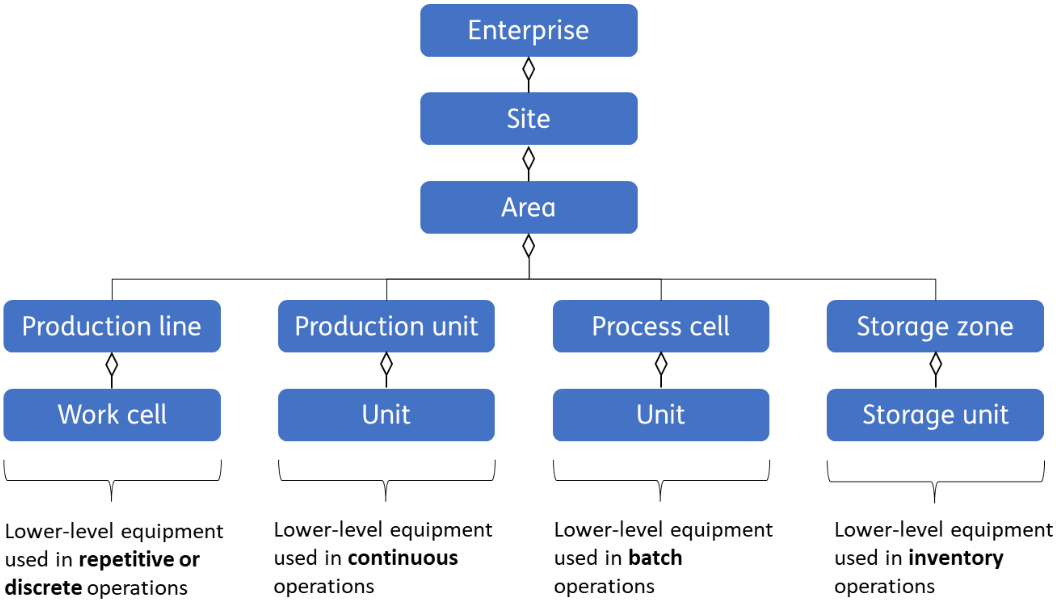

Figure 4.3: Equipment hierarchy according to the IEC 62264 standard.

An "Area" in Figure 4.3 consists of multiple production lines connected via buffers and transport. An example production line could be a printer with subsequent finishing equipment to cut, staple, or fold paper. These production lines match the "Production line" in Figure 4.3. As example, consider the production system structure shown in Figure 4.4. The gray rectangles in Figure 4.4 represent the production lines, the triangles represent the buffers, and the arcs represent the transport. Transport between buffers can be performed by a human operator or using automated transport like an automated guided vehicle or transport belt. A production line consists of equipment modules, like a printer or finisher equipment, referred to as "Work cells" in Figure 4.3, and their connections. The blue rectangles in Figure 4.4 inside the production lines represent the equipment modules. We assume the production lines to be tightly coupled, i.e. they have no or very restricted internal buffering between its modules [19]. Multiple production lines are loosely coupled via buffers and transport. If multiple production lines need to be tightly coupled, then we assume that they are combined into an aggregated production line possible including shared internal buffer(s). There can also be connections between two buffers, representing transportation of material between them.

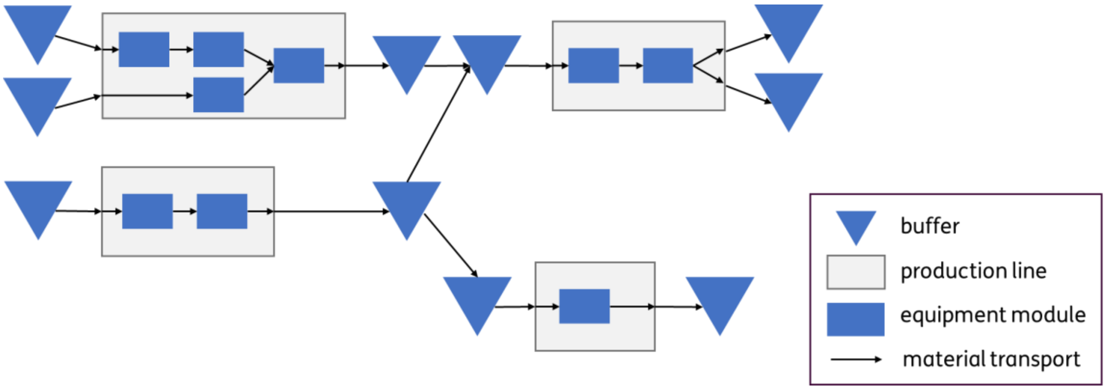

Figure 4.4: Production system structure.

The simulation model has a hierarchical structure with three levels, a sketch of which is shown in Figure 4.5. On the highest level, there is a supervisor. The supervisor assigns jobs to production lines, which form the middle level. The production line coordinates the jobs' operations to its equipment modules, which form the bottom level.

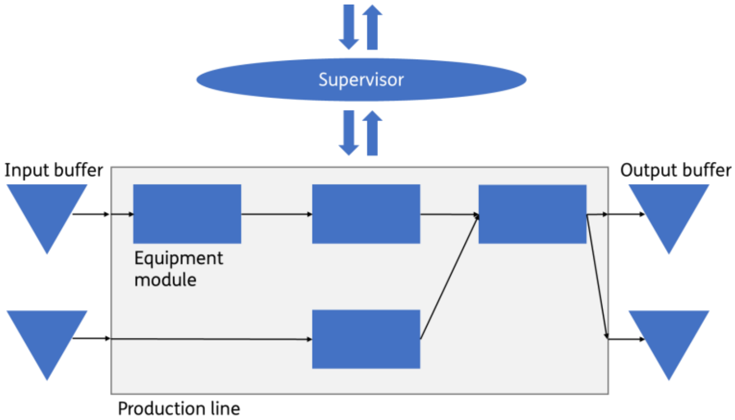

Figure 4.5: Simulation model structure.

Figure 4.6 illustrates the alignment between the supervisor and the underlying production lines and between the production lines and its equipment modules. The supervisor creates so-called allocated jobs. An allocated job corresponds to work to be performed by one production line. It is represented by a directed acyclic graph (DAG) of allocated operations, i.e., an operation and the corresponding equipment module to which it is assigned. The optimizer might require information on the equipment state if there are large sequence-dependent setup times possibly affecting the resulting performance.

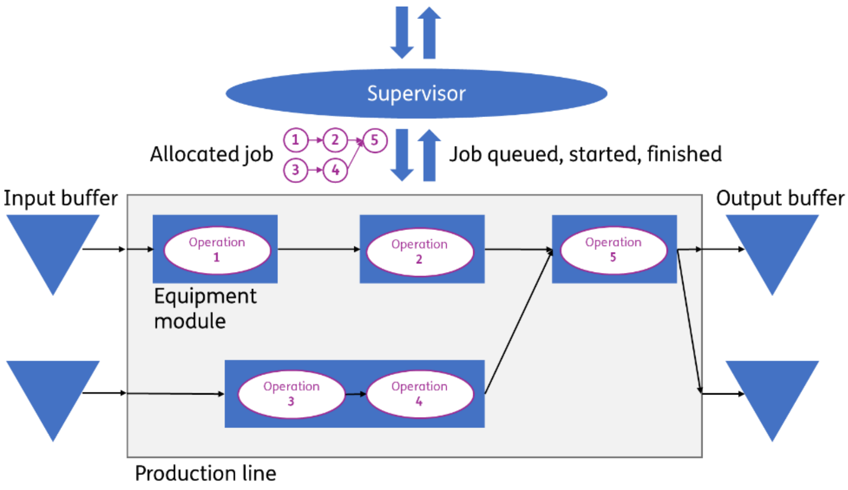

Figure 4.6: Communication between the supervisor and a production line and between the production line and its equipment modules.

The illustration in Figure 4.6 shows an allocated job, which is assigned to the single production line shown. The allocated job is a DAG, which consists of five operations and their allocation to the equipment modules. Note that multiple operations can be allocated to the same equipment module.

The interaction between the supervisor, the production lines, and the equipment modules is explained below and will be explained further using the example in Section 4.4.

- **Supervisor:** The supervisor is responsible for the overall execution of the allocated jobs. To start a job, it sends a *queue* message to the corresponding production line. The production line will acknowledge this with a *queued* message back to the supervisor, after which the supervisor can queue other allocated jobs. It can however also wait until it has received a *started* or a *finished* message from the production line before queueing the next allocated job. This depends on the dependencies between different allocated jobs.
- **Production line:** The production line is responsible for the communication with its equipment modules. It takes the first allocated job from its queue and sends a *started* message to the supervisor. It then determines the initial operations of the allocated job's DAG. In the example in Figure 4.6, there are two initial operations, i.e., operations 1 and 3. For these operations, the production line sends a *queue* message to the corresponding equipment module, which will acknowledge it with a *queued* message. The production line will also receive *started* and *finished* messages for each operation. After receiving a *finished* message, it will determine the next operations that can be queued. For instance, after operation 1 has finished, operation 2 can start. If all operations are reported as finished, the production line will send a *finished* message to the supervisor.
- **Equipment module:** The equipment module is responsible for the actual execution of operations. When it is idle, it takes the first operation from its queue and then determines the earliest time at which it can start. This depends on the preceding operations in the DAG. At this earliest point in time, it sends a *started* message to its production line, it determines the operation's duration and it executes the operation by waiting this duration. After this, it sends a *finished* message to its production line.

The aspect of timing variation is added to the model by expressing the timing of operations using probability distributions. The exact timing of an operation will be determined during simulation, by sampling from the specified distribution.

4.4 Eclipse POOSL simulation example

POOSL (Parallel Object-Oriented Specification Language) [21] (see also https://poosl.esi.nl) is an object-oriented modelling language, which is tightly integrated with the Rotalumis simulating engine in the Eclipse POOSL tool. Eclipse POOSL offers the means to describe and simulate systems. With POOSL, a system can be described in terms of communicating processes that interact via messages over a network. The messages are sent and received via ports of these processes. Processes can create and modify objects, that can be shared via messages.

This section explains an example of a POOSL simulation of a HMLV production system consisting of two production lines and a transport resource. The structure of the system is shown in Figure 4.7.

- One production line is a **printer line** with three modules: an input module, a printer module and an output module. The printer line takes empty sheets as input, prints them, and creates stacks of book blocks, i.e. the printed sheets that will form a book's content, as output.
- The second production line is a **book maker line**, which also consists of three equipment modules: an input module, a book maker module and an output module. The book maker line takes a book block and creates a book, e.g. by folding, stitching and trimming. The produced books are put on the output module.
- The third main element is a **transport line**, which is responsible for transporting (stacks of) book blocks from the printer line's output to the book maker line's input. The transport line consists of only one equipment module, a transport module.

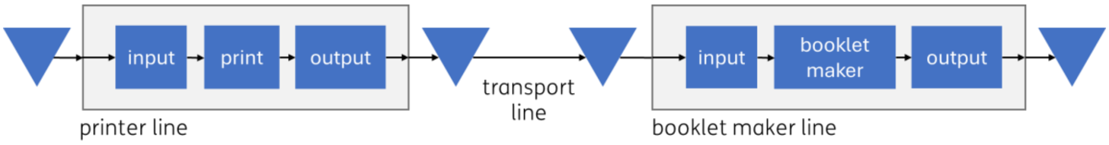

Figure 4.7: Structure of our example HMLV production system.

The overall structure of the corresponding POOSL simulation model is shown in Figure 4.8. It shows POOSL's so-called process classes and the communication channels between them. The communication channels are attached to the process classes' ports. When a message is sent on a channel via a port, it can be received by all connected process classes, which may be several.[4]

[4] One can guarantee reception by the correct process class by including recipient information in messages.

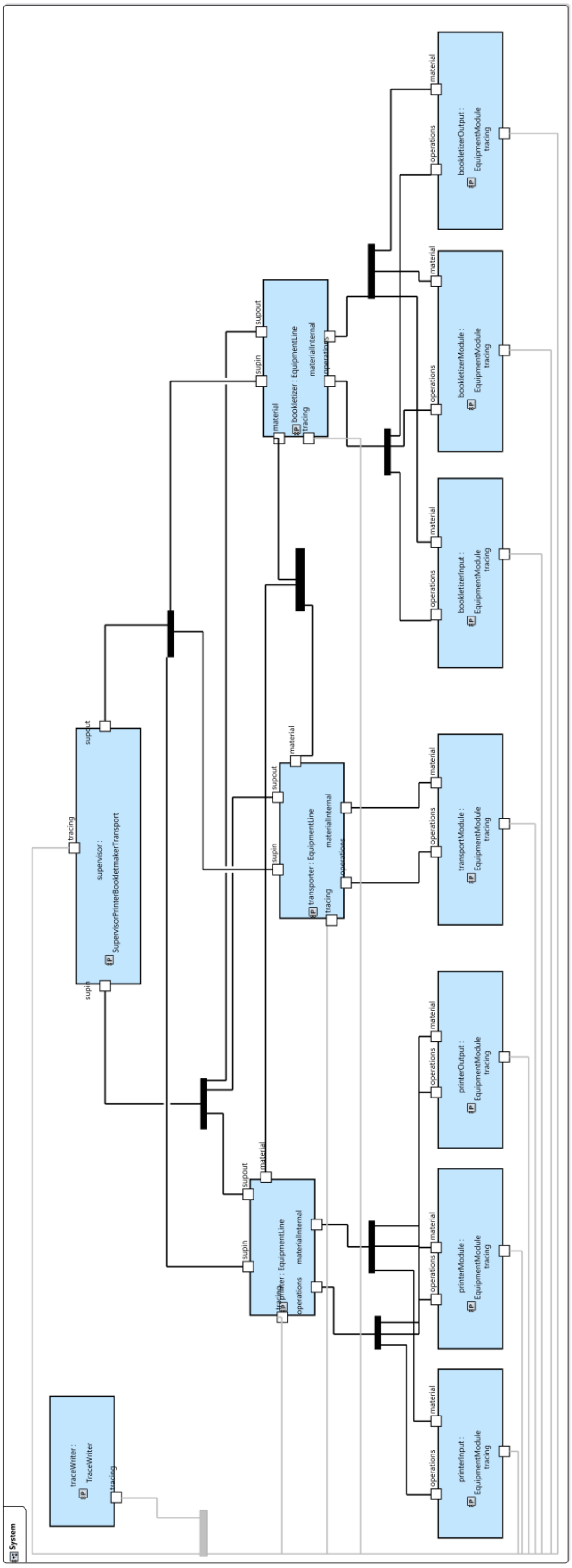

Figure 4.8: POOSL simulation model structure.

Figure 4.8 shows the hierarchical structure explained in Section 4.3. The supervisor is connected to the production lines via two communication channels: the supervisor's ports *supin* and *supout* is connected to the production lines' *supout* and *supin* ports. A production line is connected via two channels; via an *operations* channel and a *material* channel. The latter was not explained in Section 4.4; it allows communication of the materials to be used for input and output operations.

Note that all process classes are connected to a TraceWriter process class via their *tracing* port. These ports and this process class are used for the performance analysis explained in Section 4.5. This channel is shown in light grey in Figure 4.8.

A picture of the interaction between the process classes of the POOSL simulation is shown in Figure 4.9. It shows the messages being exchanged between the process classes over time. Figure 4.9 specifically shows the interaction between the book maker production line and the book maker output module and between the book maker output module and the trace writer.

Figure 4.9: Interaction between process classes of a POOSL simulation

4.5 Performance analysis

The POOSL simulation model can be executed to generate a Gantt chart that illustrates the behavior of the HMLV production system. These Gantt charts are created using Eclipse TRACE4CPS [22] (see also https://eclipse.dev/trace4cps/).

The Gantt chart in Figure 4.10 is a resource view (see Section 4.3) showing the result of producing two booklets: one job represents a three-page booklet and another a four-page booklet. After printing the sheets of one booklet, the transporter module transports the sheets from the printer production line to the book maker production line. The rectangles in the Gantt chart represent the execution of operations, the arrows the communication between supervisor, production lines and equipment modules.

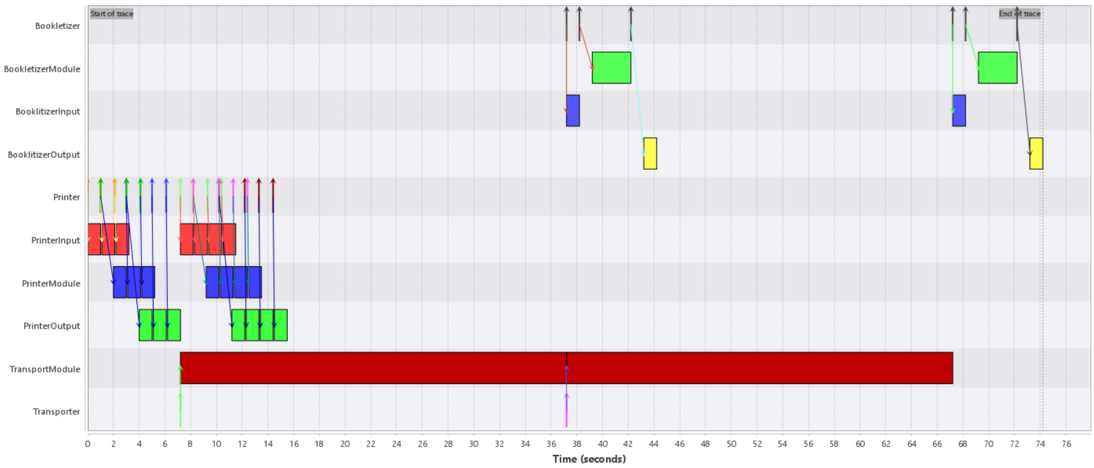

Figure 4.10: Gantt chart produced by the production system simulation.

In this example, the transport time is the dominant factor: both the printer and the book maker are idling most of the time. The transport bottleneck can partially be mitigated by combining both booklets in a single job. A combined job would involve only one transport operation. Moreover, the delay between printing the first and second set of sheets and the delay between making the first and second booklet would be much smaller.

5 Conclusion

As customers are asking for tailored products, there is a growing need to move from mass production towards customization. Realizing this customization means that high-mix low-volume production systems are gaining importance. As these HMLV production systems need to deliver a wide range of products, they have much more flexibility. As the customer orders are typically also much smaller, there are many new planning and scheduling decisions that need to be made on different time scales.

Characterizing HMLV production systems and subsequently modeling and analyzing them is key to ensure that production facilities dealing with HMLV can operate efficiently, and quickly adapt to continuously changing circumstances. Reflecting on the two research questions, we come to the following conclusions.

- **Research Question 1:** How to specify high-mix low-volume production systems?

Like any production system, a starting point for describing HMLV production systems is the 4M's model considering the characteristics of manpower, machine, method, and material. The differentiating factor of HMLV is however the flexibility, which can be linked to any of these elements. In Chapter 3 we have introduced a framework to define the concept of flexibility, and how this can be linked to flexibility on different time scales and to the different operational levels.

- **Research Question 2:** How to effectively allocate jobs to available equipment considering job and equipment constraints and uncertainties?

In this report, we have described our performance engineering methodology that can be used to effectively allocate jobs to available equipment. In our methodology, we can consider the detailed timing behavior of production lines, as well as the timing behavior resulting from multiple production lines that are loosely connected, for example via buffers. The current solution presents a simulation model to describe and analyze loosely connected production systems. In future research, we would like to extend this work by generating the simulation models from a domain model. This domain model will be an extension of the current domain model available to describe tightly coupled production systems. We also want to attach an optimizer that can compute schedules that ensure the effective allocation of jobs onto available equipment. This optimizer will use the material, job, and equipment parts of the specification, and compute an optimal allocation instead of an allocation that is manually specified. Reflecting on the second research question, we conclude that we have a vision and a partial solution, but not yet all parts to answer the question.

Next to further developing our performance engineering methodology, we want to validate what benefits it brings to have this detailed timing behavior of production lines captured, and compare our approach against other well-known approaches using commercial tooling like Anylogic [24][5] or FlexSim [25][6].

[5] https://www.anylogic.com
[6] https://www.flexsim.com

6 References

[1] O. Vermeulen, "Simulating High-Mix Low-Volume Manufacturing Systems in FlexSim," Eindhoven University of Technology, 2025.

[2] J. van Maren, "Simulating HMLV printing systems in AnyLogic," Eindhoven University of Technology, 2025.

[3] B. Buetfering, A. van der Stoel, P. K. Sriram, H. Van Landeghem, I. Ickerott, V. Wiegel and E. Alfnes, "Lean in high variety, low volume production environments - a literature review and maturity model," in *Proceedings of EurOMA*, 2016.

[4] R. Hayes and S. Wheelwright, "Link Manufacturing Process and Product Life Cycles," *Harvard Business Review,* 1 1979.

[5] S. Rajagopalan, "Make to order or make to stock: model and application," *Management Science,* vol. 48, no. 2, pp. 241-256, 2002.

[6] J. Gosling and M. M. Naim, "Engineer-to-order supply chain management: A literature review and research agenda," *International journal of production economics,* vol. 122, no. 2, pp. 741-754, 2009.

[7] M. Latzel, "High Mix / Low Volume: Possibilities & Limitations," ESSERT Robotics, [Online]. Available: https://www.essert.com/blog/hmlv/high-mix-low-volume/. [Accessed 30 01 2025].

[8] C. N. Potts and M. Y. Kovalyov, "Scheduling with batching: A review," *European Journal of Operational Research,* vol. 120, no. 2, pp. 228-249, 2000.

[9] W. T. Lunardi, E. Birgin, P. Laborie and D. Ronconi, "Mixed integer linear programming and constraint programming models for the online printing shop scheduling problem," *Computers & Operations Research,* vol. 123, 2020.

[10] D. Upton, "The management of manufacturing flexibility," *California management review,* vol. 36, no. 2, pp. 72-89, 1994.

[11] W. Terkaj, T. Tolio and A. Valente, "Focused Flexibility in Production Systems," in *Changeable and Reconfigurable Manufacturing Systems*, 2009, pp. 47-66.

[12] L. Van De Ginste, J. Goos, M. Schamp, A. Claeys, S. Hoedt, K. Bauters, A. Biondi, E.-H. Aghezzaf and J. Cottyn, "Defining flexibility of assembly workstations through the underlying dimensions and impacting drivers," *Procedia Manufacturing,* vol. 39, pp. 974-982, 2019.

[13] N. Slack, A. Brandon-Jones and R. Johnston, Operations Management (8th edition), Pearson, 2016.

[14] S. Nakajima, Introduction to TPM: Total Productive Maintenance, Cambridge, MA: Productivity Press, 1988.

[15] L. Van De Ginste, E.-H. Aghezzaf and J. Cottyn, "The role of equipment flexibility in Overall Equipment Effectiveness (OEE)-driven process improvement," *Procedia CIRP,* vol. 107, pp. 289-294, 2022.

[16] D. Scott and R. Pisa, "Can overall factory effectiveness prolong Moore's Law?," *Solid State Technology,* vol. 41, no. 3, pp. 75-82, 1998.

[17] L. M.-y. A. Pintelon and P. Nganga Muchiri, "Performance measurement using overall equipment effectiveness (OEE): Literature review and practical application discussion," *International Journal of Production Research,* vol. 46, no. 13, pp. 3517-3535, 2008.

[18] K. M. N. Muthiah and S. H. Huang, "Overall throughput effectiveness (OTE) metric for factory-level performance monitoring and bottleneck detection," *International Journal of Production Research,* vol. 45, no. 20, pp. 4753-4769, 2007.

[19] J. Verriet, "TNO 2023 R12451 Compositional Performance Prediction for Tightly Coupled Manufacturing Systems," TNO, Eindhoven, 2023.

[20] J. Verriet, "TNO 2024 R12174 Performance Prediction and Optimisation for Tighly Coupled Manufacturing Systems," TNO, 2024.

[21] J. Voeten, POOSL: An object-oriented specification language for the analysis and design of hardware/software systems, Eindhoven University of Technology, 1995.

[22] M. Hendriks, J. Verriet and T. Basten, "Visualization, transformation, and analysis of execution traces with the eclipse TRACE4CPS trace tool," *International Journal on Software Tools for Technology Transfer,* vol. 26, pp. 101-126, 2024.

[23] IEC, *IEC 62264-1:2013 Enterprise-control system integration Part 1: Models and terminology,* 2013.

[24] A. Borshchev, "Multi-method modelling: AnyLogic," in *Discrete-event simulation and system dynamics for management decision making*, 2014.

[25] W. Nordgren, "FlexSim simulation environment," in *Proceedings of the 2003 Winter Simulation Conference, Vol. 1, IEEE*, 2003.

ICT, Strategy & Policy

High Tech Campus 25
5656 AE Eindhoven
www.tno.nl

innovation
for life