



HAL
open science

Evaluating the scalability of reconfigurable manufacturing systems at the design phase

Audrey Cerqueus, Xavier Delorme

► **To cite this version:**

Audrey Cerqueus, Xavier Delorme. Evaluating the scalability of reconfigurable manufacturing systems at the design phase. *International Journal of Production Research*, 2023, 61 (23), pp.8080-8093. 10.1080/00207543.2022.2164374 . hal-03971200

HAL Id: hal-03971200

<https://hal.science/hal-03971200>

Submitted on 15 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Evaluating the scalability of reconfigurable manufacturing systems at the design phase

Audrey Cerqueus^a and Xavier Delorme^b

^a; IMT Atlantique, LS2N, UMR CNRS 6004, F-44307 Nantes, France; ^bMines Saint-Etienne, Univ Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158 LIMOS, F-42023 Saint-Etienne, France

ARTICLE HISTORY

Compiled December 15, 2023

ABSTRACT

To face market volatility, reconfigurable manufacturing systems (RMS) aim to efficiently and cost-effectively react to changes. We focus on one characteristic of RMS: the scalability (ability to adapt the volume of throughput). In the literature, the only few indicators for scalability are not always formally defined and usually only consider a partial view of scalability. Moreover, most of them are actually more suited for the configuration planning rather than for the design. However, the design of the RMS has a high impact on its scalability. We propose the first combinatorial definition of this problem and a new measure to fully assess the scalability of a system at the design phase. This measure, based on a multi-objective approach, can assess the scalability of single-product manufacturing systems, analyzing all configurations that it can implement. We present numerical experiments to compare this indicator with a state-of-the-art scalability indicator and with some classical production line design indicators, and we show that future research should focus on scalability as a specific criterion to optimize during the design of a RMS. In addition, the results obtained allow us to infer some managerial insights on the best levers to use when performing a reconfiguration for scalability purpose.

KEYWORDS

Scalability; Reconfigurable manufacturing systems; Multi-objective indicator; Combinatorial analysis; Line Balancing

1. Introduction

Nowadays, one of the main challenge industrial companies are facing is to design manufacturing systems that can deal with the volatile market and uncertain future (see e.g., Aheleroff et al., 2021; Dolgui et al., 2022). They call for reduced time-to-market as well as the ability to deal efficiently with numerous ramp-up phases. Koren

This is an Accepted Manuscript version of the following article, accepted for publication in International Journal of Production Research: “ Audrey Cerqueus, Xavier Delorme (2023) Evaluating the scalability of reconfigurable manufacturing systems at the design phase, International Journal of Production Research 61:23 , 8080-8093, DOI: 10.1080/00207543.2022.2164374 ”. It is deposited under the terms of the Creative Commons Attribution-Non Commercial License 2023 CC-BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CONTACT audrey.cerqueus@imt-atlantique.fr
delorme@emse.fr

et al. (1998) have introduced the Reconfigurable Manufacturing Systems (RMS) as an answer to this need of reactivity. Basically, RMS aim at managing shorter product lifecycles while keeping longer production system lifecycles. To achieve this goal, RMS relies on six key features (Koren et al., 2018):

- Modularity, i.e. compartmentalization of operational functions into modules
- Integrability to rapidly and efficiently integrate modules
- Diagnosability allowing the automatic identification of problems in the production system and real-time monitoring the product quality
- Customization to produce different parts in a family,
- Convertibility to enable the system to change for new products
- Scalability which is the ability to adapt the volume of production

The recent emergence of Industry 4.0 enables the implementation of such systems by providing new solutions to implement these features (Zheng et al., 2018; Ivanov et al., 2019; Napoleone et al., 2022). For example, Plug & Produce Cyber-physical systems bring modularity and integrability, Internet of Things (IoT) and Digital Twins work together to improve diagnosability, Collaborative robot (Cobot) and additive manufacturing can provide the flexibility needed for the customization. . .

Beside the use of new digital technologies, the decision support systems used by companies working with RMS have to be thought accordingly (see e.g., El Alaoui El Abdellaoui et al., 2020). According to Andersen et al. (2017) and Dolgui et al. (2022), due to the multiple variants and generations of product over the system life cycle, the design of RMS is a challenge compared to the design of traditional manufacturing systems. Koren et al. (2017) states that scalability might be the most important feature to deal with the uncertain demand or ramp-up phase. According to Putnik et al. (2013), the scalability in manufacturing systems can be defined as “the capacity for adding or removing the resources in a cost-effective manner, in order to adjust the production capacity on a system in steps or stages”. Contrary to convertibility, it concerns only the throughput volume and not a change in the specifications of the product. The scalability assessment of a system must integrate components measuring the ability to quickly adapt the system, with incremental steps, and cost-effective changes, to answer to the need to provide at any time an efficient configuration with the exact capacity needed. Such an adaptation is particularly useful during ramp-up and product end-of-life phases.

RMS are typically composed of serial stages or stations with several identical parallel resources, joined by a gantry and a conveyor to move the parts in this grid system. The resources are generally Computer Numerical Control (CNC) or Reconfigurable Manufacturing Tools (RMT) but can also be of other types (for instance workers or cobots). Figure 1 schematizes this typical organization of RMS. Such layouts are sometimes denoted in the literature as parallel-serial lines with crossover (see e.g., Freiheit et al., 2004).

A configuration is defined as the state of the system, determined by the resources used, in terms of functionality, performance and number. Two levers can be involved in a reconfiguration: adding or removing parallel resources on the stations, or processing a full reconfiguration of the system by changing the tools used in RMT or CNC machines (Putnik et al., 2013). In the former approach, the assignment of operations to stages stays the same, thus the only organizational change in the system is the addition of resources and their integration on the gantry. The latter enables us to change the balancing of the system, implying a longer setup time, as many resources can be impacted and the system must be purged of semi-finished products. In both cases,

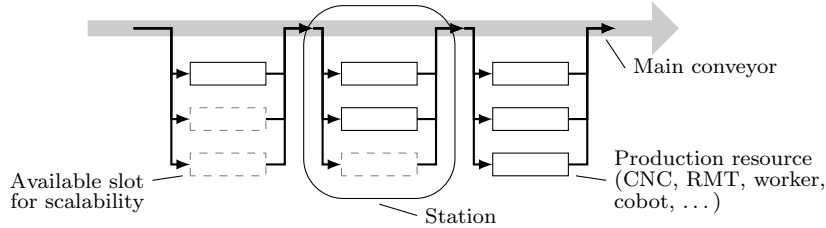


Figure 1. RMS layout as described by Koren et al. (2018)

the choice of equipment is a tactical decision that should be made when designing the system to enable operational reconfigurations.

In this article, we investigate the scalability of RMS by considering the two aforementioned levers. In particular, we consider two research questions:

- (1) What is the relationship between the balancing of an RMS and its scalability, both for the choice of the initial configuration and for further reconfigurations?
- (2) How to compare the scalability of different possible initial configurations of manufacturing system taking into account the policy to be applied for future reconfigurations and considering an unknown demand evolution?

In order to address these questions, two main contributions are provided. First, we give a formalization of the combinatorial optimization problem associated with the balancing of an RMS such as defined by Koren et al. (1998), and we analyze its complexity, which appear to be polynomial. Second, we propose a new measure to assess the scalability of a single-product system since the design phase, based on a multi-objective approach. In addition, the experiments performed to validate these contributions allow us to infer some interesting managerial insights for practitioners dealing with the design or the reconfiguration of an RMS.

The remainder of the paper is organized as the following: Section 2 presents the related literature, Section 3 formalizes the combinatorial definition of the problem, the proposed scalability measure is explained in Section 4 and experimental results are analyzed in Section 5. Section 6 concludes this paper and discusses few perspectives.

2. Related literature

The scalability of manufacturing systems has been studied in different research communities (optimization, model-based systems engineering,...). Since this paper aims to formalize the problem optimizing the scalability of a single-product manufacturing system from the design phase, we choose to focus this literature review on papers dealing with optimization approaches and scalability evaluation.

If the design of manufacturing systems has generated a lot of publications on various balancing and facility layout problems, the specific characteristics of RMS make it an original optimization problem requiring new methods. The first methods proposed were heuristics (Youssef and ElMaraghy, 2006; Dou et al., 2011; Essafi et al., 2012; Lahrichi et al., 2021) but exact models and algorithms have also been presented (Essafi et al., 2010; Borisovsky et al., 2014). A recent review on the main optimization problem associated with RMS has been provided by Yelles-Chaouche et al. (2021). In addition, works on robotic assembly line balancing (see e.g., Borba et al., 2018) or parallel stations (see Aguilar et al., 2020, for a recent survey) present some similar features

with the case of RMS.

However, all these works only focus on the generation of a single configuration minimizing the cost and do not take into account the possibility to further scale the system.

Scalability in the context of RMS has been studied on two different levels: the management of the scalability (on an operational level) and the design of a scalable system (on a tactic level).

Regarding the management of the scalability, Deif and ElMaraghy (2006); Deif and ElMaraghy (2007) were among the first to consider several reconfigurations, assessing different reconfiguration policies on various scenarios of demand evolution. Wang and Koren (2012) started to consider successive reconfigurations with a heuristic method minimizing the number of machines required. An extension of this method has been presented in Koren et al. (2017), integrating buffers in RMS. Several optimization methods (Hees et al., 2017; Hu et al., 2017, e.g.) and simulation-based methods (Hsieh, 2018; Gola et al., 2021, e.g.) have been developed to handle the scalability during the planning process. It is interesting to note that, even if some authors working on this question do not accurately explain the nature of the changes performed during a reconfiguration, most of them consider reconfigurations based on a new line balancing.

Some works have focused on the scalability since the design phase by analysing the characteristics allowing an RMS to be scalable. Koren et al. (1998) and Spicer et al. (2002) started to study the impact of various system configurations on throughput capacity and scalability. Koren and Shpitalni (2010) went further and discussed the relationship between the structure of RMS and the number of configurations available. A similar idea was considered by Delorme et al. (2016) who suggested to evaluate the solution of their heuristic by generating the set of trade-off configurations between capacity throughput and number of resources needed. Son et al. (2001) questioned the link between the balancing of a production system and its productivity and scalability. They highlighted that unbalanced RMS can sometimes allow us to obtain smaller steps of capacity changes. More recently, Napoleone et al. (2019) introduced a classification of the main root causes leading to scalability and convertibility, Rösio et al. (2019) provided a list of key enablers for scalability including the possibility to change the line balancing and to add machines, and Beauville Dit Eynaud et al. (2022) presented a framework for the design of RMS and a reconfiguration strategy to answer to a specific demand evolution. Finally, some works tried to integrate the design and the planning problems. In Moghaddam et al. (2020), the problem of designing multi-product and scalable RMS for multiple production periods has been addressed, minimizing design and reconfiguration costs while fulfilling demand. The authors presented two approaches: an up- and downgrading method based on approximate demand in each period; and RMT selections and reconfigurations based on long-term demand forecast. A similar problem of design and configuration planning was considered in Cerqueus et al. (2020) where the authors used the scalability of a mono-product RMS to minimize the energy cost in a time-of-use pricing scheme.

Finally, the evaluation of the scalability of an RMS has received little attention. In Wang and Koren (2012), a first metric based on the smallest possible incremental capacity change has been presented to evaluate a priori the scalability on the first reconfiguration. However this metric is dependent on the current state of the system and does not allow to take into account the subsequent reconfigurations. In addition, the definition of the proposed metric can be confusing because incremental capacity can be defined either as the gap in terms of resources used or in terms of capacity throughput. Both definitions are equivalent in the examples presented in the paper but

it is not true in the general case. In addition, the authors showed that the metric can be coupled with a heuristic simulating the future reconfigurations (i.e. rebalancing of the line and addition of resources) to forecast the production throughput gain of different configurations. Based on the experimental results obtained, the authors observed that a lower number of stages leads to higher throughput and gain, mainly because it leads to a more reliable system. Nishith et al. (2013) has also proposed a slightly different metric based on the same principle of potential capacity increase but they provided very few details on the calculation. Indeed, Putnik et al. (2013) stated that new performance measures of scalability were still needed. The first evaluation of the scalability based on the set of possible configurations identified has been proposed by Wang et al. (2017). Their function takes into account the gap between the minimum and maximum throughput among all the configurations, as well as the average gap between the configuration and the effort needed to go from one configuration to another (time, cost, ...). However, the authors do not indicate how to identify the set of available configurations and the numerous parameters used in the function. More recently, some authors have proposed to extend the concept of scalability either to reconfigurable supply chain (Zidi et al., 2021) or to process manufacturing (Accorsi et al., 2021). Important to note is that the metrics proposed in these papers are similar to those proposed for RMS, the former being defined as a combination of throughput capacity amplitude and of latency, and the latter being calculated by the simulation of a given curve of demand.

Overall, we can observe that scalability is rarely considered alone and most authors consider it together with convertibility. Even if both features do not have the same purpose, this choice may be explained by the fact that most authors use the same levers (i.e. adding resources and rebalancing operations) to deal with them. This observation is confirmed by Maganha et al. (2018) which report that there is a confusion among industrials between these two features.

3. Combinatorial definition of the problem

We are considering the design of a paced reconfigurable manufacturing system. A set N of operations to process is given and each operation $n \in N$ has a processing time t_n . The set of operations have to be partitioned into a sequence of stations (i.e., we have a line balancing problem), with n_0 and m_0 the upper bound on the number of operations which can be assigned to the same station and on the number of stations, respectively. The assignment of operations to stations should respect a set $Prec \subset N \times N$ of precedence relations. The workload of a station is defined as the sum of the processing times of the operations assigned and the production system is paced at a takt time defined according to the station with the highest operating time. In order to increase the production level, each station can be equipped with at most r_0 resources working in parallel and the whole sequence of stations can be duplicated while the total number of resources used does not exceed \bar{r} . The data are summarized in Table 1.

All definitions presented in this section will be illustrated on the same didactic example composed of 7 operations ($N = \{1, \dots, 7\}$), given by the precedence graph in Figure 2 on which the processing times are given above the node representing the operation. We take $n_0 = 4$, $m_0 = 4$, $r_0 = 3$ and $\bar{r} = 7$.

We call a sequence $B = (S_1, \dots, S_m)$, $S_k \subseteq N, \forall k \in 1, \dots, m$ a feasible balancing (see Figure 3) if all operations are assigned once ($\bigcup_k S_k = N$), the precedence constraints are respected ($(i, j) \in Prec \wedge i \in S_k \Rightarrow j \in \bigcup_{k' \geq k} S_{k'}$), and the number of operations

N	The set of operations
t_n	Processing time of the operation n
n_0	Maximum number of operations assigned to a same station
m_0	Maximum number of stations
$Prec$	Set of precedence constraints
r_0	Maximum number of resources assigned to a same station
\bar{r}	Maximum total number of resources assigned in the system

Table 1. Problem data

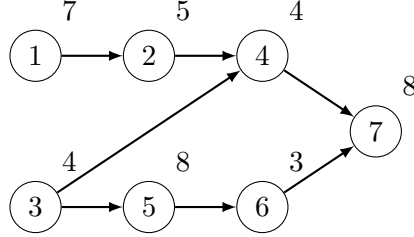


Figure 2. Precedence graph and processing times of the illustrative example

assigned per station as well as the number of stations do not exceed the upper bounds n_0 and m_0 respectively. In the following, the notation $B.S_k$ (resp. $B.m$) will be used to refer to the k^{th} component (resp. the cardinality) of a balancing B (the same notation will be used for all other sequences or tuples). Let \mathcal{B} denote the set of all feasible balancing. We can consider each S_k as a subset of all n -tuples (e_1, \dots, e_n) such that $1 \leq n \leq n_0$ and e_i are all distinct elements from the set N , and the set \mathcal{B} as a subset of all m -tuples of distinct S_k such that $1 \leq m \leq m_0$. Alternatively, the set \mathcal{B} is a subset of all $|N|$ -dimensional integer vectors with values in the range $[1, m_0]$. Therefore, $|\mathcal{B}| \leq O(\min(|N|^{n_0 m_0}; m_0^{|N|}))$.

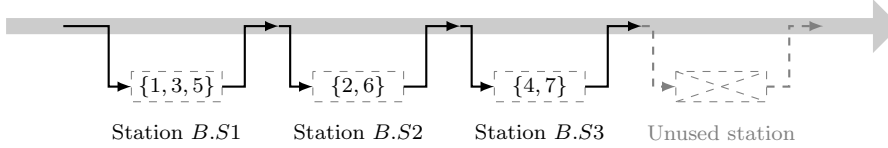


Figure 3. Example of a balancing $B \in \mathcal{B}$ for a problem with 7 operations and $m_0 = 4$

We call a sequence of 2-tuples $L = ((S_1, r_1), \dots, (S_m, r_m))$ a feasible line (see Figure 4) if (S_1, \dots, S_m) is a feasible balancing, $r_k \in 1, \dots, r_0, \forall k \in 1, \dots, m$ and $\sum_{k=1}^m r_k \leq \bar{r}$. Let \mathcal{L} denote the set of all feasible lines and $\mathcal{L}(B) \subseteq \mathcal{L}$ denote the subset of lines using the balancing B , i.e. $\mathcal{L}(B) = \{L \in \mathcal{L} | L.m = B.m \text{ and } L.S_k = B.S_k, \forall k \in 1, \dots, m\}$. The components r_k from $\mathcal{L}(B)$ correspond to a subset of all m -dimensional integer vectors, $m \in 1, \dots, m_0$, with values in the range $[1, r_0]$, which implies that $|\mathcal{L}(B)| \leq O(r_0^{m_0})$. Reciprocally, let the singleton $\mathcal{B}(L)$ denote the balancing used by the line L .

The performance of each line L is a 2-dimensional vector $z(L) = (\tau(L), \rho(L))$ which can be calculated according to (1), where $\tau(L)$ (resp. $\rho(L)$) corresponds to the takt time (resp. the total number of resources used).

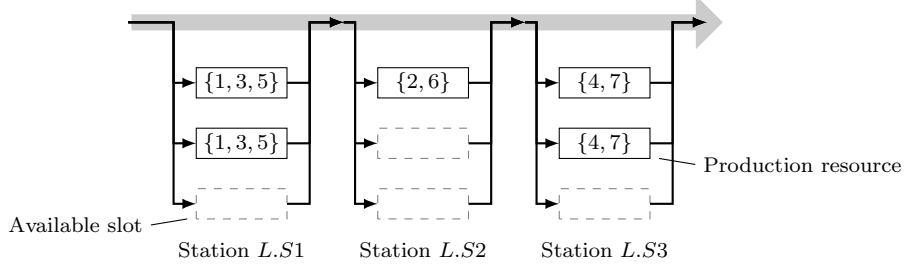


Figure 4. Example of a line $L \in \mathcal{L}$ for a problem with 7 operations and $r_0 = 3$. Here we have, $\tau(L) = \max \{(7 + 4 + 8)/2; (5 + 3)/1; (4 + 8)/2\} = 9.5$ and $\rho(L) = 5$.

$$\begin{cases} \tau(L) = \max_{k \in 1, \dots, L.m} \left\{ \left(\sum_{n \in L.S_k} t_n \right) / L.r_k \right\} \\ \rho(L) = \sum_{k=1}^{L.m} L.r_k \end{cases} \quad (1)$$

When comparing two lines from an economical perspective, the lower number of resources and the lower takt time are preferred by decision-maker, so it is a bi-objective problem and we can use the Pareto dominance (Pareto, 1896): a line L dominates another line L' if $\tau(L) \leq \tau(L')$, $\rho(L) \leq \rho(L')$ and we have $\tau(L) < \tau(L')$ or $\rho(L) < \rho(L')$, denoted $z(L) \preceq z(L')$.

We can focus only on the non-dominated lines. Let $\mathcal{L}^*(B) \subseteq \mathcal{L}(B)$ denote the subset of feasible lines using the balancing B which are not dominated by another line using B . For any given B , the set $\mathcal{L}^*(B)$ can be constructed by starting from the line $L \in \mathcal{L}(B)$ with $L.r_k = 1, \forall k \in 1, \dots, L.m$ and then increasing iteratively of one unit the number of resources of the bottleneck station (i.e., $\operatorname{argmax}_{k \in 1, \dots, L.m} \{ \sum_{n \in L.S_k} t_n / L.r_k \}$)² until either the upper bound r_0 or \bar{r} is attained. Since the number of steps of this process is bounded by \bar{r} and each step can be computed in $O(n_0 m_0)$, we have $|\mathcal{L}^*(B)| \leq \bar{r}$ and the process can be done in constant time since n_0, m_0 and \bar{r} are given constants.

We call a λ -tuple $C = (L_1, \dots, L_\lambda)$ a feasible configuration (see Figure 5) if all $L_q, q \in 1, \dots, \lambda$ are feasible lines which share the same balancing (i.e., $\forall q, q' \in 1, \dots, \lambda, L_q.m = L_{q'}.m$ and $\forall k \in 1, \dots, L_q.m, L_q.S_k = L_{q'}.S_k$) and $\sum_q \sum_k L_q.r_k \leq \bar{r}$. Let \mathcal{C} denote the set of all feasible configurations and $\mathcal{C}(B) \subseteq \mathcal{C}$ denote the subset of configurations using the balancing B , i.e. $\mathcal{C}(B) = \{C \in \mathcal{C} | C.L_1.m = B.m \text{ and } C.L_1.S_k = B.S_k, \forall k \in 1, \dots, C.L_1.m\}$. Since all the components of $\mathcal{C}(B)$ are elements from the set $\mathcal{L}(B)$ and $\lambda \leq \bar{r}$, we can deduce that $|\mathcal{C}(B)| \leq O(r_0^{m_0 \bar{r}})$. Reciprocally, let the singleton $\mathcal{B}(C)$ denote the balancing used by the configuration C .

Similarly to the lines, the performance of a configuration C is a 2-dimensional vector $z(C) = (\tau(C), \rho(C))$ which can be calculated according to (2), where $\tau(C)$ (resp. $\rho(C)$) corresponds to the takt time (resp. total number of resources used). We can naturally

²Whenever there is not a single bottleneck station, we can arbitrarily pick one of them since the next steps will consider the others. In this case, the line's takt time will only decrease when the number of resources will have been increased in all of the aforementioned stations. The set of solutions generated during these iterations will differ according to the sequence of bottleneck stations considered (they would be denoted as equivalent solutions in multi-objective optimization) but all these solutions will be dominated since they have the same takt time but a higher number of resources.

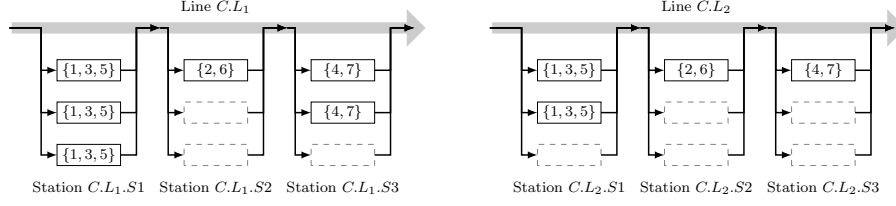


Figure 5. Example of a configuration $C \in \mathcal{C}$ with 2 lines for a problem with 7 operations and $r_0 = 3$. Here we have $\tau(C) = \frac{8 \times 12}{8 + 12} = 4.8$ and $\rho(C) = 10$.

extend the notion of Pareto-dominance to the configurations and only consider the subset $\mathcal{C}^*(B) \subseteq \mathcal{C}(B)$ of feasible configurations using the balancing B which are not dominated by another configuration using B . Since such a configuration will only use non-dominated lines from $\mathcal{L}^*(B)$, we have $|\mathcal{C}^*(B)| \leq O(\bar{r}^{\bar{r}})$. Finally, let $\mathcal{C}^* \subseteq \mathcal{C}$ denote the subset of non-dominated configurations, $\mathcal{C}^* \subseteq \bigcup_{B \in \mathcal{B}} \mathcal{C}^*(B)$.

$$\left\{ \begin{array}{l} \tau(C) = \frac{1}{\sum_{q=1}^{C.\lambda} \frac{1}{\tau(L_q)}} = \frac{\prod_{q=1}^{C.\lambda} \tau(L_q)}{\sum_{q=1}^{C.\lambda} \left(\prod_{u \in \{1, \dots, C.\lambda\} \setminus q} \tau(L_u) \right)} \\ \rho(C) = \sum_{q=1}^{C.\lambda} \rho(L_q) \end{array} \right. \quad (2)$$

Changing the system state from one configuration to another in order to answer to an evolution of market demand is called a reconfiguration. Such changes generate costs and require a reconfiguration time, which are usually non-decreasing functions of the magnitude of the change, i.e. the gap between the old configuration \hat{C} and the new one C . The minimal change occurs when the same balancing is kept ($\mathcal{B}(C) = \mathcal{B}(\hat{C})$) since such a reconfiguration does not impact the tools or modules used. On the contrary, whenever there is a change in the balancing, some tools or modules must be replaced. We call \mathcal{N}_p the set of neighborhoods (3) based on the number of operations movements from one initial balancing. Starting from an initial configuration $C \in \mathcal{C}$ these neighborhoods can thus define the subset $\bigcup_{B \in \mathcal{N}_p(\mathcal{B}(C))} \mathcal{C}(B)$ of possible future configurations reachable over the lifecycle of the RMS for a given threshold p of changes allowed.

$$\left\{ \begin{array}{l} \mathcal{N}_0(B) = \{B\} \\ \mathcal{N}_1(B) = \{B' \in \mathcal{B} \mid B'.m = B.m \text{ and } \exists n \in N, k_1, k_2 \in 1, \dots, B.m, \\ \quad k_1 \neq k_2, B'.S_{k_1} = B.S_{k_1} \cup \{n\}, B'.S_{k_2} = B.S_{k_2} \setminus \{n\}, \\ \quad B'.S_k = B.S_k \forall k \neq k_1, k_2\} \cup \{B\} \\ \mathcal{N}_p(B) = \{B' \in \mathcal{B} \mid B'.m = B.m \text{ and } \exists ((n_1, k_1^1, k_2^1), \dots, (n_p, k_1^p, k_2^p)), \\ \quad \forall i \in 1, \dots, p, n_i \in N, k_1^i, k_2^i \in 1, \dots, B.m, \forall k \in 1, \dots, B.m, \\ \quad B'.S_k = B.S_k \cup \{n_i \mid k_1^i = k\} \setminus \{n_i \mid k_2^i = k\}\} \\ \mathcal{N}_{|N|}(B) = \{B' \in \mathcal{B} \mid B'.m = B.m\} \end{array} \right. \quad (3)$$

Given the balancing B_1 in Figure 3, the neighborhood $\mathcal{N}_1(B_1)$ is composed of seven balancing:

$$\begin{array}{lll} B_2 : B_2.S_1 = \{3, 5\}, & B_2.S_2 = \{1, 2, 6\}, & B_2.S_3 = \{4, 7\}, \\ B_3 : B_3.S_1 = \{1, 2, 3, 5\}, & B_3.S_2 = \{6\}, & B_3.S_3 = \{4, 7\}, \\ B_4 : B_4.S_1 = \{1, 3, 5\}, & B_4.S_2 = \{6\}, & B_4.S_3 = \{2, 4, 7\}, \\ B_5 : B_5.S_1 = \{1, 3, 5\}, & B_5.S_2 = \{2, 4, 6\}, & B_5.S_3 = \{7\}, \\ B_6 : B_6.S_1 = \{1, 3\}, & B_6.S_2 = \{2, 5, 6\}, & B_6.S_3 = \{4, 7\}, \\ B_7 : B_7.S_1 = \{1, 3, 5, 6\}, & B_7.S_2 = \{2\}, & B_7.S_3 = \{4, 7\}, \\ B_8 : B_8.S_1 = \{1, 3, 5\}, & B_8.S_2 = \{2\}, & B_8.S_3 = \{4, 6, 7\}. \end{array}$$

According to these definitions, the optimization problem of maximizing the scalability of an RMS at the design stage can be expressed as (4), where f_s refers to a scalability metric evaluating the set of configurations available for future reconfiguration and p is a parameter defining the maximal level of changes allowed. Considering that checking the feasibility of each balancing can be done in polynomial time, and as long as the function f_s can also be calculated in polynomial time, an approach based on the enumeration of all feasible balancing and configurations will be polynomial, i.e. theoretically efficient (Garey and Johnson, 1979), since we have $|\mathcal{B}| \leq O(|N|^{n_0 m_0})$, $|\mathcal{C}| \leq O(|N|^{n_0 m_0} r_0^{m_0 \bar{r}})$ and $|\mathcal{C}^*| \leq O(|N|^{n_0 m_0} \bar{r}^{\bar{r}})$ with n_0 , m_0 and \bar{r} being given constants. However, in practice these constants usually take large values and such an enumeration approach will only be applicable for small problems.

$$\max_{B \in \mathcal{B}} f_s \left(\bigcup_{B' \in \mathcal{N}_p(B)} \mathcal{C}(B') \right) \quad (4)$$

Note that the design of a dedicated manufacturing system (DMS) would correspond to a specific case where the set of configurations available would be restricted within a subset of $\mathcal{C}_{DMS} \subset \mathcal{C}$ such that only one resource is assigned per station ($C.L_q.r_k = 1, \forall q \in 1, \dots, C.\lambda, k \in 1, \dots, C.L_q.m$).

After the design phase, each reconfiguration will correspond to another optimization problem (5), where \hat{C} is the initial configuration of the RMS and T is the new takt time deduced from current market demand. When choosing a new configuration from

the available set $\mathcal{N}_p(\mathcal{B}(\hat{C}))$, there can be several configurations with the same minimal value of p but a rational decision-maker will prefer to select the non-dominated one. As a consequence, we can discard the dominated configurations from the search space.

$$\begin{aligned} \min_{C \in \mathcal{C}(B)} p \\ \text{s.t.} \quad & \left| \begin{array}{l} B \in \mathcal{N}_p(\mathcal{B}(\hat{C})) \\ \tau(C) \leq T \\ p \in 0, \dots, |N| \end{array} \right. \end{aligned} \quad (5)$$

In the next section, we will focus on the definition of the scalability measure f_s which both take into account the whole set of reachable configurations and is calculable in polynomial time.

4. Hypervolume based indicator for scalability

Based on the literature review, we extracted three main properties to evaluate the scalability level of a set of reachable configurations:

- (1) all the configurations should be as economically efficient as possible (i.e., we are seeking configurations with the lowest possible values for both the number of resources and the takt time).
- (2) they should cover a large range of takt time to be able to answer both small and large demands.
- (3) the configurations should be well-spread on the two-dimensional performance space (i.e., we prefer to have regular steps to increase the capacity throughput incrementally).

The first one is omnipresent as soon as we consider the design of a manufacturing system. The second one is linked to the work of Wang et al. (2017) taking into account the gap between the minimum and the maximum throughput. The third one comes from Wang and Koren (2012) searching for the smallest incremental capacity change as well as the adaptation “in steps” in the definition of scalability of Putnik et al. (2013). These three characteristics are actually very similar to those considered in the evaluation of a multi-objective approximate method, where sets of potentially non-dominated points have to be evaluated and compared. We can thus use multi-objective indicators to assess the quality of a set of reachable configurations (and the associated values of z).

Among the numerous multi-objective indicators, the most commonly used is probably the hypervolume (Zitzler et al., 2003). For any subset of configurations $\tilde{\mathcal{C}} \subseteq \mathcal{C}$, let us consider $F(\tilde{\mathcal{C}}) = \{(\tau(C), \rho(C)), C \in \tilde{\mathcal{C}}\}$, the set of points which are the images by z of each configuration from $\tilde{\mathcal{C}}$. The hypervolume of $F(\tilde{\mathcal{C}})$ is the area between the points $(\tau(C), \rho(C))$ and a reference point P . Its formula is given by Equation (6).

$$\mathcal{H}(\tilde{\mathcal{C}}) = \mathcal{A} \left(\left(\bigcup_{C \in \tilde{\mathcal{C}}} (\tau(C), \rho(C)) - \mathbb{R}_{\leq}^2 \right) \cap (P + \mathbb{R}_{\leq}^2) \right) \quad (6)$$

\mathcal{A} is the geometrical area. \mathbb{R}_{\leq}^2 denotes the dominance cone $\{y \in \mathbb{R}^2 : y \preceq (0, 0)\} \cdot y +$

\mathbb{R}_{\leq}^2 stands for the vector addition of $y \in \mathbb{R}^2$ and any vector of \mathbb{R}_{\leq}^2 , it defines the area whose points weakly dominate y . $y - \mathbb{R}_{\leq}^2$ is defined similarly as the area whose points are weakly dominated by y .

Of course, the reference point P have to be such all $(\tau(C), \rho(C))$ have lower values on both τ and ρ . For our application, we can take

$$P = \left(\max_{K \subseteq N, |K| \leq n_0} \left\{ \sum_{j \in K} t_j \right\} + \varepsilon_\tau, \bar{r} + \varepsilon_\rho \right)$$

with $\varepsilon_\tau, \varepsilon_\rho \in \mathbb{R}^+$. Indeed, the takt time cannot be greater than the sum of the n_0 highest processing times, since there are no more than n_0 operations per stations, and the number of resources used is bounded by \bar{r} . Since the number of resources used is necessarily an integer, we can take $\varepsilon_\rho = 1$. To have a proportional gap for ε_τ , we take $\varepsilon_\tau = \frac{\max_{K \subseteq N, |K| \leq n_0} \{\sum_{j \in K} t_j\}}{\bar{r}}$.

Figure 6 graphically illustrates the notion $y + \mathbb{R}_{\leq}^2$ and the hypervolume of a set of points $F(\tilde{\mathcal{C}})$.

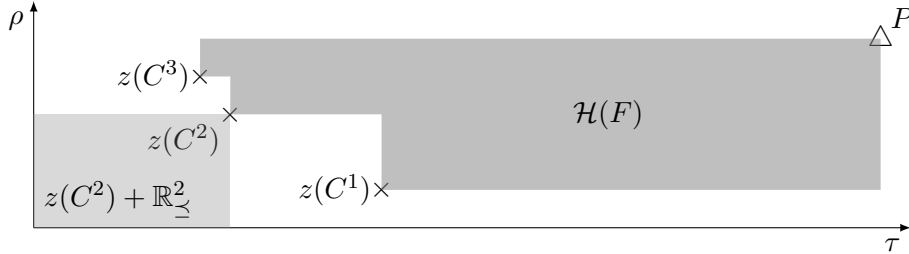


Figure 6. Illustration of the notion $y + \mathbb{R}_{\leq}^2$ and the hypervolume of a set of points $F(\tilde{\mathcal{C}})$ for a set of 3 configurations $\tilde{\mathcal{C}} = \{C^1, C^2, C^3\}$ (depicted by \times).

Then we can define the scalability function $f_s(\tilde{\mathcal{C}}) = \mathcal{H}(\tilde{\mathcal{C}})$. The hypervolume based scalability indicator f_s , in this bi-objective context, is the sum of $|\tilde{\mathcal{C}}|$ rectangles. Thus it is computed in $O(|\tilde{\mathcal{C}}|)$. Considering that the dominated configurations can be discarded from the set $\tilde{\mathcal{C}}$, the cardinality of the set is bounded by \bar{r} since the performance on the number of resources used ρ is integral and bounded by \bar{r} by definition. Thus f_s can be computed in constant time, $O(\bar{r})$, and used to solve the problem (4).

However, the values taken by the function f_s are difficult to interpret by the decision-maker and do not really allow him to have a precise idea of the quality of a set of configurations. Moreover, the values of f_s can vary significantly from one instance to another which makes it difficult to compare the results obtained for various instances. As a consequence, it can be useful to have a relative value to compare the set of reachable configurations with a reference set for the instance considered. By definition, the subset of non-dominated configurations considering any possible balancing \mathcal{C}^* has the highest value for \mathcal{H} . We can thus evaluate the quality of a set of reachable configurations by comparing $\mathcal{H}(\tilde{\mathcal{C}})$ with the value of the hypervolume of the ideal set $\mathcal{H}(\mathcal{C}^*)$. We define the relative hypervolume based indicator f'_s in Equation (7). f'_s takes values in $[0,1]$ and a low value indicates that the set of reachable configurations are close from those of the ideal set of configurations, indicating its good quality. Since $\mathcal{H}(\mathcal{C}^*)$ is constant for a given instance, maximizing f_s and minimizing f'_s are equivalent optimization problems.

$$f'_s = 1 - \frac{f_s(\tilde{\mathcal{C}})}{\mathcal{H}(\mathcal{C}^*)} \quad (7)$$

5. Experimental results

The goals of this section are to: (i) to compare the hypervolume based indicator for scalability to classical line balancing indicators and to the average of the scalability measure proposed by Wang and Koren (2012), to see if the same RMS design are considered scalable by both metrics and what makes the difference for other designs; (ii) to determine some characteristics of scalable systems, in order to help decision maker in the design of their systems, and (iii) to analyse of the impact of the length of the neighborhoods on the scalability of the system, to determine the gain in reconfiguring a system by re-balancing it.

For the instances considered, we enumerated all balancing satisfying the precedence constraints, under the following assumptions:

- There are no more than three resources on each station.
- The total number of resources is bounded by the number of operations in the system.
- There are no more than 50% of the operations assigned to the same station.

The benchmark of instances used has been extracted from Scholl (1999) and Otto et al. (2013). Since our experimental process is based on a total enumeration of all feasible balancing, we chose to focus on a restricted set of instances: the ten smallest instances of Scholl (1999) and ten from the instances with 20 operations of Otto et al. (2013). For the latter, we arbitrarily chose the instances numbered $\{50, 100, \dots, 500\}$, and checked that they are representative of the whole set of instances, in terms of values for the parameters on the structure of the graph, the order of strength (density of the precedence graph) and time distribution. All experiments were implemented in C++ and ran on an Ubuntu system using an Intel Xeon 6230 chip and 5 MO of RAM. For the numerical experiments, we limited the computational time allocated to the balancing enumeration to an hour. Only the five smallest instances completed the process under the timeout. For the others, we randomly deleted operations until the resulting instance completes the enumeration phase within the time limit. The initial and reduced size of the instances are given in Table 2.

In Table 3, the correlations between the scalability indicator f'_s and the average incremental measure of Wang and Koren (2012) are presented for all considered instances. These correlations are computed on the sets of reachable configurations extracted from the same balancing (i.e. \mathcal{N}_0).

Since the scalability indicator presented in Wang and Koren (2012) is a local evaluation of the scalability, we adapted it by computing the average of the smallest incremental capacity in percentage. Sorting the non-dominated configurations by number of resources, for two successive configurations, the increment is the difference between the inverse of the takt time of two configurations. In the following, this indicator is denoted avgScal. The two indicators avgScal and f'_s are to be minimized.

Table 3 shows that there is almost no correlation between the two scalability indicators. This reflects a major difference in the computation of the two indicators regarding what is a good increment. By definition, the measure of Wang and Koren

Inst. name	Orig. size	Red. size	Inst. name	Orig. size	Red. size
Mertens	7	-	Otto-50	20	11
Bowman8	8	-	Otto-100	20	12
Jaeschke	9	-	Otto-150	20	10
Jackson	11	-	Otto-200	20	11
Mansoor	11	-	Otto-250	20	12
Mitchell	21	13	Otto-300	20	10
Roszieg	25	13	Otto-350	20	9
Heskia	28	9	Otto-400	20	10
Buxey	29	12	Otto-450	20	15
Sawyer30	30	12	Otto-500	20	14

Table 2. Description of the benchmark instances original and reduced size (number of operations), “Otto-i” denotes the instances from Otto et al. (2013) named “instance_n=20-i”

Instance	avgScal	Instance	avgScal
Mertens	0.289	Otto-50	0.297
Bowman8	0.039	Otto-100	0.209
Jaeschke	0.251	Otto-150	0.167
Jackson	0.154	Otto-200	0.210
Mansoor	0.138	Otto-250	0.194
Mitchell	0.090	Otto-300	0.137
Roszieg	0.201	Otto-350	0.070
Heskia	0.105	Otto-400	0.224
Buxey	0.024	Otto-450	0.060
Sawyer30	0.071	Otto-500	-0.107

Table 3. Correlation of the hypervolume with the average scalability from Wang and Koren (2012)

(2012) gives better values when the difference between the takt time of two consecutive configurations is small. The increments being averaged, avgScal is lower if the spread of the τ values for the whole set of reachable configuration is tighter. On the contrary, the f'_s indicator favors the biggest increment it is possible to obtain when adding a resource, i.e. the largest possible capacity increment for the lowest possible resource cost. Moreover, the hypervolume computation gives better values if the values of ρ and τ are spread on a large range. Thus it is not surprising to see that these two indicators are not correlated.

In Table 4, we present the correlations between these scalability indicators and the classical line balancing indicators (takt time, number of stations, idle time, smoothness). We still suppose that the sets of reachable configurations are \mathcal{N}_0 , i.e. all configurations of a reachable set are based on the same balancing. All indicators are to be minimized, thus a value close to 1 indicates a strong positive correlation.

	takt time	nb stations	idle time	smoothness
avgScal	0.648	0.010	0.783	0.793
f'_s	-0.403	0.979	0.393	0.160

Table 4. Correlation between the scalability indicators and classical line balancing indicators.

Table 4 shows that there are positive correlations between avgScal and all the classical line balancing indicators, except for the number of stations. On the contrary, there is no correlation between the classical line balancing indicators and the indicator f'_s ,

except for the number of stations.

As mentioned earlier, the avgScal favors small capacity increments. When the system is well-balanced, adding one resource will lead to a small increment. Indeed, by adding a resource on the bottleneck station, the takt time will be reduced only to achieve the workload of the second bottleneck station, which is most likely to be close to the previous takt, since the stations are well-balanced. Thus the correlation of avgScal with the smoothness ratio was expected. Even if there is a correlation between avgScal and the smoothness, this correlation is not complete, which is coherent with Son et al. (2001) stating that there can be unbalanced scalable systems.

Moreover, a well-balanced system has a low idle time, thus the correlation is similar for this indicator. The indicator avgScal is also correlated with the takt time since the well-balanced systems offering the smallest capacity increments are those with a low takt time, leading to a long sequence in the sequential process of adding resources on the same line. The absence of correlation with the number of stations is due to the fact that the number of stations alone is not sufficient to obtain well-balanced systems, and for the unbalanced balancing, some increments are high. In Koren et al. (2017), the authors stated that the scalability is correlated to the number of stations, which does not reflect with the indicator avgScal.

However this assertion seems correct for the indicator f'_s , since it presents a strong positive correlation with the number of stations. This correlation can be explained by the fact that having a low number of stations allows to derive a larger set of configurations. Indeed, generally from a balancing, we can derive a configuration for any number of resources between the number of stations of the balancing and the maximum number of resources. The hypervolume being highly impacted by the number of configurations, it tends to have higher values when the number of stations in the balancing is low.

f'_s is slightly negatively correlated with the takt time. It can be seen as a consequence of the strong positive correlation with the number of stations, since a low number of stations implies a high takt time, also explaining that the correlation is negative. However, it is mitigated by the fact that not all balancing with a high number of stations has a low takt time. For balancing with a high takt time and a high number of stations, the value of f'_s is very variable, some can have highly productive configurations for the number of resources but others not.

The idle time being the product of the number of stations and the takt time, a similar explanation can be given for the positive correlation between the idle time and f'_s .

It is interesting to note that f'_s is not correlated with the smoothness, which means that it is as likely to have a scalable system with a well-balanced system as it is for an unbalanced one. This is a stronger statement than the one made in Son et al. (2001) saying that some unbalanced systems could be scalable.

Indeed, Table 4 shows that the classical line balancing indicators cannot be used to evaluate the scalability of a RMS, save for the number of stations. However, even if there is a strong correlation between the number of stations and f'_s , we looked more in-depth at the relationship between these two indicators and we statistically analyzed the repartition of the f'_s values for the balancing with the lowest number of stations and the neighborhood \mathcal{N}_0 (Table 5).

This table shows a significant difference between the minimum and the maximum value of f'_s , since the maximum is 2 to 4 times more than the minimum value. The first quartile represents on average an increase, regarding the minimum, of 1 percentage point, 2 percentage points for the median, 4 percentage points for the third quartile

Instance	min	1st quartile	median	3rd quartile	max	average
Mertens	9.49	11.65	12.95	15.54	28.57	14.36
Bowman8	6.57	8.01	9.69	12.00	17.16	10.26
Jaeschke	7.73	8.80	9.32	10.64	19.05	9.99
Jackson	5.63	6.13	6.88	7.73	14.09	7.30
Mansoor	4.93	5.48	6.37	8.07	12.85	7.00
Mitchell	4.32	5.26	5.85	7.48	11.31	6.42
Roszieg	4.29	4.98	5.76	6.85	12.23	6.31
Heskia	5.33	6.39	7.25	10.97	17.30	8.61
Buxey	4.32	5.17	5.85	7.83	14.36	6.61
Sawyer30	3.93	5.89	6.98	9.38	17.30	7.86
Otto-50	5.51	6.23	6.87	8.10	15.01	7.55
Otto-100	6.38	7.81	8.36	9.33	13.80	8.94
Otto-150	6.91	8.12	8.88	10.29	15.64	9.42
Otto-200	5.53	6.64	7.84	10.21	15.57	8.58
Otto-250	5.68	6.45	7.47	9.12	14.43	8.05
Otto-300	6.69	7.70	8.67	10.55	17.80	9.33
Otto-350	6.70	7.92	9.19	11.80	18.59	10.11
Otto-400	7.01	8.03	8.91	9.87	16.54	9.40
Otto-450	4.52	5.32	5.75	7.01	12.22	6.37
Otto-500	3.76	4.63	4.97	6.13	10.01	5.51
Average	5.76	6.83	7.69	9.44	15.69	8.40

Table 5. Statistical repartition of the value of f'_s (in percentage) restricted to the balancing with the lowest number of stations for each instance

and 10 percentage points for the maximum. This shows that not all balancing with the lowest number of stations has a good value for f'_s . Thus minimizing the number of stations offers no guarantee regarding the scalability of the system, which means that having a low number of stations is a necessary condition to have good scalability but not a sufficient one.

All the previous experiments have focused on \mathcal{N}_0 , i.e. deriving configurations without re-assigning the operations to the stations. In the literature, most authors have often considered reasonable to allow some operations to be reassigned during a reconfiguration of RMS. In Table 6, we aim to evaluate the gain that could come from the movement of one operation. It contains the minimum and average values of f'_s of all balancing considering \mathcal{N}_0 and \mathcal{N}_1 .

First of all, it can be observed that the average minimum value for \mathcal{N}_0 is 5.76%. This value represents the gap between the scalability level achieved with an optimal balancing when future reconfigurations will not allow to change the balancing and the ideal scalability level obtained if any balancing could be considered for the reconfigurations (i.e., reconfigurations can fully redesign the system without considering its initial or current configuration). It corresponds to the cost of imposing that all the configurations of a reachable set must share the same balancing. This cost seems really low by comparison with the additional costs of reconfiguration times associated with a change of tools. By comparing the values with the minimum values of f'_s when focusing only on the balancing with the lowest number of stations (Table 6), it can be noted that the minimum always corresponds to a balancing with the fewest stations. This emphasizes that to obtain scalable systems, the number of stations must be minimal.

When allowing one movement of operations (neighborhood \mathcal{N}_1), this average drops

Instance	min f'_s on \mathcal{N}_0	avg f_s on \mathcal{N}_0	min f'_s on \mathcal{N}_1	avg f'_s on \mathcal{N}_1
Mertens	9.49	34.98	1.73	27.62
Bowman8	6.57	34.87	1.45	30.36
Jaeschke	7.73	36.55	2.71	32.04
Jackson	5.63	36.89	0.96	33.45
Mansoor	4.93	38.60	1.63	35.81
Mitchell	4.32	39.84	0.74	37.39
Roszieg	4.29	39.74	1.22	37.00
Heskia	5.33	32.21	0.67	28.37
Buxey	4.32	37.81	0.69	35.24
Sawyer30	3.93	37.83	0.41	35.11
Otto-50	5.51	37.30	0.51	33.80
Otto-100	6.38	37.29	0.41	33.74
Otto-150	6.91	34.65	0.70	30.58
Otto-200	5.53	37.78	0.52	34.15
Otto-250	5.68	37.59	0.81	34.16
Otto-300	6.69	32.62	0.67	28.54
Otto-350	6.70	34.32	0.41	29.90
Otto-400	7.01	34.98	0.45	30.83
Otto-450	4.52	40.72	0.69	38.28
Otto-500	3.76	40.06	1.14	38.15
Average	5.76	36.83	0.93	33.23

Table 6. Minimum and average value of f'_s (in percentage) among all balancings, for \mathcal{N}_0 and \mathcal{N}_1

to 0.93%, meaning that reachable configurations are an approximation of good quality of the ideal set. The cost of reconfiguration compared to the ideal set is expected to be largely lower, since the ideal set could imply moving a large number of operations.

The average value of f'_s over all balancing is around 37 % for all instances when considering \mathcal{N}_0 and is only reduced to 33 % when considering \mathcal{N}_1 . It highlights that the potential gain of optimizing this indicator during the design phase w.r.t. picking a random solution for scalability (i.e. optimizing other criteria only), remains very close, above 30 percentage points, whatever the neighborhood considered. Moreover, this potential gain is much higher than the one obtained by switching from neighborhood \mathcal{N}_0 to \mathcal{N}_1 (around 4 percentage points).

Figure 7 shows the evolution of the minimum value of f'_s with respect to the different \mathcal{N}_i , for the 5 smallest instances of our benchmark. We can notice that the minimum f'_s quickly decreases when i increases and stabilises after $i = 2$ or 3. This shows that allowing only of small number of movements is already enough to achieve the best configurations.

Finally, we can note that for Jackson and Mansoor instances, the curve stabilises to a value close but not equal to 0 (respectively 0.07% and 0.01%). This reveals that the ideal set of configurations for those instances cannot be obtained entirely with configurations based on a balancing with the same number of stations.

6. Conclusion

Despite scalability being considered to be one of the main characteristics of RMS, its evaluation at the design phase has not yet received a lot of attention. The existing

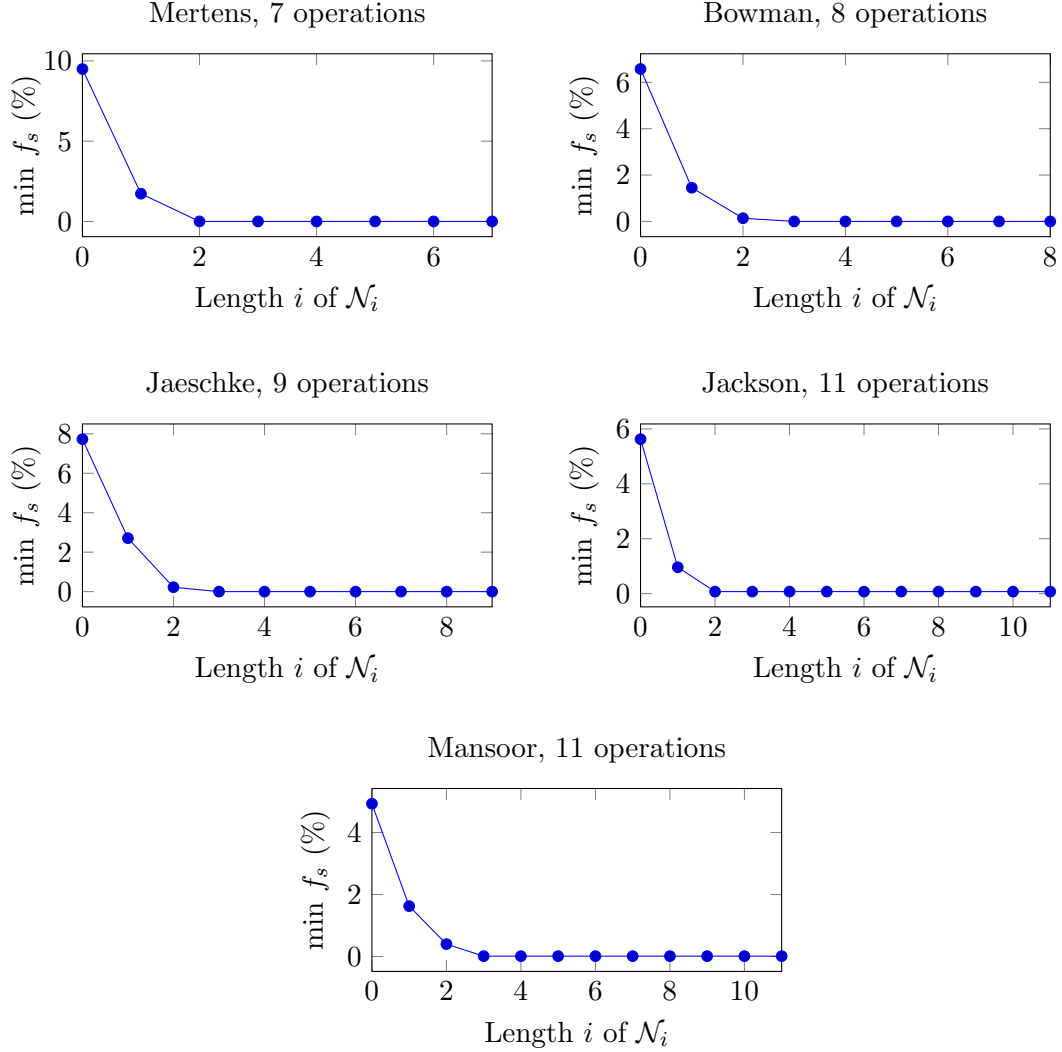


Figure 7. Evaluation of the minimum value of f'_s regarding the length i of the neighborhood \mathcal{N}_i

measures in the literature are actually difficult to use, due to some imprecisions in the definitions, to the multitude of parameters, or to the dependence to the initial state of the system.

In this paper, we propose a new measure based on a classical multi-objective metric to assess the scalability level of a balancing by taking into account all the configurations which can be achieved. An important theoretical result is that, due to the nature of the problem defined, we proposed a method to enumerate the whole set of balancing in polynomial time.

The experimental results show that this hypervolume based indicator allows a different evaluation than the smallest increment measure proposed by Wang and Koren (2012). They also reveal that the usual line balancing criteria are mainly unrelated with the scalability, enhancing the need for dedicated metrics which could be used to optimize the scalability at the design step. From a managerial point of view, to obtain a scalable system, the decision maker does not need a balanced system, as it is as likely to have a scalable system when implementing a balanced system as with an unbalanced one. This offers a stronger statement than in Son et al. (2001). It is also

interesting to note that, similarly to the conclusions reported in Koren et al. (2017), in our setting, the designer should prefer a lower number of stages, as it is strongly and positively correlated with the scalability, even without taking the reliability into account.

Moreover, the experimental results demonstrate that the cost of keeping the same balancing for the reachable configurations of a design is low if the scalability measure is taken into account during the design. In addition, this cost rapidly decreases when allowing few operations to be moved to a different machine when reconfiguring. Indeed, allowing the move of a single operation is enough to achieve a scalability level with a gap lower than 1% from the optimum. From a managerial point of view, we can conclude that, contrary to a reconfiguration for convertibility, when implementing a reconfiguration for scalability, decision-makers should focus on the first reconfiguration lever consisting in the addition of resources rather than on rebalancing the operations. Even if they want to use the second lever and change some of the tools used, they could restrict their search at one operation moved (i.e. one or two tools impacted) which will make the reconfiguration problem easier to solve and the reconfiguration itself easier to process.

Thus further works should seek to develop methods to optimize the scalability indicator when designing RMS, along with some classic economic performance measures. The proposed measure can also be used to determine further the characteristics of scalable systems and identify the parameters affecting scalability. As the scalability characteristic may not be totally independent from the five other key features of RMS, studying these interactions would be another research axis.

Data availability statement

The data that support the findings of this study are available from the corresponding author, [AC], upon reasonable request.

References

- Accorsi, R., Bortolini, M., Galizia, F.G., Gualano, F., Oliani, M., 2021. Scalability analysis in industry 4.0 manufacturing, in: Scholz, S., Howlett, R., Setchi, R. (Eds.), *Sustainable Design and Manufacturing 2020*, Springer, Singapore. pp. 161–171.
- Aguilar, H., García-Villoria, A., Pastor, R., 2020. A survey of the parallel assembly lines balancing problem. *Computers and Operations Research* 124. In press.
- Aheleroff, S., Mostashiri, N., Xu, X., Zhong, R.Y., 2021. Mass personalisation as a service in industry 4.0: A resilient response case study. *Advanced Engineering Informatics* 50, 101438.
- Andersen, A.L., Brunoe, T.D., Nielsen, K., Rösiö, C., 2017. Towards a generic design method for reconfigurable manufacturing systems: Analysis and synthesis of current design methods and evaluation of supportive tools. *Journal of Manufacturing Systems* 42, 179–195.
- Beauville Dit Eynaud, A., Klement, N., Roucoules, L., Gibaru, O., Durville, L., 2022. Framework for the design and evaluation of a reconfigurable production system based on movable robot integration. *International Journal of Advanced Manufacturing Technology* 118, 2373–2389. doi:.
- Borba, L.M., Ritt, M., Miralles, C., 2018. Exact and heuristic methods for solving the

- robotic assembly line balancing problem. *European Journal of Operational Research* 270, 146–156.
- Borisovsky, P.A., Delorme, X., Dolgui, A., 2014. Balancing reconfigurable machining lines via a set partitioning model. *International Journal of Production Research* 52, 4026–4036.
- Cerqueus, A., Gianessi, P., Lamy, D., Delorme, X., 2020. Balancing and configuration planning of RMS to minimize energy cost, in: Lalic, B., Majstorovic, V., Marjanovic, U., von Cieminski, G., Romero, D. (Eds.), *Advances in Production Management Systems. Towards Smart and Digital Manufacturing. APMS 2020*, Springer, Cham. pp. 518–526.
- Deif, A.M., ElMaraghy, W.H., 2006. A control approach to explore the dynamics of capacity scalability in reconfigurable manufacturing systems. *Journal of Manufacturing System* 25, 12 – 24.
- Deif, A.M., ElMaraghy, H.A., 2007. Assessing capacity scalability policies in RMS using system dynamics. *International Journal of Flexible Manufacturing Systems* 19, 128 – 150.
- Delorme, X., Malyutin, S., Dolgui, A., 2016. A multi-objective approach for design of reconfigurable transfer lines. *IFAC-PapersOnLine* 49, 509–514.
- Dolgui, A., Ivanov, D., Peron, M., Sgarbossa, F., 2022. Expected trends in production networks for mass personalization in the cloud technology era, in: Mourtzis, D. (Ed.), *Design and Operation of Production Networks for Mass Personalization in the Era of Cloud Technology*. Elsevier, pp. 13–37.
- Dou, J., Dai, X., Meng, Z., 2011. A GA-based approach for optimizing single-part flow-line configurations of RMS. *Journal of Intelligent Manufacturing* 22, 301–317.
- El Alaoui El Abdellaoui, M., Bricard, E., Grimaud, F., Gianessi, P., Delorme, X., 2020. Scalable, reconfigurable simulation models in industry 4.0-oriented enterprise modeling, in: Bae, K.H., Feng, B., Kim, S., Lazarova-Molnar, S., Zheng, Z., Roeder, T., Thiesing, R. (Eds.), *Proceedings of the 2020 Winter Simulation Conference*, IEEE. pp. 2755–2766.
- Essafi, M., Delorme, X., Dolgui, A., 2012. A reactive GRASP and path relinking for balancing reconfigurable transfer lines. *International Journal of Production Research* 50, 5213–5238.
- Essafi, M., Delorme, X., Dolgui, A., Guschinskaya, O., 2010. A mip approach for balancing transfer line with complex industrial constraints. *Computers and Industrial Engineering* 58, 393–400.
- Freiheit, T., Shpitalni, M., Hu, S.J., 2004. Productivity of paced parallel-serial manufacturing lines with and without crossover. *Journal of Manufacturing Science and Engineering* 126, 361–367.
- Garey, M.R., Johnson, D.S., 1979. *Computers and intractability: a guide to the theory of NP-completeness*. Freeman, New York.
- Gola, A., Pastuszak, Z., Relich, M., Sobaszek, L., Szwarc, E., 2021. Scalability analysis of selected structures of a reconfigurable manufacturing system taking into account a reduction in machine tools reliability. *Eksploatacja i Niezawodność – Maintenance and Reliability* 23, 242–252.
- Hees, A., Schutte, C.S., Reinhart, G., 2017. A production planning system to continuously integrate the characteristics of reconfigurable manufacturing systems. *Production Engineering* 11, 511–521.
- Hsieh, F.S., 2018. Design of scalable agent-based reconfigurable manufacturing systems with petri nets. *International Journal of Computer Integrated Manufacturing* 31, 748 – 759.

- Hu, Y., Guan, Y., Han, J., Wen, J., 2017. Joint optimization of production planning and capacity adjustment for assembly system. *Procedia CIRP* 62, 193 – 198.
- Ivanov, D., Dolgui, A., Sokolov, B., 2019. The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research* 57, 829–846.
- Koren, Y., Gu, X., Guo, W., 2018. Reconfigurable manufacturing systems: Principles, design, and future trends. *Frontiers of Mechanical Engineering* 13, 121 – 136.
- Koren, Y., Hu, S.J., Weber, T.W., 1998. Impact of manufacturing system configuration on performance. *Annals of the CIRP* 47, 369 – 372.
- Koren, Y., Shpitalni, M., 2010. Design of reconfigurable manufacturing systems. *Journal of Manufacturing Systems* 29, 130–141.
- Koren, Y., Wang, W., Gu, X., 2017. Value creation through design for scalability of reconfigurable manufacturing systems. *International Journal of Production Research* 55, 1227 – 1242.
- Lahrichi, Y., Grangeon, N., Deroussi, L., Norre, S., 2021. A new split-based hybrid metaheuristic for the reconfigurable transfer line balancing problem. *International Journal of Production Research* 59, 1127–1144.
- Maganha, I., Silva, C., Ferreira, M.D., 2018. Understanding reconfigurability of manufacturing systems: An empirical analysis. *Journal of Manufacturing Systems* 48, 120–130.
- Moghaddam, S.K., Houshmand, M., Saitou, K., Fatahi Valilai, O., 2020. Configuration design of scalable reconfigurable manufacturing systems for part family. *International Journal of Production Research* 58, 2974–2996.
- Napoleone, A., Andersen, A., Pozzetti, A., Macchi, M., 2019. Reconfigurable manufacturing: A classification of elements enabling convertibility and scalability, in: Ameri, F., Stecke, K., von Cieminski, G., Kiritsis, D. (Eds.), *Advances in Production Management Systems. Production Management for the Factory of the Future. APMS 2019*, Springer, Cham. pp. 349–356.
- Napoleone, A., Negri, E., Macchi, M., Pozzetti, A., 2022. How the technologies underlying cyber-physical systems support the reconfigurability capability in manufacturing: a literature review. *International Journal of Production Research* DOI: 10.1080/00207543.2022.2074323.
- Nishith, M., Rishi, G., Sharma, S., 2013. Flexibility measurement criteria with respect to reconfigurable system properties. *International Journal of Engineering Research and Applications* 3, 1711–1716.
- Otto, A., Otto, C., Scholl, A., 2013. Systematic data generation and test design for solution algorithms on the example of salbp-gen for assembly line balancing. *European Journal of Operational Research* 228, 33–45.
- Pareto, V., 1896. *Manuel d'économie politique*. F. Rouge, Lausanne.
- Putnik, G., Sluga, A., ElMaraghy, H., Teti, R., Koren, Y., Tolio, T., Hon, B., 2013. Scalability in manufacturing systems design and operation: State-of-the-art and future developments roadmap. *CIRP Annals - Manufacturing Technology* 62, 751 – 774.
- Rösiö, C., Aslam, T., Srikanth, K.B., Shetty, S., 2019. Towards an assessment criterion of reconfigurable manufacturing systems within the automotive industry. *Procedia Manufacturing* 28, 76–82.
- Scholl, A., 1999. *Balancing and Sequencing of Assembly Lines*. Contributions to Management Science, Physica-Verlag HD.
- Son, S.Y., Olsen, T.L., Yip-Hoi, D., 2001. An approach to scalability and line balancing for reconfigurable manufacturing systems. *Integrated Manufacturing Systems* 12,

500 – 511.

- Spicer, P., Koren, Y., Shpitalni, M., Yip-Hoi, D., 2002. Design principles for machining system configurations. *CIRP Annals* 51, 275 – 280.
- Wang, G.X., Huang, S.H., Yan, Y., Du, J.J., 2017. Reconfiguration schemes evaluation based on preference ranking of key characteristics of reconfigurable manufacturing systems. *International Journal of Advanced Manufacturing Technology* 89, 2231–2249.
- Wang, W., Koren, Y., 2012. Scalability planning for reconfigurable manufacturing systems. *Journal of Manufacturing Systems* 31, 83 – 91.
- Yelles-Chaouche, A.R., Gurevsky, E., Brahimi, N., Dolgui, A., 2021. Reconfigurable manufacturing systems from an optimisation perspective: a focused review of literature. *International Journal of Production Research* 59, 6400–6418.
- Youssef, A.M.A., ElMaraghy, H.A., 2006. Modelling and optimization of multiple-aspect RMS configurations. *International Journal of Production Research* 44, 4929–4958.
- Zheng, P., Sang, Z., Zhong, R.Y., Liu, Y., Liu, C., Mubarak, K., Yu, S., Xu, X., et al., 2018. Smart manufacturing systems for industry 4.0: Conceptual framework, scenarios, and future perspectives. *Frontiers of Mechanical Engineering* 13, 137–150.
- Zidi, S., Hamani, N., Kermad, L., 2021. New metrics for measuring supply chain reconfigurability. *Journal of Intelligent Manufacturing* , 22p.
- Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., da Fonseca, V.G., 2003. Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Transactions on Evolutionary Computation* 7, 117–132.