Assembly system design and operations for product variety

S.J. Hu (2)a,*, J. Ko b, L. Weyand c, H.A. ElMaraghy (1)d, T.K. Lien (1)e, Y. Koren (1)f, H. Bley (1)g, G. Chryssolouris (1)h, N. Nasr g, M. Shpitalni (1)i

a Department of Mechanical Engineering, The University of Michigan, Ann Arbor, MI, USA
b Department of Industrial & Management Systems Engineering, University of Nebraska-Lincoln, NE, USA
c Intelligent Manufacturing Systems Centre, University of Windsor, Windsor, Canada
d Department of Production and Quality Engineering, Norwegian University of Science and Technology, Norway
e Laboratory for Manufacturing Systems and Automation, University of Patras, Greece
f Center for Integrated Manufacturing Studies, Rochester Institute of Technology, Rochester, NY, USA
h Laboratory for CAD and Lifecycle Engineering, Technion – Israel Institute of Technology, Israel

ABSTRACT

Assembly is the capstone process for product realization where component parts and subassemblies are integrated together to form the final products. As product variety increases due to the shift from mass production to mass customization, assembly systems must be designed and operated to handle such high variety. In this paper we first review the state of the art research in the areas of assembly system design, planning and operations in the presence of product variety. Methods for assembly representation, sequence generation and assembly line balancing are reviewed and summarized. Operational complexity and the role of human operators in assembly systems are then discussed in the context of product variety. Challenges in disassembly and remanufacturing in the presence of high variety are presented. We then conjecture a future manufacturing paradigm of personalized products and production and discuss the assembly challenge for such a paradigm. Opportunities for assembly system research are summarized at the end of the paper.

© 2011 CIRP.

1. Introduction

Mass customization has become a prevalent paradigm of manufacturing since the late 1980s as it seeks to provide customized products at near mass production cost [153]. As a result of the paradigm shift from mass production to mass customization, the number of varieties offered by consumer product manufacturers has increased significantly over the past several decades. For example, the number of distinct vehicle models in the U.S. increased from 44 in 1969 to 165 in 2006 [185,186]. Within each model, there can be many choices on the powertrain and interior combinations. Another example is the number of styles of running shoes, which increased from 5 in the early 1970s to 285 in the late 1990s [39]. Such increases were motivated by the desire to provide high variety and highly customized products in response to the diversification of consumer needs and preference, and the fierce competition in the global market. As manufacturers try to adapt their product offering to satisfy segmented markets, more varieties were created based on certain base designs.

Variety can be achieved at different stages of product realization, during design, fabrication, assembly, at the stage of sales, or through adjustment during the usage phase (see Fig. 1). Designed-in variety incorporates customer design inputs and such products tend to be personalized, one-of-a-kind products. Variety can also be added during the fabrication process, for example, through machining, or rapid prototyping. Many biomedical products are fabricated with high variety to respond to the high human variability.

Assembly is one of the most cost effective approaches to high product variety. With proper design of a Product Family Architecture (PFA) [179], each functional module of the product is provided with several variants so that the assembly combination will provide high variety in the final products (see Fig. 2, where the total number of variety is $3 \times 2 \times \cdots \times 3$). Such an approach enabled the production of customized products at near mass production cost, which was cost-effectively accomplished by designing the basic product options and allowing the customers to select the assembly combination that they most prefer. The economy of scale is achieved at the component level, while economy of scope of high variety is achieved in the final assembly by using flexible/reconfigurable manufacturing systems.

Variety can also be created during the time of sales or use. For example, golf clubs can be cut to length at the time of purchase in order to fit an individual’s height and swing pattern. Seat heights on bicycles can be adjusted at the time of use. These adjustments are made based on mass produced products.

Since assembly is a cost effective approach to variety, this paper reviews the state of the art research in the design and operations of assembly systems in support of product variety and identifies opportunities for future research. Specifically, we first review the

Please cite this article in press as: Hu SJ, et al. Assembly system design and operations for product variety. CIRP Annals - Manufacturing Technology (2011), doi:10.1016/j.cirp.2011.05.004

* Corresponding author.

0007-8506/$ – see front matter © 2011 CIRP.
doi:10.1016/j.cirp.2011.05.004
literature in assembly representation and sequence generation for a family of products, and methods and algorithms for designing and balancing assembly systems in the presence of variety. We also review algorithms for planning, scheduling and operating such assembly systems. Since variety causes complexity in manufacturing and assembly systems, we also discuss models of product variety induced manufacturing complexity and discuss their applications. Finally, we discuss a future paradigm in personalized products and the associated assembly challenges. Opportunities for research in assembly systems in support of product variety are summarized together with the conclusions.

2. Assembly representation and sequence

The design of an assembly system requires methods to represent the assembly components and hierarchy, and to generate the sequences of assembly. Methods and algorithms for assembly representation and sequence generation are reviewed in this section.

2.1. Representation of product assembly

Here we review the most commonly used assembly representation methods, including liaison and precedence graphs, and discuss how these methods are adopted for representation of products with variety.

2.1.1. Common representation methods for product assembly

Several methods are available to represent the relationship among component parts in an assembly, and such representation can be quite useful during system conceptual design and assembly sequence planning. One of the commonly used assembly representation methods is the Bill-of-Material (BOM). A BOM generally lists all parts, subassemblies and materials, and also includes other information such as quantities, costs and manufacturing methods. A BOM usually has a tree-graph or tabular structure with hierarchical level codes [81]. A variety of BOM graphs, such as Network BOM [140], have also been used to represent the functional relations of parts and subassemblies. The BOM has been a standard communication tool in industry for design, manufacturing and purchasing, and has been integrated to Computer-Aided Design (CAD) and Enterprise Resource Planning (ERP) systems.

Another commonly used assembly representation is the graph-theoretic description of components and their physical connections, such as the liaison graph and adjacency matrix. A liaison graph is a graphical network wherein nodes represent parts and lines between nodes represent certain user-defined relations between parts. These relations, represented using edges in a graph, are called “liaisons”, which represent the physical contact or joining between components [195]. Any assembly step is characterized by the establishment of one or more of the liaisons of the assembly. Fig. 3 shows the components of a laptop computer and the corresponding liaison diagram. The assembly process is complete once all liaisons are established.

The liaison graph has also been used for generating assembly sequences. For instance, liaison graphs were used to deduce the assembly task precedence in generating all feasible sequences [40]. An AND/OR graph representation was used to develop a correct and complete algorithm to generate all feasible assembly sequences [78–80]. A cut-set method was also used to generate all feasible assembly sequences for the concurrent design of products and assembly lines [11].

In addition, other diverse aspects of the assembly have been represented by a variety of means. The precedence graph has been used extensively to represent the constraints on processing orders among assembly tasks, e.g. what tasks must be completed before other tasks. Such precedence relations are particularly useful in assembly line balancing problems. Most precedence graphs use assembly tasks (realization of liaisons) rather than components.

Please cite this article in press as: Hu SJ, et al. Assembly system design and operations for product variety. CIRP Annals - Manufacturing Technology (2011), doi: 10.1016/j.cirp.2011.05.004
but if a base component is first established, then the task of adding each subsequent component can simply be represented by the component on the precedence graph (see Fig. 4 for example). Mechanical assembly features such as tolerance and kinematics were also included in assembly representation [154]. A general constraint model was proposed to express process constraints more systematically and efficiently [180]. An ontology-based representation was proposed to identify differences in joints by using a region-based theory and semantic web rule language [97]. Disassembly was also used widely either to analyze assembly sequence [79] or to concurrently plan assembly and disassembly using directed graphs [200].

2.1.2. Assembly representation methods for product variety

The increasing product variety has led to new approaches in assembly representation, as summarized in several review papers. A comprehensive review was conducted on graph-based assembly representations such as graphs of location, virtual link, constraint, operation and functions for integrated product and process design [203]. Product family and platform designs were also reviewed in several survey papers [88,94].

The assembly representations for product variety appeared in the literature in diverse forms. Among these, the PFA has been one of the extensively studied topics. A PFA was used to measure market position, commonality and manufacturing economy [89,91,179]. The BOM has also evolved further to represent a variety of products, in particular, product families in a more convenient way. Common approaches are to introduce the concept of Generic Bill of Material (GBOM) [77,147]. These GBOMs use functional and structural relations among components to represent product variants. A variety of representation methods were used including tabular forms and programming language based notations. Other hierarchical representations were also used to represent product families. For instance, generic subassemblies for a product family was used for integrated product family and assembly system design [41].

Liaison graphs have also been adapted to represent product variety. One such development is the product family liaison graph that combines the liaison graphs of product variants by representing common components over different variants as a single node. Thus, for a family of products, the liaison graph can be modified to include both common and variant parts in the assembly. A product family liaison graph was used to identify maximal common subassemblies and a product-family assembly sequence [73].

The assembly representations have developed to incorporate more diverse aspects of the product variety. For example, the similarity and dependency in assembly modularity were expressed using cost criteria in terms of tool or fixture change [110]. The relations of the cost with product and process variety were also investigated in the product family design [204].

2.1.3. Current and further research directions

The current assembly representations are limited in terms of the comprehensiveness of assembly information. For example, the usual BOM cannot directly represent the complex physical assembly processes. On the other hand, the assembly representations based on the liaison graphs are not suitable in representing hierarchical functional structures. A next generation information system [135] is desirable to provide designers and manufacturing engineers with more comprehensive information with convenient data management features. A new graph-theoretic assembly representation incorporating product and process information is possible to overcome the above problem.

Another important issue is the assembly representation for collaborative development of product families. Nowadays more and more design and assembly work is conducted as collaborative designs across globally distributed design teams, companies and software modules. Therefore, an assembly representation enabling interoperability across different locations and software platforms are critical. Some proposed concepts include e-Assembly system for collaborative assembly representation [30] and web-based collaboration system [87]. An example of industrially available collaborative system is the TeamCenter offered by Siemens. Research in this area should be extended to provide more efficient assembly representation tools for product variant customization. The reason is that globalized design and manufacturing often require the variants for local markets to be generated by regional design teams that use different assembly software and supply bases.

The standardization of assembly representation is also a critical issue for interoperable and collaborative designs. Such efforts include ISO Standard 10303 for Product Data Representation and Exchange and other related standards by ISO working group TC 184/SC 4 [170], and the US National Institute of Standards and Technology (NIST) Core Product Model (CPM) and Open Assembly Model (OAM) [56].

In addition, more advanced concepts of variety need to be developed, especially for modeling the relation between customers and products for customization and personalization [197]. These new representations will provide tools for more buyer centric marketing and assembly plans and enable manufacturers to evaluate the benefits of such customization approaches. More complete integration of information on the environmental impact is also desirable. Current life cycle assessment (LCA) tools often are not conveniently integrated with assembly representation, in particular, for variant evaluation and management. Assembly representation systems inherently integrated with LCA databases will greatly speed up the environmental assessment in product variant design. However, some product characteristics of importance to the overall environmental performance of the product cannot be easily represented in such LCA databases. For example, energy use of the product during use stage, product disassemblability, etc., are not easily incorporated in the LCA database and may lie on a meta level above the information on choice of components and subassemblies. Another direction of research is to incorporate uncertainty information of product performance and reconfiguration as part of product variant modeling. Although some software tools are available for uncertainty modeling, their use has been limited.

2.2. Assembly sequence

The sequence of assembling a set of parts plays a key role in determining the quality of the assembled product, as well as assembly process design issues, such as the needs for fixturing, ability for in-process testing, and the number of assembly steps. Determination of all possible assembly sequences is an important and critical stage in the total design process of a product. One of the pioneers in assembly sequence research is Bourjault. Bourjault’s early work used rules that are determined by a series of “yes” or “no” questions, which are answered by studying the mating of components for an assembly [23]. Bourjault represented a product by using the information contained in a part list and an assembly drawing to form a liaison graph, where the components are the...
nodes and the liaisons are the determined mates. All assembly sequences are determined algorithmically using the liaison graphs.

De Fazio and Whitney [40] extended Bourjault’s work by simplifying the determination of the set of rules, or precedence constraints, by using specific questions about liaison precedence. Questions are specifically asked about “what liaisons must be done prior to doing liaison i” and “what liaisons must be left to be done after doing liaison i”. De Fazio and Whitney’s work significantly reduced the question count for determining all possible assembly sequences. Their later work with Baldwin et al. [11] took advantage of using a computer as aid for automatic assembly sequence generation. Other work that takes advantage of a computer aid in determining all assembly sequences is the work of Khosla and Mattikalli [96]. They developed a methodology that uses software to automatically determine the assembly sequence from a 3D model of the assembly [96]. Further, Kanai et al. [95] developed a computer aided Assembly Sequence Planning and Evaluation system (ASPEN) that takes all the solid-model components of a product and automatically determines all feasible sequences by decomposition and determines the optimum sequence using Methods Time Measurement (MTM) as time standards for operating time determination. Choi and Zha developed computer aided automatic assembly sequence generation with their work on automated sequence planning [32]. They use the creation of an AND/OR graph and the identification of leveled feasible sub-assemblies to determine the assembly sequence.

The works of de Mello and Sanderson built upon previous research by treating an assembly sequence generation problem as a disassembly sequence problem [79,80]. The problem is then further decomposed into sub-problems where subassemblies are joined one at a time.

Gupta and Krishnan [73] created an algorithm to determine the largest subassembly in an assembly problem for a product family where some components differ. They used De Fazio and Whitney’s algorithm for finding all assembly sequences and implemented their searching algorithm to find the maximum generic subassembly. Dini et al. [44] made use of the genetic algorithm to create and evaluate assembly sequences. They created a fitness function which takes into account geometrical constraints of the assembly and other optimization aspects and using their genetic algorithm decreased the time for computation. Marian et al. [127] also attempted to optimize the assembly sequence planning problem by using genetic algorithms. Wang and Ceglarek [182] used graph theory by developing a methodology that generates all the sequences for a k-ary assembly process. The authors used a k-piece graph to represent assemblies without precedence constraints and a k-piece mixed-graph to represent assemblies with precedence information. Using this approach, all feasible sub-assemblies can be identified, and all of the assembly sequences for a k-ary assembly process are generated iteratively.

Exploring the choices of assembly sequence is a very difficult task for two reasons. First, the number of possible sequences can be large for even a small number of parts and can increase staggeringly with increasing parts counts. Second, seemingly minor design changes can drastically modify the available choice of assembly sequence. Up to now, almost all assembly sequence generation algorithms are based on sequential tasks. Consideration of assembly hierarchy allows parallel assembly sequence and hybrid system configurations [119] and such choices can be explored to simplify assembly sequence generation and system design.

3. Assembly system design for product variety

Upon the availability of a set of feasible assembly sequences, then the design of an assembly system is accomplished with the creation of system configurations and balancing of the assembly systems by assigning tasks to the proper stations. Methods and algorithms for these key steps are reviewed in this section.

3.1. Assembly system configurations

Assembly systems can be designed using various configurations. The moving assembly line introduced by Ford [58] had a serial layout. Such systems, known as serial lines or flow lines, were used for high volume production of a single product type with dedicated machines and material handling systems. Since then, assembly systems have become much more sophisticated and complex, not only to accommodate more complex products but also to provide the flexibility needed to handle the increasing variety of products resulting from the trend toward mass customization. Different configurations are being used and they are described below.

System configurations are classified primarily into two different types: synchronous configurations, whereby each part undergoes the same sequence of operations regardless of its path through the system, and asynchronous configurations, whereby parts may undergo different operation sequences, depending on their path through the system [171]. In synchronous systems, parts move from one operation to the next at a constant pace. Therefore, synchronous systems are more appropriate for mass production. Asynchronous systems are more commonly used in assembly systems, especially whenever subassemblies are used. The main assembly line is typically serial with feeders from other subassembly serial lines, as shown in Fig. 5. Note that all the synchronous systems depicted in Fig. 5 are special cases of the flexible-general configuration. In general, for machining applications, symmetrical configurations are generally used. Yet in this era marked by mass customization, personalization and more complex products, more complex configurations are frequently used, particularly in assembly [99].

System configurations are determined by (1) the arrangement of the machines and (2) the relations (connections) among them. For example, in Fig. 6, different machine arrangements are shown in (a) and (b), so that (a) and (b) represent different configurations. Similar machine arrangements are shown in (c) and (d), but the connections among the machines are different, and therefore the
configurations are different as well. In the figure, (e) and (f) have both the same arrangements and the same connections, and thus represent the same configuration.

Koren and Shpitalni reviewed the design of reconfigurable manufacturing systems with detailed discussion of system configurations [103]. Webbink and Hu proposed an algorithm for generating assembly system configurations and matching these configurations to assembly sequences [187].

In general, the number of configurations given a certain number of machines can be quite high. For example, given six machines, there can be 170 different configurations, including serial, parallel and hybrid configurations [103]. Not all these configurations are used in manufacturing since planning and operations can be quite difficult for some of the non-symmetric configurations. An interesting theoretical question is: how many configurations exist. This is a difficult and complex question that does not have a simple answer. Shpitalni and Kumaz [165] developed a counting procedure for agile systems. For a system consisting of n machines in m operations, the procedure involves generating all possible arrangements and then calculating all possible connections for each arrangement. For example, a 15-machine system arranged in approximately five operations has 1001 possible arrangements. One of these arrangements {1, 3, 5, 7, 9} (machines in the first operation, second operation, etc.) has 196 configurations. To determine the total number of possible configurations, the number of connections for each of the 1001 arrangements must be calculated separately and these must then be summed up. In addition, Shpitalni and Remennik [166] examined the number of practically used paths in reconfigurable manufacturing systems with crossover. They showed that the practical number of different paths is far lower than the maximum theoretical number and that this number decreases with shorter lines and with higher machine reliability.

Another aspect of configuration in automatic assembly is the need for part feeder configuration. For one single flow configuration there will exist several different total configurations depending on the set of feeders used for a particular product or product family. A configuration change will take place every time a different set of feeders are put into operation. This is a common way of reconfiguring systems for products with similar assembly sequence but differences in part composition. So each possible flow configuration must be multiplied by the number of viable part feeder sets to obtain the total number of configurations.

The impact of different configurations on system performance can be profound. Koren et al. [102] analyzed the performance of difference configurations in terms of quality, throughput and convertibility. Freiheit et al. [61] analyzed the throughput performance of systems configurations with and without crossover. The impact of configuration and material handling on system throughput was studied by Freiheit et al. [60].

3.2. Assembly line balancing

Assembly line balancing is to search for the optimum assignment of assembly tasks to stations given precedence constraints according to a pre-defined single or multi-objective goal. These objectives vary from a single objective of minimizing the number of stations for a given cycle time, or minimizing the cycle time for a given number of stations in a serial line, to optimizing line efficiency and imbalance simultaneously in a non-serial line [156].

Balancing for a single product type may be solved in the form of two problem types [164]: “Type I consists of assigning tasks to workstations such that the number of stations is minimized for a given production rate. Type II is to maximize the production rate, or equivalently, to minimize the sum of idle times for a given number of stations”. Balancing of assembly lines where product variety exists involves a deeper elaboration of an initial rough assembly line layout to achieve a desired cycle time.

3.2.1. Line balancing for variety

Often, two approaches were suggested for assembly lines for multiple product models: (1) a multi-model assembly line where different product models are considerably distinctive, therefore production is executed in batches of each product model, and (2) a mixed-model assembly line where the product model variants are significantly similar that they can be assembled simultaneously on the same line [24]. The applications of these lines are: automotive, furniture, electronics industries, etc.

The line balancing for multi-model production poses new challenges. For example, to assemble a variety of product models without building individual lines for each product model, different product models are assembled in a single assembly line in mixed-model production. In mixed-model assembly balancing, however, the different assembly process characteristics of different models result in new problems such as drift and model sequencing that do not exist in simple single-model balancing.

Drift: The term drift represents the deviation from the optimal cycle time. One major goal of line balancing is to achieve a similar cycle time at each station, but this is nearly impossible in practice if lots of product variants which need different assembly operations have to be produced. The deviations can be negative or positive as shown in Fig. 7.

The negative drift represents the time span during which the worker of a station does not perform any assembly activity regarding one special product variant. A negative drift that is extremely high is often caused by a high variance of assembly processes. A negative drift does not comply with the requirement of a production which is close to the maximum capacity of a station and it reduces the total efficiency of the line [53,157,189].

The positive drift represents the time span during which the worker of a station exceeds the predefined cycle time regarding one special product variant. A positive drift that is extremely high is often caused by a high variance of assembly processes. It not only puts pressure on the worker at the station concerned, but also has a negative impact on the total production. Stations including a large positive drift can easily become “bottlenecks” within assembly lines [71,189].

The majority of the research addressing the drift focused more on the positive drift (also called work overload) than the negative one, because the former is considered more costly in terms of lost productivity and impact on other stations such as line stoppage. There exists a great deal of studies proposing methods to reduce the positive drift. A workload stability problem was solved using heuristics [14,128,176,199,206]. A case study examined how the overload was handled in a US industry [37]. Some research also considered the stochastic assembly times and their effect on work overload [205]. A bypass sub-line was modelled to minimize the line stoppage due to work overload in some product models [134]. In fact, a great deal of research on overload minimization is related to mixed-model sequencing.

3.2.2. Integrated approach to mixed-model line balancing and sequencing

The use of mixed-model assembly lines is widespread due to their advantages in reducing inventories, eliminating transfer costs

---

1 This section is based in part on results of the research project “Flexible Assembly Processes for the Car of the Third Millennium (MyCar)” funded by the Commission of the European Union.

![Fig. 7. Positive and negative drift in accordance with [53,189].](image-url)
between models and meeting ever-changing customer demands more efficiently. Nevertheless, the use of mixed-model lines makes it almost impossible to balance the line properly, due to differences between models characteristics. Gilad et al. showed that operating a number of parallel assembly lines, each assembles different models, the efficiency of the assembly system can be improved and its operating costs reduced [75].

The operation of the mixed-model assembly line usually involves model (build) sequencing problem that determines the order by which models are released to the line. The reason for the need of such model sequencing is that the multi-model line balancing solutions can minimize only the potential long-term workload fluctuations (drifts) due to different models. Therefore, the mixed-model balancing and sequencing problems are solved together.

3.2.3. Input needed for line balancing

In order to be able to perform line balancing, process information and precedence constraints regarding possible assembly sequences of components have to be available. Line balancing can be performed only after the process planning has been completed, whereas there are different approaches to the identification of precedence constraints. According to [15], these approaches can be divided into at least three categories:

- Approaches dealing with assembly topology (liaison graph, joining area graph, AND/OR graph, etc.)
- Approaches dealing with geometry (e.g. analyses based on (dis)assembly simulations considering parts as solid objects)
- Approaches dealing with the analysis of incorporated additional assembly related information (referred to as assembly features) which can be functional, semantic, relational, technological, etc.

All in all, assembly features represent a promising approach that can help identifying precedence constraints and achieving an automated assembly planning [15–17,19]. However, lots of old and new publications – for instance [44,49,51,76,85,209] – are dealing with precedence constraints and highlighting the importance of identifying and documenting those constraints within assembly planning, whereas the general acceptance of the existing approaches is partly low in the industrial practice. Line balancing experts often trust their experience only.

3.2.4. Line balancing modeling and solution approaches

Assembly lines are divided into stations that have to be clocked as equally as possible so that the predicted volume of products can eventually be realised. The expected demand determines the cycle time of the line and of the stations [157]. As to the personnel planning in the context of line balancing, equal workloads with equal execution times should be assigned to the workers [116].

A. General mathematical formulations and solution methods for mixed-model assembly lines

A variety of mathematical formulations have been used to model the mixed-model assembly line balancing problem. The list includes binary integer programming [68], mixed integer programming [162], goal programming [67] and non-linear integer programming [8]. Diverse solution methods have also been used. They include branch-and-bound methods [27], genetic algorithms [145,167], simulated annealing methods [129], greedy algorithms [92], and ant colony algorithms [181]. Due to the computational complexity, more and more heuristics algorithms have been proposed.

B. Industrial approach dealing with the average assembly time and drift

A line balancing approach dealing with average calculations implies that processes, which were generated within the process planning, are assigned to stations so that an average cycle time that is acceptable can be achieved. For alternative parts, the average assembly time has to be acceptable. This approach is often favoured in the industrial practice. The consideration of an average assembly time provides first hints regarding the workload and the required number of workers, but it does not give any variant-specific information. Therefore, subsequent variant-specific drift analyses are necessary. Fig. 8 shows the average assembly time of alternative steering wheels that are different as far as their assembly processes are concerned. The steering wheels have to be assembled in one station where further parts, which are not variant-specific, have to be assembled as well. As shown in the example, the predefined cycle time \( t_{\text{acceptable}} \) is observed in case of the average assembly time of steering wheels. This result would still be valid if the product variants consisting of a steering wheel \( Y \) were considered individually. For product variants consisting of a steering wheel \( Y \), a negative drift would occur, whereas this drift would still be acceptable. The assembly of a steering wheel \( X \) would lead to a positive drift instead which exceeds the predefined cycle time \( t_{\text{acceptable}} \). This drift would not be acceptable. The information regarding the positive drift is not directly available in connection with the average calculation. Consequently, this example shows that average calculations can be only the first step of the procedure of finding an optimal balancing. Variant-specific aspects play an important role in today’s assembly planning and there is no doubt that further effort has to be spent into the topic of an adequate variant management which is able to deal with product, process and resource variants in a transparent, non-ambiguous way [18,21,22,202].

3.2.5. Results of line balancing and software

Today’s line balancing deals with lots of different aspects. Traditionally, the main goal is an optimal allocation of processes to stations so that no or at least acceptable drifts occur. Further goals include the optimization of logistic aspects and walking distances within the line. Modern digital line balancing tools offer possibilities not only for allocating processes to stations, but also allow a simultaneous visualisation of the utilisation of the material.

**Fig. 8.** Line balancing approach dealing with average calculations according to [185].

Please cite this article in press as: Hu SJ, et al. Assembly system design and operations for product variety. CIRP Annals - Manufacturing Technology (2011), doi:10.1016/j.cirp.2011.05.004
zones (see Fig. 9) as well as the calculation of walking distances of employees which are linked to the location of the part boxes within the material zones. For instance the DELMIA tool named “Automatic Line Balancing” offers the above-mentioned possibilities. Moreover, it has to be considered that the line balancing also affects the final need of resources [190,193]. Consequently, minimizing the final need of resources or the environmental impact can be another goal, whereas this approach only makes sense in cases where expensive resources are used.

Several other software implementations are available, commercial or non-commercial. A commercial software package OptLine uses a genetic algorithm [54,148]. Many non-commercial software implementations are available, although the majority of them are for problem specific or for single model problems [7,93]. A heuristic algorithm was also implemented as an MSAccess based software for an industrial case study [113].

3.2.6. Current shortcomings and future research topics
Despite extensive research on assembly line balancing, still further research is desirable. Such research needs were summarized in several review papers [24], including rebalancing, feeder line, disassembly and material supply. In addition to these general research needs, several other important aspects need to be studied further especially for multi-product assembly lines. One of the topics is the consideration of task-sequence dependent inter-task times between product models [143], since the high variety may involve more sequence dependent setup times that cannot be ignored anymore. Other topics include the balancing in complex configurations other than serial lines [99] and the effect of stochastic product changes on line balancing [100]. The integration of balancing and sequencing is still a topic worth more attention. Robust line balancing is also an important topic, since product mix and volume change frequently. For the more extensive application of academic research in industry, practical algorithms need to be developed and implemented as software.

A serial line might also possess several parallel identical stations in a single assembly stage to satisfy a desired cycle time. It is indicated that for such an arrangement, the workload differences increase drastically among those identical stations when product variety increases in a mixed products assembly. Consequently, frequent product rescheduling and process re-sequencing are often needed [24].

One specific industrial problem is that there is often a gap between process planning and line balancing. Process planners partly identify precedence constraints, but they mostly do not document the obtained information in the digital planning environment. Consequently, the persons who are dealing with line balancing have to start from scratch regarding precedence constraints due to the existing information gap. Moreover, it has to be considered that the creation of precedence graphs is not trivial – at least if complex products with a high product variety have to be considered. Different approaches to creating precedence graphs exist, but their suitability is partly case-specific and they are often not used in the industrial practice [15].

One further challenge is the volume flexibility. Manufacturers are nowadays confronted with a turbulent global market, and therefore it is very difficult to predict the needed production volume of each product variant. The goal of volume flexibility is to create production lines which are prepared for the predicted demand, but which can also be easily and profitably adapted to other demand scenarios [157,188,191]. Line balancing research has to overcome this challenge by starting volume flexibility investigations [192]. The topic of volume flexibility is very complex due to the fact that not only OEMs have to deal with this challenge – suppliers are affected as well [163]. In addition, optimized assembly stations layout is needed to eliminate non-value added assembly process elements such as walking and waiting [98]. The serial line can be bent in a U-shaped line for more homogenous workloads of assembly stations, however balancing these workloads requires production scheduling integration with tasks assignment to stations [133]. Few researchers focused on balancing more complex assembly line shapes such as asymmetric lines [99,187] and cellular layouts [115], however, the lack of integration of the layout design with assembly line balancing is still a problem. All these are compounded by the increasing number of product variants. Integration of line balancing with equipment selection is another area of research for assembly system design [177].

All in all, much research is still necessary for highly flexible production lines which are effective and economic. For the development of highly flexible production lines, cost aspects are very important too [136]. Cost aspects have to be considered as early as possible in the planning phase [194].

3.3. Assembly system design for delayed product differentiation
Adopting postponement strategies is one of the solutions for managing variety in assembly systems [114,174,175]. In particular, Delayed Product Differentiation (DPD), a form postponement, is implemented in assembly systems to defer the differentiation of final product configurations. In a mixed-product assembly, the points of product differentiation represent specialization stages in the assembly system where each product starts to develop its own unique identity, thus, becoming differentiated from other variants in the family.

Delayed differentiation plays an important role in obtaining a cost effective manufacturing method in the final stages of production. Zinn [208] has identified four different reasons for
delayed differentiation: labelling, packaging, assembling and manufacturing. In Zinn’s terminology, manufacturing means that raw materials are transported to the warehouse where both parts are manufactured and the final assembly is performed to customer orders. This variant really is nothing but a distributed manufacturing system. The notion of delayed assembly is more interesting. Here the product is assembled to customer order and shipped immediately thereafter. The manufacturer then keeps stock of manufactured standard parts and sub-assemblies that are assembled according to an individual customer’s order. In this way a large number of customized products may be assembled from a small set of basic components.

The idea of delayed differentiation is not totally new though. Henry Ford introduced the idea of knock-down bodies in the early 1920s. The main idea was to send the components of the cars tightly packaged to assembly plants in different parts of the world. This saved freight cost and time [57]. It also gave as side benefits the opportunity of assembling cars more specifically to the local customer’s order.

Assembling directly to order reduces the risk of loss due to obsolescence. Completed products will become obsolete more often than many of their components, thus there is a definite advantage in not assembling before a customer actually has ordered a product. In an assemble-to-order environment, delayed differentiation can allow production of some standard components while the customer specific features and components can be added as late as possible.

The principle of delayed differentiation can also enable higher automation in assembly. Nevins and Whitney showed in 1978 that robotic assembly could be the most cost efficient assembly method for production volumes above approximately 125,000–300,000 units per year [144]. Later this limit has moved down to less than 50,000 products per year.

Lien has shown that by designing a product with many common base components the earlier stages of an assembly process can be automated, leaving the individual variation to the last stage that will be manual [123]. The automation relies on similarity in the assembly operation so that no reconfiguration of the automatic section of the assembly line is required. Change of the components in magazines or feeders is acceptable as long as it does not lead to long and costly changeover periods. This principle has been applied with success in the manufacturing of consumer white goods for more than two decades. The economical batch sizes of individual products can be as low as 100 in product families where the total yearly volume of the whole family is higher than 100,000 per year.

To take advantage of material flow simplicity by delayed differentiation it is sometimes necessary to redesign the product. The main objective of this redesign is to obtain commonality in as many of the base components as possible within a product family. Such redesign requires the joint effort of product designers and production engineers. When properly performed, such redesign leads to considerable simplification of the assembly process and reduction in total assembly time [122].

Delayed differentiation strategy influences the layout of the assembly line. In the system shown in Fig. 10, the logical material flow has an inverted T-shape. The physical implementation has been an L-shape where the vertical bar represents an automated serial line while the horizontal bar represents an automatic distribution system to a set of manual assembly station that operates in parallel. The material supply to the manual stations has a control system that ensures stable supply to each worker where the workers are pacing the automatic line [121]. Balancing of assembly system configurations with delayed differentiation was presented in a recent paper by Ko and Hu [99].

System configuration for delayed differentiation assembly lines involves the determination of points of product differentiation. Selecting these points is intertwined with the products family design since integrated product modules will be assembled on the assembly stations between successive points of differentiation. Determining points of differentiation has been traditionally associated with strategic objectives such as capital investment and inventory holding costs as well as Work-in-Process level (WIP); but only very few points of differentiation were considered in the literature [59,64,72]. Dynamic programming was used to locate the positions of maximum of two points in the assembly line [31,82], Swaminathan and Tayur [175] and Swaminathan and Lee [174] realized assembly system synthesis by considering assembly processes precedence constraints in positioning points of product differentiation.

Some studies connect system configuration, where points of differentiation are determined followed by establishing product families, to assembly line balancing. AlGedawy and ElMaraghy [5] proposed an integrated design model of assembly systems where layout, assembly sequencing, task assignment and station location decisions are combined in the optimization process for a mixed-model product assembly. This model adapts Cladistics – a comparative data analysis tool used in Biology to establish the evolutionary course of living organisms – to perform commonality analysis of the studied products. The cladistics analysis is further modified to incorporate functionalities not needed in biological analysis, namely the assembly processes precedence constraints and the required production rates for each product variant. The branching points of the resulting most parsimonious and optimal cladistic tree of the product variants and families represent the best location of points of product differentiation without any limitation on their number. Fig. 11 illustrates an assembly line produced by this new DDP assembly line design model for a family of automobile engine idlers and belt-tensioners [5].

### 3.4. Performance evaluation

Quality, cost and delivery have been the most important measures of manufacturing systems performance, but performance metrics in terms of environmental sustainability and social/ operator impact are gaining attentions. Quality is usually measured in terms of closeness to design specification of certain key product characteristics [102]. Hu [83] presented a “stream of variation” model in predicting the quality of automotive assembly by combining engineering structural model with statistical analysis where variation due to part compliance were considered. Camelo et al. extended such a model to multi-stage assembly systems [28] and developed diagnostic approach for assembly systems when multiple process faults exist [29]. Hu and Stecke [84] applied such models to assembly systems with different configurations and compared the performance of assembly system configurations. However, very limited work exists in analyzing assembly system quality when multiple products are produced in the same system.

Throughput is another important measure of assembly system performance. Significant research has been conducted in this area. Gershwin’s book [66] provides an excellent introduction to analytical methods for analyzing manufacturing system throughputs. Li et al. [117] presented reviews of recent advances and future research topics in manufacturing throughput analysis. Colledani and Tolio [25] presented a decomposition based methods for analyzing manufacturing system throughput in designing configuration and reconfiguration. Li and Huang [118] presented a
Other flexibility measures of manufacturing system performance include scalability [172] and convertibility [125]. Scalability is the ability to easily modify the production capacity of a system by adding or subtracting manufacturing resources (e.g. machines) and/or changing components of the system. Convertibility measures the ease of a manufacturing system in changing its functionality, for example, in response to different product types. Wiendahl et al. [196] provided a review of manufacturing changeability at all levels of the enterprise.

Cost is one final measure of a manufacturing system. A generalized cost model for assembly processes which is based on the Activity Based Costing (ABC) technique and combines major assembly cost factors into a single model can be used as a tool to support decision making during both the design and operation of assembly systems. Such a model is applied to an automotive case study [130]. A model has been proposed for dealing with the uncertainty of a customer’s potential acceptance of a delivery date for his/her ordered product. The method estimated the probability, as to whether a customer will actually place the order once s/he has received a potential delivery date for a product. In this instance, the manufacturing resources should be committed to this order. The Bayesian networks method is adopted and an automotive industrial case study is discussed [126].

The use of advanced simulation techniques, for the performance evaluation of assembly processes and systems, has considerably reduced the need for physical experimentation. Virtual Reality (VR)-based methods and tools have been specifically used for the evaluation of human-related performance aspects in assembly systems [150]. An immersive VR-based simulation environment has been developed to support process experimentation and verification, concerning factors that cannot be described analytically, and therefore, do not affect the process in a predetermined way [33]. The hybrid use of immersive interaction techniques and digital mannequin technologies, within an integrated simulation environment, have further facilitated the assembly performance evaluation, over a range of different human populations [34].

4. Assembly system operations

4.1. Process planning for assembly with high variety

Assembly is more than just putting parts together: it is the capstone process in manufacturing [195]. Assembly process planning is an important function in production planning and control of discrete-part manufacturing. The level of granularity in assembly planning refers to the amount and level of detail to be included. Assembly planning could be further classified into macro and micro planning. Macro planning is concerned mainly with tasks such as set-up planning, the identification of the assembly tasks and their sequencing to transform one configuration of parts into another [78,112]. Micro-level planning on the other hand focuses on finer details for each operation, such as detailed assembly steps, tool trajectory, collision avoidance, end-effector selection and generation of executable robot programs in case of robotic/flexible automated assembly.

Hommel de Mello and Sanderson [80] and Jones et al. [93] surveyed, classified and evaluated the different criteria in the literature and in available planning systems and applications. In addition to the popular objectives such as optimizing cost, time, performance, non-value adding tasks such as re-orientation, re-fixturing and tool changes, assembly specific criteria have been considered. Assembly planning is an NP hard problem. In the last two decades, increased application of non-traditional optimization methods, such as Genetic Algorithm (GA), Simulated Annealing (SA), Tabu Search, Ants Colony, Neural Networks. were witnessed. Artificial Intelligence based search methods were used for assembly sequence generation and assembly process planning for manual and automatic assembly and disassembly operations (see [10] for more detail).
As product variants increase, variant-oriented planning of their assembly processes becomes an important logical enabler to support such a change on both the assembly system and product level. Product versions are developed over time in response to the need for changes. Derivatives and variations in function, form and configuration lead to new product classes. This gives rise to product families and platforms that contain variants of the products and their parts, components and configurations, which evolve over time. It is essential to manage these product changes and variety in order to mitigate the resulting design planning and production complexities, as well as to prolong the life of their respective assembly systems and use their capabilities more effectively to produce the desired products variants.

4.1.1. Declarative knowledge representation for assembly planning

The amount and type of detail to be included in assembly planning is important. Whitney [195], and Cunningham et al. [38] adopted a unique view of integrating assembly planning within product design activities. Xu et al. [198] proposed a systematic approach based on process knowledge customization and meta-modelling of manufacturing resources. Interfacing with the CAD models of assembly is the first crucial step in assembly planning, since ultimately it is desired to be able to fully automate assembly planning and generate plans based on the geometric and relational input data and models of products and their variants. Ideally, preliminary assembly drawings, cross-sections, parts lists, as well as relational data such as liaison diagrams are needed to be able to start assembly planning and analysis as early in product design and development cycle as possible. Algorithmic methods such as the popular Cut-Set to generate precedence data needed for assembly planning are then invoked. The different graphical representations and data structures used to model precedence relationships and sequences include modeling and representation methods and algorithmic methods to generate implicit and explicit data representation structures.

4.1.2. Reconfigurable assembly planning

The time spent on design and assembly planning for a new product is generally considerable [112]. For products with long life cycles produced in large quantities and often assembled manually, such time investments are justifiable. However, present market conditions and higher product variety lead to much shorter life cycles and smaller production volumes. Therefore, the time spent on product development including assembly planning activities must be considerably reduced. Furthermore, the continuously evolving product and part families require new variety-oriented assembly planning concepts, models and tools. Azab and ElMaraghy [9] introduced a novel approach, which transforms the nature of process planning from sequencing to insertion (Fig. 12). This approach allows master plans of existing products to be re-configured on the fly to meet the requirements of new products/variants assembly operations, while minimizing the differences between the new and old plans. Therefore, instead of generating new plans for new variants from scratch, only portions of the master assembly plans, corresponding to new features/operations are generated and positioned optimally within the overall assembly plan.

A performance index was also formulated to evaluate the extent of reconfiguration of the process plan. The Plan Reconfiguration Index is used to evaluate the quality of the re-configured process plans. It is a measure of the extent of reconfiguration and changes that occur due to variety. This represents a new direction in process planning and a novel criterion aimed at minimizing the resulting disruption in down-stream activities on the shop floor [9]. This new Reconfigurable Process Planning (RPP) approach enables local reconfiguration of assembly plans when needed, where needed and as needed, while minimizing the extent of change/reconfiguration on the shop floor and the costs associated with making changes to existing facilities, tooling, labor training and quality concerns. It is applicable to macro assembly planning and was demonstrated in the assembly of a family of household electric kettles.

These novel models and algorithms satisfy the need to plan and re-plan assembly processes frequently for different product variants. This approach is applicable to companies with progressive design changes and low- to mid-volume assembly. They provide important logical enablers to cope with the continuously evolving products and their variants.

Co-evolution of product families and assembly systems has been proposed as a methodology for the joint design and reconfiguration of product families and assembly systems over several product generations [26]. The method includes the initial concurrent design of product family and assembly systems, and the subsequent reconfiguration of the product family and the assembly systems. Such a method capitalizes on the opportunities offered by the modular product architectures and reconfigurable assembly systems for design and assembly system reuse. A general methodology for co-evolution of product, process and production systems is described in a CIRP keynote by Tolio et al. [178].

4.2. Scheduling in mixed model assembly

Some recent research carried out on scheduling mixed model assembly lines concentrates on the determination of policies that can guarantee the maximum satisfaction of the performance criteria, set by each company. The performance of mixed model assembly lines characteristics, against these criteria, is dependent on the demand patterns that the production system is subjected to. For instance, a new scheduling policy, which is based on maximizing the slack time of the remaining tasks in the manufacturing resources queues, has proved to be providing improved tardiness performance for specific workload patterns [149]. The use of simulation in scheduling assembly systems is proving to be indispensable, since the NP-hard nature of the scheduling problem does not allow the identification of the optimal solution, within an acceptable time frame. Toward this goal, hybrid scheduling approaches, involving discrete event simulation capabilities, are being investigated [111]. The scheduling process, in mixed model assembly lines, operated by humans, is typically faced as a job rotation problem. The job rotation enables production systems to cope with the fluctuating market demand by exploiting the benefits of flexible workforce. It provides employees with a more engaging working environment, resulting in far less monotonous and repetitive tasks. A dynamic job rotation tool has been discussed that allows for the efficient allocation of assembly tasks to suitable operators, at any point of time, leading to more balanced workload distribution and thus, achieving a dynamic line balancing. A hierarchical approach to multiple criteria and decision making algorithms is used for the implementation of the tool. The tool generates alternative rotation schedules and evaluates them against any predefined criteria [131]. Additionally, a method has been proposed for dynamically querying supply chain partners to provide real time or near real time information, regarding the availability of parts, required for
the production of highly customizable products. This method utilizes Internet-based communication and real-time information from RFID sensors. The feasibility of this approach and its implementation are demonstrated in a typical automotive case [138].

As described in Section 3, model (build) sequencing problems are solved together with line balancing for mixed-model lines. The reason for the need for such integrated sequencing and balancing is that the multi-model line balancing solutions can minimize only the potential workload fluctuations due to model differences; the actual fluctuation in a smaller time scale should be mitigated by proper build sequences. The model sequencing is usually a short-term problem whose planning periods are typically in the order of days. The lot size is usually one per model in the case of mixed-model sequencing. The usual objectives of this model sequencing are to level out the fluctuation of workload and part usage. Some research papers solved the mixed-model balancing and sequencing problems in integrated ways [13,37]. The practical applications of the model sequencing include car sequencing in the automotive industry usually to limit the successive assembly of the same product models [69,70,151,169], and lean assembly lines to level the material consumption of different models for just-in-time part supply [12,43,45,46,132]. For a comprehensive review, refer to a survey paper by [25] that classified the mixed-model sequencing problems based on the characteristics of stations, lines and objectives.

4.3. Management of product variety induced complexity

As product variety increases, the assembly process in mixed-model assembly systems can become quite complex, which in turn influences the system performance. Such variety-induced complexity exists in many aspects of the assembly systems, including part supply, production scheduling, manual assembly operations, etc. Four main sources of product varieties may lead to increases in complexity of assembly processes [63]: (1) different basic functions within the same family of products, (2) adaptation of the same function to different requirements, (3) offering optional functions, and (4) non-functional requirements. One possible way to cope with this challenge is to model how product variety complicates the assembly process and operations and in turn influences the system performance.

Manufacturing complexity has been an active area of research in the past two decades. For example, complexity measures were defined based on the data collected at plants and the relationship between complexity and system performance was investigated. For example, MacDuffie et al. [124] established an empirical relationship between complexity and manufacturing system performance based on data collected from 70 assembly plants worldwide as part of the study by the International Motor Vehicle Program at M.I.T. The authors studied product mix complexity based on product variety (product mix and its structure) in assembly plants and according to the different level of product variety, three types of product mix complexity were defined: model mix complexity, parts complexity, and option complexity. In addition, MacDuffie et al. [124] also identified significant negative correlation between the complexity measures and the main bottleneck score based on statistical analysis.

Nakawaza and Sub [139] defined a complexity measure as the amount of effort needed to make a part in order to achieve a certain geometric precision and surface quality in machining. Based on that, Sub [173] defined complexity as a measure of uncertainty in achieving the specified functional requirement in the context of product design. Deshmukh et al. [42] derived an information-theoretic entropy measure of complexity for a given combination and ratio of part types to be produced in a manufacturing system. Fujimoto and Ahmed [62] derived a complexity index based on the assemblability of a product, which was defined as the uncertainty of gripping, positioning, and inserting parts in an assembly process. Fujimoto et al. [63] proposed a complexity measure based on product structure using information entropy in different assembly process planning stages. They claimed that impact of product variety on manufacturing systems could be reduced by reducing the complexity. Papakosta et al. [149] defined a manufacturing execution complexity based on the intrinsic structure of the system and the uncertainty related to the operations of the system.

Measuring the complexity of product assembly based on their assembly features and requirements supports the design of both products and systems for reduced complexity. It also helps rationalize the choice of various assembly processes, sequences, equipment and system configuration and layouts. Samy and ElMaraghy [159,160] defined product complexity as the degree to which the individual parts/subassemblies have physical attributes that cause difficulties during the handling and insertion processes in manual or automatic assembly. They developed a product complexity model that incorporates the quantity and diversity of information as well as the design for assembly (DFA) principles for assembled products into an earlier operations complexity model by ElMaraghy and Urbanic [52]. Design for Assembly (DFA) principles were used to define a relative point scale of the different assembly attributes used in both manual and automatic handling and insertion assembly processes. The product assembly complexity model is expressed as:

$$C_{\text{product}} = \frac{F_p}{N_p} + C_{\text{product}}\log_2(N_p + 1) + \frac{F_h}{N_h} \log_2(N_h + 1) \quad (1)$$

where $C_{\text{product}}$ is product assembly complexity, $N_p$ and $N_h$ are the total numbers of parts and fasteners respectively, $F_p$ and $F_h$ are the number of unique parts and fasteners respectively, and $C_{\text{product}}$ is the product assembly complexity index.

The second term of Eq. (1) represents the diversity and quantity of information related to fasteners, $F_p$, $F_h \geq 1$. The above complexity metric was demonstrated using two case studies of the assembly of an automotive piston and a family of three-pin electric power plugs. Higher complexity is proportional to longer assembly time in case of manual assembly, and more equipment cost in case of automatic assembly. The complexity metric can be used at early design stages to guide designers in selecting part features to reduce total product complexity.

4.3.1. Complexity induced by product variety

A new complexity measure was proposed by Zhu et al. [207] to characterize the manufacturing complexity caused by product variety in mixed-model assembly lines. This complexity, called operator choice complexity, takes the form of Shannon’s information entropy and integrates both product variety and assembly process information. Consider a certain station in a mixed-model assembly line, i.e. station $j$, where the operator needs to perform several assembly activities in a sequential manner. The assembly activities involve various choice-selection processes, such as part choice, fixture choice and tool choice. Suppose there are $K$ assembly activities at the station. For $k$th activity, $k = 1, \ldots, K$, there are $N_k$ different variants the operator needs to choose. The demand fraction of variant $v$ is $q_{vk}$, $v = 1, \ldots, N_k$, which also represents the probability that variant $v$ is selected at $k$th activity. The operator choice complexity for $k$th activity, $H^k_0$ is defined as

$$H^k_0 = \sum_{v=1}^{N_k} q_{vk} \log_2 q_{vk} \quad (2)$$

The definition of operator choice complexity is consistent with experimental observations found in classic human cognitive studies. Hayman [86] studied the performance of human choice activities by measuring how quickly a person can make a decision in response to different stimuli. He found that the average reaction time was approximately a linearly increasing function of information entropy conveyed by the stimulus. In one assembly station, the information entropy conveyed by stimulus is equal to the choice complexity of the station, $H^k_0$, defined by Eq. (2). Therefore, the larger choice complexity at one assembly station, the longer time the operator needs to make the selection. The total complexity of one assembly station is the sum of operator choice complexity.

Please cite this article in press as: Hu SJ, et al. Assembly system design and operations for product variety. CIRP Annals - Manufacturing Technology (2011), doi: 10.1016/j.cirp.2011.05.004
complexity of all these $K$ assembly activities.

$$H_j = \sum_{i=1}^{K} H_{ij}^2$$

(3)

Based on the complexity definition, introducing flexibility into the assembly lines can reduce complexity. For example, using common tools or fixtures for different variants can help to simplify the assembly process. Therefore choices (of the tools, fixtures, and assembly procedures) will be eliminated if flexible tools, common fixtures, or shared assembly procedures are adopted to treat a set of variants. Because fewer choices are needed to choose, complexity is reduced.

The complexity measures defined by Zhu et al. [207] focus on serial assembly lines. Wang and Hu [183] extended the complexity definition to mixed-model assembly systems with different configurations, including serial, parallel and hybrid of both. They also developed an approximate throughput model by taking into consideration the complexity-based operator reaction time and fatigue effect, based on which the impact of complexity on assembly system performance can be analyzed.

The use of hybrid system configurations enables variant differentiation through the parallel stations in a system. Through assigning modules to different stations and differentiating variants at different parallel stations, the number of options the operator needs to choose at assembly stations can be reduced and mix of these options can be adjusted. Therefore, the complexity of mixed-model assembly systems can in turn be reduced. A mathematical formulation based on mixed-binary non-linear programming (MBNP) is developed to find the mixed-model assembly system with minimum complexity [184].

4.3.2. Assembly system structural complexity

Modern assembly systems are comprised of a large number of entities (machines, transporters, buffers, data, infrastructure, etc.) and have high dimensionality, redundancy, interactions, and uncertainty. Their modules and characteristics are driven by the assembled products and their variants. Such systems are becoming more complex to operate, control and manage. Some of this complexity is due to the inherent structural characteristics of the equipment and layout configuration [48]. ElMaraghy [47] emphasized the importance of considering a manufacturing system itself as a product, that has a life cycle, and whose design, configuration, operation and complexity should be managed. Classification systems such as the OPTIZ Group Technology code were developed only for manufactured parts. The equivalence for manufacturing systems did not exist until the development of the Structural Classification and Coding System (SCC) by ElMaraghy [48], which captures the structural complexity inherent in the various types of equipment in a manufacturing system as well as their layout. Kuzgunkaya and ElMaraghy [109] applied it to assessing the structural complexity of manufacturing system configurations. The original equipment Chain type Structural Classification Code (SCC) has been extended [50] to include the assembly specific structural features of typical equipment used in products assembly systems such as assembly machines, transporters, buffers, parts feeders and handling equipment. Assembly equipment controls, programming, operation, power source, and sensors were also incorporated.

A code-based complexity metric has been developed [161] and used to measure the overall assembly system complexity and account for the number, diversity and information content within each class of the assembly system modules caused by the assembled products variety. Assembly machines complexity, for example, is represented by:

$$C_M = \left[ \frac{n_M}{N_M} + f_M \right] \log_2 (N_M + 1)$$

(4)

where $C_M$ is the machine complexity, $N_M$ is the number of assembly machines, $n_M$ is the number of unique assembly machines (an indicator of diversity), and $f_M$ is the average complexity index of the $N_M$ assembly machines. Similar formulas were developed to represent the complexity of material handling and buffer equipment.

The total assembly system complexity is represented by:

$$C_{\text{system}} = w_1 C_M + w_2 C_{\text{AHMS}} + w_3 C_B$$

(5)

where $C_{\text{system}}$ is the total assembly system complexity, $C_M$, $C_{\text{AHMS}}$, $C_B$ are machine, material handling, and buffer equipment complexities respectively. The weights, $w_i$ represent the relative importance of the complexity of the three equipment classes.

The developed assembly system complexity metric can be used by system designers to compare and rationalize various system configuration alternatives and select the least complex assembly system that meets the requirements.

5. Roles of human operators in assembly

The role of human operators in assembly has been quite different in the varying types of production systems that exist. One extreme is the high volume assembly line in which manual operators work on sequential assembly tasks at a machine paced rate. This was seen as an efficient way of assembly work when it was introduced by Henry Ford on the Model T assembly line [57]. However, it had its clear limitation in its dependence on all workers who were thought of as being able to perform at full rate throughout the working day. The opposite end of the scale is the complete assembly in one workplace by one operator only. Although it is not feasible for large products like cars, it is often used for a less complex assembly even in medium size production volumes. The advantage of the individual workplace solution has been claimed to be that it gives higher degree of work satisfaction and a much larger flexibility [55].

Most modern assembly systems are found to be between these two extremes [105]. The introduction of robotic solutions along with traditional automatic assembly lines has removed human operators from the monotonous and often stressful high pace work on sequential assembly lines. However, still many assembly operations are so complex that human assembly workers are the most efficient solution [20]. In some cases, manual operations are the only options. Studies have also shown that manual workers perform better over time when they are given varying tasks [55]. Thus, the assembly line principle introduced by Ford is not an optimal solution. Different ideas on how to combine the efficiency of the sequential assembly line with the versatility and flexibility of the human operator in operator paced assembly systems have led to new concepts of efficient flexible assembly systems. These systems combine the effectiveness of robots and other machines where they perform best with human operators.

A good example of different automation degrees can be found in today’s car manufacturing (see Fig. 13). As to the press shop, body shop and paint shop, there is usually a high automation degree. Humans are mostly responsible for observing the production and logistic aspects only. As to the powertrain area and final assembly area, the situation is different. The automation degree is usually low and humans are responsible for the assembly due to complexity and flexibility aspects.

Humans are much better than machines when it comes to adapting to small variations in assembly tasks. Therefore, they will give a more stable output even if there are slight variations of each of the products they assemble. Nevertheless, there is a trend regarding an increase of the degree of automation – especially in the final assembly area [155,190]. Examples concerning automatic solutions in the final assembly area are given in Fig. 14. Standardization and modularization are enablers for effective automatic solutions.

In the automotive assembly, many components are heavy. Even tasks that cannot be fully automated can be performed semi-automatically by using passive, cooperating robots. These robots
In order to avoid that complex final tasks will become a source of frequent stops in an assembly process, a different philosophy can be applied. An assembly line that uses a simple sequential process at the stable early stages and splits into parallel operations to perform the complex final operations has been proven to be more efficient than simple sequential lines. In the sequential line, automatic operations can be used with great efficiency. Stable operations can be obtained as the tasks are standardized and have little or no variation between product variants. For the final operations, manual work is the best alternative. The workers can easily change their tasks in order to assemble different variants of the product. Letting each of the parallel operations complete the assembly task will allow the workers to operate at their individual pace [123].

It is not always easy to separate the tasks that can be performed automatically from the tasks that require manual assembly along an assembly line. The approach has then been to investigate the possibility for combining manual and robotic assembly in one work station [36]. The rules for the operators’ safety interdict the use of work places where operators have to put parts of their body inside the workspace of a robot operating in automatic mode [146,158]. In several research projects, methods to overcome this problem have been discussed and investigated [6,137,201]. The most promising approach seems to be the use of vision systems to block the robot from moving into areas where human hands, arms or other parts of the body are observed. An example of such systems is shown in Fig. 15. Although projects have shown that such methods are viable, they have not yet been put into industrial use.

A different approach is the development of lightweight robots with force sensing and sufficiently low kinetic energy to stop an unintended harmful interference between humans and robots in human–robot combined assembly operations [1]. In recent years, such robots have entered the market. They will enable new exciting possibilities for safe human–robot operations if the robots can be controlled so that the robot can never damage any part of the human body by contact force. The robot can then be an active supplier of components and perform many insertion tasks in collaboration with a human.

Fig. 14. Examples concerning automatic solutions.

Fig. 15. (a) A manual–robotic workplace observed by an electronic camera. (b) Areas identified as belonging to the human operators, used to prevent the robot to move into these areas [106].
In some of these concepts of robot and human cooperation, one possibility is to let the robot serve the human. The human operator can always set the pace to suit his or her capacity at any time. This can lead to a reduction of work induced damage to the muscle and skeleton system of the worker, which in turn increases the overall productivity of the individual [120].

6. Remanufacturing issues for products with high variety

The production process termed “Remanufacturing” can be utilized for the manufacture of some types of high-variety products, subject to logistical and economic constraints. In remanufacturing, used products are returned to a like-new condition, thereby recovering most of the added value from original production (including energy, raw materials, and labor) and reducing the environmental impact of the product. Studies have shown that remanufacturing provides an 85 percent energy savings over the production of new products and an associated reduction in the use of scarce natural resources [141,142]. During the remanufacturing process, recovered products, termed cores, are systematically disassembled into their basic elements. Following disassembly, components are cleaned and inspected for wear or damage. Damaged or worn components are replaced or restored and feature upgrades can be incorporated as required or requested by resellers. Finally the product is reassembled and re-qualified. For this reason, remanufacturing differs from other recovery processes such as recycling in its completeness: a remanufactured machine or component should match the same customer expectations, performance, reliability, and life cycle as new machines.

Remanufacturing incorporates the organizational principles of reverse logistics, which is defined as the “the process of planning, implementing, and controlling the efficient, cost effective flow of raw materials, in-process inventory, finished goods and related information from the point of consumption to the point of origin for the purpose of recapturing value or proper disposal” [74]. Fig. 16 depicts how remanufacturing extends the useful lifecycle of successive generations of built products beyond that offered by new manufacture. Because remanufactured products typically retail at a 30–50% discount from new builds, profit margins are tight; consequently the production infrastructure is typically arranged to output a standardized product with very limited capacity for customization.

It follows that the challenges to the use of remanufacturing to create products with high variety are primarily logistical and economic. High variety implies extensive inventories of customizing parts that must be sourced, sorted and maintained, generating additional costs over the price of standard cores and components. Furthermore, additional parts require labor-intensive manual cleaning and individualized inspection before they can be safely inventoried. Testing techniques for additional parts must be adjusted to accommodate different materials, possibly requiring supplementary test equipment and operator training. Finally, reassembly and re-qualification of occasional, low volumes of customized products will always be greater than the same operations for a steady high-volume production flow.

While remanufacturing traditionally cannot exist for products with high variety due to the added complexity and cost of processes, it is technically possible to do so if products are designed with remanufacturing in mind and the level of customization is at a level that allows for standard product configuration. Variety customization is then added to the base product configuration. In remanufacturing, the added product customization elements would follow the same strategy of new product manufacturing. Remanufacturing would only be carried out for the base product configuration. Customization will then be added to the product, during the remanufacturing process, with new customized components. While this might result in recovery of smaller percentage of material content in products, the level of customization added might increase the value of the product to the end user.

In this scenario, the economics of remanufacturing will be more challenging than that of limited variety products. However, with expected higher value of high variety products, this can result in a new model for remanufacturing that allows for a way to deal with products with fast pace of technology advancement to be able to achieve higher level of recovery and value to the end user.

Many models of Xerox copying machines, for example, allow for the ready substitution of different and/or upgraded user-interfaces, body panels and covers, and other visible cosmetic exterior components while retaining the same essential core imaging and paper-transport mechanisms, frame, power supplies, and other non-visible interior parts that can be repeatedly remanufactured. Other candidates for high-variety remanufacturing include certain office cubicles and related furnishings, which can be readily customized with replacement fabric panels, desktop finishes, and contemporary trim to accommodate changing expectations and tastes.

The general worldwide trend, fueled by global consumer demand, toward “greener” products and production practices should accelerate the marketing of a much wider variety of products designed for remanufacturability. As this occurs, the increased supply of cores and availability of smart automated assembly lines as may well enable the emergence of high-variety remanufacturing as a new industrial sector.

7. The next paradigm–product personalization and its assembly challenges

The manufacturing industry has evolved through several paradigms since its birth two centuries ago (Fig. 17). The first paradigm was “Craft Production”, which creates exactly the product the customer requests but at a high cost. “Mass Production” allowed low-cost manufacturing of large volumes of products, enabled by interchangeability and dedicated manufacturing systems. However, the product variety offered by such production was very limited, as evidenced by the famous quote...
from Henry Ford, “Any customer can have a car painted any color that he wants so long as it is black” [58]. In the late 1980s, “Mass Customization” [153] emerged as a new paradigm in response to consumer demands for higher product variety and manufacturers started to offer larger numbers of product “options” or variants of their standard product. This was cost-effectively accomplished by designing the basic product options and allowing the customers to select the assembly combination that they prefer most. Such an approach allows the manufacturer to achieve economy of scale at the component level, and use reconfigurable assembly systems to create high variety for the economy of scope of the final assembly.

Consumers’ desire to influence and participate in the design of products will lead to a new paradigm of manufacturing: personalized products and production. Personalization of products and services to gain market share has been advocated as a business strategy for the past twenty years [108]. For example, Levi Strauss & Co. launched “Original Spin” in several stores in the U.S. in 1997 to market personalized jeans. Stores were setup with white-light laser scanning systems to produce a digitized image of a customer’s body shape for tailoring the jeans. But Levi closed down “Original Spin” in 2003 due to lack of custom manufacturing support (Levi closed its manufacturing plant in the US before 2003), insufficient infrastructure and training, and poor customer experience (e.g. response time and service) [152].

Various companies are now marketing personalization even though their practices are best described as mass customization. For example, IndiDenim offers “custom jeans as unique as you” by providing a potential buyer many choices of “fabric, features and fit” (http://www.indidenim.com). Nike offers customized sports shoes on a number of base models where the customer can choose the color and finish for “base, eyestay, tip, swoosh, heel, and lace and lining” (http://nikeid.nike.com). Such customization is achieved through delaying differentiation in manufacturing; essentially the shoes are all made in one gray color with the customized color and finish achieved at the end of the production line.

7.1. Differences between customization and personalization

As noticed above, the terms of customization and personalization have been used without differentiation in the literature and public media. However, we believe they represent very different paradigms of manufacturing with clearly different goals and consumer participation. Table 1 summarizes the key differences between these two paradigms in terms of production goals, customer involvement and the requirements for manufacturing systems. Mass customization is to achieve economy of scope through market segmentation by designing variants according to a product family architecture and allowing customers to choose the design combinations. The goal of personalized production is for value differentiation by engaging customers in the design process and achieving efficacy of the products. Personalization is enabled by a product architecture where module differentiation can be achieved to allow space for customer design involvement. Koren [101] also delineates the key differences between customization and personalization from the process steps of Design, Make and Sales.

While product family design methodologies for mass customization were based on products that consisted of common modules and customized modules [91,168], we envision that a personalized product will typically have an open architecture and will consist of three types of modules: common modules that are shared across the product platform; customized modules that allow customers to choose, mix and match; and personalized modules that allow customers to create and design (see Fig. 18). All these modules will have standard mechanical, electrical and informational interfaces to allow easy assembly and disassembly. Based on the anticipated cost of the product, some designs may not contain all three types of modules but instead be composed of just the customized and personalized modules.

One example of this approach is the offering from the office chair manufacturer HAG which enables customers to bring in leather from their own wild game hunting for seat and backrest, and personally selected wood for the armrest.

7.2. Technology enablers for personalization

To make personalized production a cost effective reality, several enabling technologies must be developed. These technologies include:

- **Methods and tools for understanding and capturing consumer needs and preference**: The variability in terms of customer physical characteristics and preferences are very high. How such variability can be construed into design domains and parameters is a significant challenge. Quantitative models from human factors, behavioral economics, marketing and psychology can be integrated together to express such variability in a quantitative manner that is useful for design synthesis.

- **Designing by non-designers**: Currently available design methods and tools are for “designers”. As consumers are involved in the design process, new enabling environments must be created to allow non-designers to design aspects of products. These environments should facilitate the involvement of non-designers, or a group of them, in exploring a feasible design space without creating frustration or a sense of being overwhelmed.

- **Cyber system for collaboration**: Systems must be developed to allow open collaboration and data sharing across consumers, between consumers and manufacturers.

- **On-demand manufacturing systems**: Manufacturing systems must be designed to be responsive to the needs of individuals. Such a system may be distributed in its configuration, and may consist of dedicated machines, reconfigurable assembly systems, as well as special purpose, fast response systems for fabricating and assembling personalized modules.

- **Assembly system**: We believe that the assembly systems for personalized products will consist of subsystems of varying degrees of flexibility and reconfigurability due to the three types of product modules, common modules, customized modules, and

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Key differences between mass customization and personalized production, in comparison with mass production.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal</strong></td>
<td><strong>Mass production</strong></td>
</tr>
<tr>
<td>Customer involvement</td>
<td>Economy of scale</td>
</tr>
<tr>
<td>Production system</td>
<td>Buy</td>
</tr>
<tr>
<td></td>
<td>Dedicated Mfg System (DMS)</td>
</tr>
</tbody>
</table>

Please cite this article in press as: Hu SJ, et al. Assembly system design and operations for product variety. CIRP Annals - Manufacturing Technology (2011), doi:10.1016/j.cirp.2011.05.004
personalized modules. It is possible that a dedicated assembly system will be used for assembling the common modules, reconfigurable assembly systems for the customized modules. But the systems for assembling the personalized modules must be truly flexible, with grippers and fixtures that are infinitely adaptable within the design constraints. These systems will be configured and integrated in a way to enable delayed differentiation.

8. Summary and conclusions

Assembly is an important manufacturing process for cost effective product variety. Significant research has been done in the design and operations of assembly systems in support of high product variety, but many opportunities exist for future research. Some of these opportunities are highlighted below:

- An enabling method for assembly system design is the development of assembly representation methods that can support collaborative design and development. This is becoming increasingly important as globally distributed manufacturers and suppliers collaborate.

- Currently available methods for sequence generation assume sequential tasks. Consideration of assembly hierarchy and parallel assembly allows other non-sequential sequence choices which may lead to simplified sequence generation and innovative system configurations.

- Most assembly line balancing algorithms are for model, serial assembly lines. Balancing of non-serial, complex configuration considering product differentiation requires new, efficient algorithms.

- Design of reconfigurable assembly systems by incorporating both machines and people can lead to cost effective product flexibility and scalability.

- As the manufacturing paradigm evolves from mass customization to personalization, new assembly design methods in support of open product architecture and personalization design needed to be developed.

- Personalized products require the support of on-demand manufacturing systems. Such a system may consist of a combination of dedicated machines, flexible machines, and rapid manufacturing systems integrated in new types of system configurations. Design and planning of such manufacturing system represent new opportunities for research.

Acknowledgments

The authors wish to thank Professor S. Kaka and other reviewers for their review and constructive suggestions. Help provided by Robert Riggs is also greatly appreciated.

References


Please cite this article in press as: Hu, S.J. et al., Assembly system design and operations for product variety. CIRP Annals – Manufacturing Technology (2011), doi:10.1016/j.cirp.2011.05.004


