A Survey of the Literature on the Multiple or Parallel Assembly Lines Balancing Problem

Amaia Lusa (amaia.lusa@upc.edu)

Institute of Industrial and Control Engineering (IOC), Universitat Politècnica de Catalunya (UPC), Avda. Diagonal 647 p11, 08028 Barcelona (Spain)

Abstract

This paper focuses on production systems that consist of multiple parallel assembly lines. The main literature contributions are briefly described and used to present a summary of the state of the art. The advantages and disadvantages of adopting multiple lines are discussed and the multiple assembly line balancing problem and its relevant characteristics are described.

Keywords: multiple assembly lines, parallel assembly lines, assembly line balancing problem

Biographical note: Dr. Amaia Lusa is a lecturer of Industrial Engineering and operations research at the Universitat Politècnica de Catalunya (UPC). She has a degree in Industrial Engineering from the UPC. She holds her PhD on manpower planning (annualised working hours planning problem). Her research activities, at the Institute of Industrial and Control Engineering of the UPC, are focused on modelling and solving, by means of optimisation techniques, industrial engineering problems. She is author or co-author of books and papers.

1. Introduction

This paper collects and analyzes literature contributions related to the design and management of production systems that consist of multiple assembly lines. The survey is based on the analysis of existing literature and offers a structured overview of parallel assembly line design and management problems and the methods and solutions proposed to date. The survey also identifies features and problems that have not yet been studied and opens new lines of research.
An assembly line is a production system made up of a series of workstations that are connected by a conveyor belt or a similar system that transports the object that is being assembled. A set of tasks is performed at each workstation and the final assembled product is obtained at the end of the line. In contrast to the traditional process (dating from before the 1920s), in which all tasks are performed at a single workstation (i.e., the entire assembly is performed by just one person), these types of systems allow companies to achieve high productivity levels. As a result, assembly lines are used in most high-consumption industries in which the products are highly standardized and are also increasingly used in companies that produce or assemble small quantities of products with a low level of standardization.

Of the different problems that arise in the design and management of an assembly line, balancing problems are probably the most frequently studied in the literature. Balancing an assembly line consists in determining the number of stations (i.e., the number of workers) and assigning some of the tasks or operations into which the overall process is divided to each of these stations. The solution to the balancing problem may be affected by different conditions (incompatibility of tasks, space constraints, etc.) and, in all cases, it must be guaranteed that the precedence constraints are fulfilled (for organizational or technical reasons, some tasks cannot be performed until others have been completed) and that the time needed to perform the tasks assigned to a station does not exceed the available time (cycle time).

In the literature, the Simple Assembly Line Balancing Problem (SALBP) is undoubtedly the most commonly studied balancing line problem. In this problem it is assumed that there is one homogeneous product (no variants exist) to be assembled; all the details of the production process (the tasks and tools or machines) are known; the available time is the same for all stations (the cycle time of the line, which is known); the operation times are known and deterministic; no additional constraints are considered (besides precedence constraints); and the stations are serially distributed. The objective is to maximize the line efficiency (the sum of processing times divided by \( m \cdot c \), being \( m \) the number of stations and \( c \) the cycle time) which, depending on the version of the SALBP, means to minimize the number of stations with a given cycle time (SALPB-1), to minimize the cycle time (i.e., to maximize the production rate) with a given number of stations (SALBP-2) or to simultaneously minimize the cycle time and the number of stations (SALBP-E). When there is a known cycle time and number of stations (the pure version of SALBP) the objective is to find a feasible solution. Previous versions can be complemented by considering the secondary objective of smoothing the workstation times (Rachamadugu and Talbot, 1991; Merengo et al., 1999; Pastor and Corominas, 2000). Further details on the SALBP can be found in the surveys
contained in the literature (Buxey et al., 1973; Baybars, 1986; Shtub and Dar-EI, 1989; Ghosh and Gagnon, 1989; Erel and Sarin, 1998; Scholl, 1999; Rekiek et al., 2002; Scholl and Becker, 2006).

However, the SALBP is often a simplification that bears little resemblance to a real-life situation. For this reason, there is a growing body of literature on the Generalized Assembly Line Balancing Problem (GALBP), which usually includes more realistic features (multi-model or mixed-model lines, tool selection problems, parallel stations, U-shaped lines, stochastic times, constraints affecting the assignment of a certain task to certain stations, etc.) and overcomes some of the limitations of the SALBP. The most common objectives of the GALBP are to minimize the overall cost of the assembly line and to maximize profit. A detailed discussion of the costs that should be considered in assembly line balancing problems can be found in Amen (2000 and 2006) and in the survey on GALBP carried out by Becker and Scholl (2006). Furthermore, the use of more realistic problems with characteristics that were not previously considered has given rise to a new classification system for assembly line balancing problems (Boysen et al., 2006).

Recently, many companies have established multiple lines, with the lines having longer cycle times and each station being responsible for more tasks than under the traditional model (Chiang et al., 2007). Indeed, among the several suggestions for future research in assembly systems identified by Ghosh and Gagnon (1989), one is to allow some strategic design considerations, such as the number of lines, to be variables. However, the problem in which two or more assembly lines (parallel or multiple lines) may be present has not been widely studied as yet. This type of system has a higher equipment cost (Daganzo and Blumenfeld, 1994) but is both more reliable and more flexible (Rekiek et al., 2002; Becker and Scholl, 2006; Chiang et al., 2007) and allows for better balances, reducing the total number of required stations (Chiang et al., 2007). It also increases productivity (Chakravarty and Shtub, 1988; Boysen et al., 2006; Chiang et al., 2007), worker satisfaction and product quality (Ellegard and Jonsson, 1992). This type of system raises new problems regarding the number of assembly lines that should be opened, their characteristics (the number of stations and the cycle time) and the assignment of products and workers to the lines, in addition to the standard problem of balancing each assembly line (i.e., to determine, if necessary, the number of stations, and to assign tasks to stations).

The paper proceeds as follows: Section 2 describes the main literature contributions regarding the balancing problem with multiple or parallel assembly lines; Section 3 describes and depict the typical configurations of parallel assembly lines; Section 4 discusses the advantages and disadvantages (or
limitations) of adopting a production system based on multiple or parallel assembly lines; Section 5 presents the problems associated with balancing multiple or parallel assembly lines and the main characteristics and the key factors that must be considered; Section 6 contains two tables that provide a summary of the most relevant literature contributions on the multiple or parallel assembly lines balancing problem; and Section 7 presents the conclusions and proposals for new lines of research.

2. Main contributions

As stated in the Introduction, few studies deal with multiple lines, an understanding of which would help to detect features and problems that have not yet been covered and open new lines of research. Some general surveys and reviews (Ghosh and Gagnon, 1989; Erel and Sarin, 1998; Rekiek et al., 2002; Scholl and Becker 2006; Becker and Scholl, 2006; Boysen et al., 2006) discuss the possibility of using parallel lines, but they provide few references. In this section, studies that address some of the problems encountered in the parallelization of assembly lines are presented in chronological order, and comments are provided.

Lehman (1969) presents a heuristic procedure for assigning a set of products to a given set of parallel lines with the aim of minimizing the overall assembly cost. The study considers the balance delay cost (idle time due to balancing), the sequence delay cost (which is considered here for the very first time and is the operator time lost due to mixed-model sequencing) and the learning cost (the relation between worker productivity and the number of times a task is repeated), which is computed using learning curves. In addition to the heuristic, another of the main contributions of this paper is the proposal of an analytical model for measuring and evaluating the assembly costs.

Wyman and Moberly (1971) analyze the results provided by a simulation model to compare three parallel lines configurations. They study the influence on the results of a number of parameters, including the queue capacity, the failure rate, the repair rate and the number of stages.

Globerson and Tamir (1980) present a simulation model to examine the influence of the number of lines on human behaviour (learning, absenteeism and turnover) and, by extension, on the system response time (the time required to complete a unit from the beginning of the line to the end) and the percentage of work completed. The simulation model they used had been designed for a specific process (an insurance policy issuing process) and the results showed that, in
certain cases, a compromise strategy consisting of a few parallel lines is better than an extreme strategy (e.g., a single line of \( n \) stations or \( n \) parallel lines with one station each). The total number of operators (stations) is fixed and the number of stations is the same for each line in each case (e.g., one line with 12 stations, two lines with six stations each, etc.). It is assumed that the entire job can be divided into tasks in such a way that they can be assigned to the desired number of stations. Learning curves are used to estimate the time required to perform each task and stochastic operation times are also considered. The model estimates absenteeism and turnover over the standard time of the operation (which is related to how repetitive or monotonous that particular task is).

Chakravarty and Shtub (1988) propose an analytical model that can be used to create an optimal job design (i.e., an optimal number of lines), which is the design that minimizes the system response time. The model uses a set of mathematical equations to link the system parameters to absenteeism, turnover and learning, and these parameters to the system response time. It is assumed that the job can be divided into any number of tasks, that the total number of workers is fixed and that all the lines are identical. The results of a computational experiment show that a good job design is particularly important in new systems, in which the learning effect is more significant.

Ahmadi et al. (1992) propose three heuristic procedures to determine near-optimal solutions for a dynamic allocation problem (to assign models to lines). The number of lines is assumed to be known and all the lines are identical. The objective is to minimize a function of changeover costs and penalty costs (the costs incurred by failing to produce all of the requirements). The allocation problem is viewed as a problem of scheduling multiple types of jobs on parallel machines.

Süer and Dagli (1994) propose a heuristic procedure to solve the problem of assigning \( n \) different models to a varying number of lines with a varying number of stations in each line over a given period of time, such that the total number of available resources (stations) is not exceeded. The number of lines and the number of resources assigned to each model must be determined for each time period. The evaluation criteria are the average system response time and the makespan (which is computed using six different scheduling generation policies, once the assignment of models to lines is known).

The system may consist of any number of lines and each line may contain any number of stations; the only limitation is the total number of available resources (stations). All of the operations are assumed to be either manual or requiring light, low cost tools, which allows the aforementioned authors to assume that the number of lines and stations in a line may vary with time, even though once a
model has been assigned to a line the number of stations remains unchanged until
the entire batch is completed. The simultaneous assembly of a model on more
than one line is not permitted and each model must meet specific demands.

As many alternatives as the number of resources available are considered for
each model (one line with one station, with two stations, etc.). A priority matrix is
calculated for each combination model-alternative. A priority matrix is also
computed for each combination. Each element corresponds to the number of
stations (i.e., resources) needed to achieve the best line efficiency (which depends
on the cycle time). For each model (matrix file), the best values for the number of
stations are placed in the first column, the second best value is placed in the
second column, and so on. From the priority matrix, six different rules are used to
assign models to resources (stations). Next, different schedule generation
techniques are used to compute the makespan (the time in which the required
quantity, or each model, is completed). Consequently, the procedure produces 36
different solutions, the best of which is selected in view of the average system
response time and the makespan. The main drawback of the procedure proposed
by Süer and Dagli (1994) is the assumption that the entire job can be divided into
any number of operations and the fact that no precedence constraints are
considered.

Pinnoi and Wilhelm (1997) propose mixed integer linear programming (MILP)
models for various assembly line design problems. The authors consider four
different station configurations for parallel lines: (1) single machine; (2) parallel
machine (two identical machines in parallel); (3) collateral machines (two non-
identical machines in parallel); and (4) collaborative machines (which are similar to
collateral machines but can accommodate complementary tasks). Each product is
assigned to a dedicated line (i.e., a line that is used for only one type of product) or
to a flexible line (i.e., a line in which different types of products are processed),
although a maximum of one flexible line is admitted due to the cost involved.
Operation times and cycle times for each product are assumed to be known and
constant. The aim is to minimize the total cost (the fixed and variable costs of
machines and stations plus the fixed costs associated with each line opened plus
the assembly setup cost for each product assigned to the flexible line). The
authors do not solve the models but instead propose using cutting planes to solve
them.

Süer (1998) proposes a three-phase heuristic procedure for a case in which a
single model is to be assembled and the high production volume means that the
number of workers required is greater than the number of tasks. The problem
consists in determining the number of assembly lines (it is possible to use identical
lines, although it is not necessary) and the number of workers to assign to each
station (parallel stations are considered) with the aim of minimizing the total number of workers.

In first phase, which is called assembly line balancing, tasks are grouped at one station, two stations, three stations, and so on, up to a value of \( n \) stations (which would be the point at which one task is assigned to each station). The groups are obtained using a heuristic procedure. In the second phase, the parallel stations are determined: various alternatives are generated for each number of workers by considering various station configurations (one station, two stations, and so on, up to a value of \( n \) stations). The number of workers to be assigned to each station is determined using an integer linear programming model that is designed to maximize the production rate. The alternative that provides the best production rate for each number of workers is selected for the final phase, which consists in determining the parallel lines. An ILP model is used to determine the number and configuration of parallel lines (i.e., the number of lines for each of the alternatives selected in the previous phase) in order to minimize the total number of workers. As a result, some alternatives can be doubled, tripled, and so on. The authors provide a numerical example but not a complete computational study.

Miltenburg (1998) address the problem of balancing U-shaped lines in a multiple U-line facility. The U-line assembly line balancing problem (UALBP), which was introduced and modelled by Miltenburg and Wijngaard (1994), considers an assembly line arranged in an U-shape. Stations can be arranged so that during the same cycle two workpieces at different positions on the line can be handled. The difference to SALBP is that a station \( k \) can contain not only tasks whose predecessors are assigned to one of the stations \( 1, \ldots, k \), but also tasks whose predecessors will be finished until the product returns to station \( k \) for the second time (Monden, 1998). A station which handles the same workpiece in two different cycles is called a crossover station. With a U-line, the number of required stations can be smaller than with a regular line (for more details on the UALBP see Miltenburg and Wijngaard, 1994; Nakade et al., 1997; Sparling and Miltenburg, 1998; Urban, 1998; Scholl and Klein, 1999; Miltenburg, 2000; Erel et al., 2001; Chiang and Urban, 2006; and Urban and Chiang, 2006).

In his paper, Miltenburg (1998) considers the problem of balancing \( N \) U-lines with identical given cycle time and the objective of assigning tasks to a minimum number of stations while satisfying constraints on cycle time (operation and travel times are considered for each station), precedences, station-type and location. It is assumed that an existing building is used and that the areas in the building where the U-lines will be located are specified. The layout of the U-line and location of tasks restricts the ability of operators to cross from one side of the line to the other without interfering with nearby stations. A secondary objective is to concentrate
the idle time in one station so that improvement efforts can be focused there. Three type of stations are considered: (1) regular (those performing a single group of tasks organized in a continuous sequence, as in a straight line; no more than five or six tasks are allowed to be assigned to a regular station); (2) crossover (includes two groups of tasks located on different sides of a U-line); and (3) multiline (includes tasks from two adjacent U-lines). In the paper, no more than one multiline station is permitted between adjacent U-lines, and it is located to include tasks from the “end” and the “start” areas of the adjacent U-lines (being the “start” area where materials enter and finished parts leave and the “end” the area adjacent to the “start” of the next U-line). With a Dynamic Programming algorithm, optimal solutions for small problem instances are obtained (any number of lines but no more of 22 tasks for each line and no wide, sparse precedence graphs).

Sparling (1998) deals with the problem of balancing a JIT production unit, which is defined to be a group of U-lines, all producing parts for the same product line, operating at the same cycle time and located sufficiently close together that multiline stations may be built connecting combinations of the U-lines. Access to the U-lines is restricted to the “start area” of each U. Stations are not permitted to cross paths and so the number of multiline stations associated with any U-line cannot exceed two. A multiline station can include tasks from at most \( k \) U-lines. The problem, which is called the N U-line Balancing with Travel (NULB-T) Problem, is modelled (considering travel times and cycle time, precedence and location constraints), and heuristic solution algorithms are presented for both the general case, where U-line locations are not fixed, and for the fixed location case.

Gökçen et al. (2006) propose heuristic procedures and a mathematical model for the multiple or parallel assembly line balancing problem. The number of lines is known in advance and the aim is to balance more than one line in one step whilst minimizing the total number of stations. The authors consider three different cases: the passive case and active cases 1 and 2. In the passive case, two parallel assembly lines perform the same production process (with the same operation times and precedence graph) and cross-trained workers. To solve this case, the authors propose a procedure consisting in (1) solving the SALBP with any method and (2) reducing the number of workers by joining stations with an idle time equal to or greater than \( c/2 \) (where \( c \) is the cycle time). In other words, two parallel assembly lines share certain stations, which can in fact be seen as a single assembly line with parallel stations.

In active case 1, multiple assembly lines perform different processes (i.e., each line has its own precedence graph and operation times) but with the same cycle time. Gökçen et al. (2006) propose a heuristic algorithm for assigning tasks to
stations so that some of them perform tasks that belong to two adjacent lines (again, some stations are shared by two lines).

Active case 2 considers different parallel lines which perform different products with different cycle times. By multiplying the task operation times by a certain factor the authors consider the same cycle time for all lines and apply the procedure designed for active case 1. Consequently, the number of units produced in a cycle may not be the same for all lines. The authors also propose an ILP model for solving this case, which is designed to minimize the total number of stations.

The authors present the results of a computational experiment that solves 95 instances generated from 14 known instances. The ILP model is solved for instances with no more than 30 tasks. In 65 of the 95 instances the heuristic procedure gives a number of stations that is smaller than the number that would be obtained if each line were balanced separately.

Lan (2007) proposes a two-phase procedure to determine the preferred configurations from a list of feasible configurations for a production line, and the number of parallel machines to assign to each station, the production rate and the system operation time for each one. The number of machines of each type is limited and the quantity and due date of a single order of one product (whose production process is automatic and fixed) must be satisfied. The procedure provides a maximum profit solution that considers the income generated by the order and the following costs: material, setup, operation, maintenance and inventory holding costs (it is assumed that units produced must be stored until the entire order has been completed and the due date is reached) and fixed costs.

The first phase determines whether it is possible to design a system that will produce the required quantity within the allotted time (the maximum production that could theoretically be obtained is computed). If this is not feasible, the order is rejected; otherwise, the next phase begins. The second phase enumerates all feasible configurations (production lines) and calculates the theoretical production time and profit for each one. These are then used as the starting point of an iterative procedure that determines the operation time (i.e., the production) of each configuration, taking into account the previously calculated profits (the maximum production is assigned to the best configurations), and the due date of the order (the theoretical production times are modified to achieve the total quantity ordered).

Chiang et al. (2007) face the problem of balancing a just-in-time production system consisting of multiple U-lines. The main differences with the previous works done
by Miltenburg (1998) and Sparling (1998) is that in this case the number of lines is to be determined and that the U-lines are supposed to be independent, that is to say, there are not multiline stations connecting assembly lines. One of the purposes of Chiang et al. (2007) is to introduce and characterize the Multiple U-Line Balancing (MULB) Problem. Three objective functions are considered: (1) to minimize the total number of stations with a given global cycle time (MULBP type I); to minimize the cycle time with a given number of stations (MULPB type II); and (3) to minimize the cost (cost-minimization problem), which is considered as a linear function of the number of lines and the number of stations.

The MULB problem with identical lines (i.e., to determine the number of lines, the number of stations in each line and the assignment of tasks to stations) is efficiently solved by means of mixed integer linear programming (MILP) for up to 45 tasks (using a standard optimization software).

The problem with multiple U-lines of varying sizes is modelled by means of a nonlinear program whose linearization would result in a formulation impractical for large problems. Instead, Chiang et al. (2007) propose a procedure to solve large MULB problems which consists of two main steps: (1) to identify all the feasible nondominated solutions for the individual U-lines (a solution is nondominated if no other solution exists with the same number of stations and cycle time); and (2) to solve an integer linear program (ILP) to determine the number of lines of each type in the solution (the three aforementioned objectives are considered). As the authors point out, solving the ILP is extremely easy so the trick lies in efficiently identifying the set of feasible nondominated solutions (Chiang et al., 2007, propose an algorithm to do so). The proposed procedure is shown to optimally solve instances with over 100 tasks. This analysis is then extended to explicitly incorporate the effect of line configuration and task assignment on equipment requirements; a branch-and-bound algorithm is developed that can be used to identify optimal solutions for moderate-sized problems, or as a heuristic for large problems.

3. Typical configurations of parallel lines

The system configuration (i.e., the arrangement of the lines and the interconnection among the lines) has a significant impact on different key performance criteria, such as the investment cost of machines and tools, the quality, the throughput, the capacity scalability, the number of product types and the system conversion time (koren et al., 1998). Some typical machining system configurations are given by Spicer et al. (2002), who compare different options in
terms of throughput, line balancing, machine investment cost, and capacity scalability. Despite the importance of configuration, most papers dealing with multiple or parallel lines balancing problems assume a given configuration. Based on the work of Spicer et al. (2002) and the papers dealing with parallel or multiple assembly line balancing problems described in previous section, the following configurations, depicted in Figures 1 to 3, and the hybrids that could arise from these, can be considered as the most typical ones.

a) Independent straight or U-shaped pure parallel lines. The lines can be identical or different and each one can produce the same model (or group of models) or not (also, a model can be assigned to one or more than one line). Most of the papers described in section 2 consider this type of configuration.

b) Parallel straight or U-shaped lines with multiline stations: multiline stations are those that perform tasks of more than one line (i.e., stations shared by more than one line). Miltenburg (1998), Sparling (1998) and Gökçen et al. (2006) include the use of multiline stations in their works.

c) Parallel straight lines with crossover before some operations: crossover is where a product manufactured on a line blocked due to a failure downstream can be transferred to another line (Spicer et al., 2002). Note that this kind of crossover would be more difficult in a system with U-shaped lines.

Allowing multiline stations may result in a reduction in the total number of required stations. However, the main disadvantages are that the system requires that the connected lines are somehow synchronized and that a failure in a multiline station may affect the performance of more than one line. Crossover, which is not included in papers dealing with multiple or parallel line balancing problems, may be advisable when machines are likely to fail. Obviously, location or space issues may constrain the possibility of including crossover.
Figure 1. Independent lines

Figure 2. Multiline stations

Figure 3. Straight lines with crossover
Even though it is not shown in previous figures, any configuration may have parallel stations (stations that perform the same tasks). Installing parallel stations is required when the maximal task time is greater than the desired cycle time, but it can be also a way of decreasing the cycle time of the line and, therefore, increasing production rate (Buxey, 1974). Furthermore, having a system with parallel stations reduces failure sensitivity and, also, it often allows for better balances. Of course, as in the case of parallel lines, having parallel stations causes additional fixed costs that must be considered (readers are referred to Becker and Scholl, 2006, for more details on parallel stations).

4. Advantages and disadvantages of multiple or parallel lines

The parallelization of assembly lines is a strategy that may provide numerous advantages to both the company and the workers. However, there are certain limitations and drawbacks that should also be considered. These advantages and disadvantages are summarized below, based on the literature mentioned above.

The main advantages of parallel lines over a single line are as follows.

- They reduce failure sensitivity (Rekiek et al., 2002; Becker and Scholl, 2006; Chiang et al., 2007): because there is more than one assembly line, production is not interrupted when one of the lines breaks down (the others keep working).

- They increase system flexibility (Rekiek et al., 2002; Becker and Scholl, 2006; Chiang et al., 2007): each line can have its own cycle time (i.e., production rate), which makes it easier to adapt production to changes in the quantity and composition of the demand. Also, a multiple line configuration may enable a large number of alternate optimal solutions (Chiang et al., 2007).

- They improve the balance of the overall production system and thus its productivity (Boysen et al., 2006; Chiang et al., 2007): the inherent flexibility of a system that contains several parallel assembly lines provides a better balance (better line efficiency), decreases idle time and/or changeover time and increases the productivity of the system as a whole. Furthermore, the need for fewer stations, and the associated reduction in idle time, will result in lower labor costs, floorspace requirements and work-in-process inventory (Chiang et al., 2007). According to Chiang et al. (2007), when U-lines and
reduced cycle times are used, the benefits (saving in the number of stations) of multiple lines are significant.

- They increase the cycle time, which produces job enlargement or enrichment at each of the stations: if there are multiple parallel assembly lines, the cycle time of each one can be increased without decreasing the productivity that could be obtained by using a single line. An increase in cycle time can provide the following advantages:

  - Worker satisfaction increases (Ellegard and Jonsson, 1992). This reduces absenteeism (which is a cause of fluctuations in the production rate) and turnover (which decreases the production rate) and increases system productivity (Chakravarty and Shtub, 1988). The cycle time increase should be moderate, because a substantial increase would mean that the number and variety of operations that workers had to perform would prevent them from becoming specialists and would ultimately have a negative influence on productivity (in an extreme situation, a cycle time increase would lead to a very low production system in which a single worker performed all of the operations).

  - According to Ellegard and Jonsson (1992), improvements in quality can be obtained. This is arguable, because even though the increase in worker satisfaction might improve the quality of the work they perform the resulting decrease in worker specialization could have the opposite effect.

The most important drawbacks and limitations of adopting parallel lines are as follows.

- The cost of equipment increases: it is clear that if some stations (with the corresponding tools and/or machines) have to be duplicated, greater investment will be required and operation costs may increase.

- Each line must contain a minimum number of stations: otherwise there would be considerable material supply problems (due to physical distribution).

- A large amount of space is required: each operation may require space for the corresponding materials and each machine requires its own space.

- If the cycle time also increases (job enlargement):
The learning effect decreases (because the number of repetitions per period is reduced). In some cases this could even cause a decrease in productivity (Chakravarty and Shtub, 1988).

If the number of operations performed by a worker increases, the time needed to change from one task to another may no longer be considered negligible.

If one considers the drawbacks and limitations outlined above, it seems clear that the best solution will be the one that strikes a balance between a potential increase in cost and the increase in productivity and flexibility and that satisfies the financial and space constraints. Consequently, the number of lines to be installed is a key feature that must be determined when a production system with multiple or parallel lines is being designed and balanced.

5. Multiple or parallel line balancing problems

As mentioned in the Introduction, the number and complexity of decisions that must be taken regarding the design and management of assembly lines increases when the possibility of installing multiple or parallel lines is considered. Papers in the literature deal with four types of problems:

1) Deciding upon the number of lines to be installed (Süer and Dagli, 1994; Pinnoi and Welhelm, 1997; Süer, 1998; Lan, 2007; Chiang et al., 2007).

2) Assigning models to lines (Lehman, 1969; Ahmadi et al., 1992; Süer and Dagli, 1994; Pinnoi and Welhelm, 1997).

3) Balancing the lines: determining the number of stations for each line and assigning tasks to stations (Pinnoi and Welhelm, 1997; Süer, 1998; Miltenburg, 1998; Sparling, 1998; Gökçen, 2006; Lan, 2007; Chiang et al., 2007).

4) Evaluating the performance of the production system in order to compare different configurations (Wyman and Moberly, 1971; Globerson and Tamir, 1980; Chakravarty and Shtub, 1988; Süer and Dagli, 1994; Freiheit, Koren and Hu, 2004; Freiheit, Shpitalni and Hu, 2004; Li, 2004; Lan, 2007).

Since the assembly line balancing problem is NP-hard (the SALBP is NP-hard, so if parallel lines are considered the problem will be also NP-hard), most authors propose solving it using heuristic procedures. Some of these procedures are
based on simulation (Wyman and Moberly, 1971; Globerson and Tamir, 1980), others incorporate analytical models to estimate costs and model absenteeism or learning (Chakravarty and Shtub, 1988), and some use mathematical models in certain steps (Süer, 1998; Gökçen et al., 2006). Pinnoi and Welhelm (1997) are the only authors who propose dealing with the problem using mixed integer linear programming models (MILP), although they do not actually solve the models presented.

An important feature to consider in any attempt to solve the problem is the evaluation criterion (or the objective function that is supposed to be optimized). The criteria considered by the different authors are given below (some of them consider a function that takes into account a combination of more than one criterion).

- To minimize the number of stations (Süer, 1998; Miltenburg, 1998; Sparling, 1998; Gökçen et al., 2006; Chiang et al., 2007).

- To minimize the overall cycle time or, equivalently, to maximize the production rate (Wyman and Moberly, 1971; Li, 2004; Chiang et al., 2007).

- To maximize the efficiency of the lines (the percentage of productive time relative to the capacity). Idle times decrease efficiency; it is caused by two main factors:
  - Balance delay: idle time due to the balancing or the assignment of tasks to stations (Kilbridge and Wester, 1961; Sarker and Shanthikumar, 1983)
  - Sequence delay: idle time due to the sequence of products to be assembled in a flexible (multi-model or mixed-model) assembly line (Lehman, 1969)

- To minimize the makespan if there is an order to be delivered (Süer and Dagli, 1994).

- To minimize the system response time (the average time taken to complete a unit from the beginning to the end of the line) (Globerson and Tamir, 1980; Chakravarty and Shtub, 1988; Süer and Dagli, 1994).

- To minimize the cost or maximize the profit. There are several aspects that involve a cost (a detailed discussion of this can be found in Amen, 2000 and 2006, and in Becker and Scholl, 2006). The following are some of the most common costs discussed in papers that deal with multiple or parallel lines.
- Fixed cost of installing a line (Pinnoi and Welhelm, 1997; Chiang et al., 2007).

- Costs (fixed plus variable) associated with the stations (either workers or machines) (Sarker and Shanthikumar, 1983; Pinnoi and Welhelm, 1997; Lan, 2007; Chiang et al., 2007).

- Penalty cost for failing to cover a given demand or order (Ahmadi et al., 1992).

- Learning cost: if a worker has to perform different tasks—either because the cycle time is long or because the line is a flexible one in which different models require different operations—then the number of repetitions per period is considered to be low and the learning effect causes a decrease in productivity that can be translated into what is known as the learning cost [Nadler and Smith (1963) estimate this cost using learning curves in which the time needed to complete a task depends on the number of times that task has been performed].

- Flexible lines (multi-model or mixed-model lines) can lead to setup or changeover costs (Ahmadi et al., 1992; Pinnoi and Welhelm, 1997; Lan, 2007).

6. Summary

Tables 1 and 2 provide a summary of the most important literature contributions to the multiple or parallel assembly line designing and balancing problems. Table 1 presents the most interesting papers and gives a brief description of the problem covered and the solutions proposed or implemented. The following information is included for each of the selected papers.

- **Problem:** indicates the problem addressed in the paper.

- **Characteristics:** indicates the most relevant features of the problem.

- **Evaluation criteria:** indicates the criteria or combination of criteria that the authors use to evaluate the solutions to the problem.
- **Approach:** indicates the approach or technique that is used to solve the problem (heuristic method, simulation, mathematical programming, etc.).

In addition to the evaluation criteria, there is a group of features and characteristics that define and condition the difficulty of the problem to be solved, the solving procedure and the evaluation criteria themselves. In Table 2, the papers selected are classified according to a group of key features that must be defined in every multiple or parallel assembly line balancing problem. The summary allows us to detect the areas that have not yet been studied in the literature and to open new lines of research. The following is a list of the main features that were considered when Table 2 was created.

- **Problem:** includes a brief description of the problem that is addressed in the paper.

- **Job division:** the division of jobs into tasks or operations. Two possibilities are considered: (i) the tasks are predetermined or (ii) the entire job can be divided into tasks in such a way that they can be assigned to the desired number of stations (free).

- **Number of products:** 1 (single model) or >1 (multi-model or mixed-model case).

- **Number of lines:** whether the number of lines is fixed or must be determined (in this case, the number of lines can or cannot be bounded).

- **Type of line:** the lines can either be identical or different (some configurations may be duplicated but this is not necessary) and, in the case of multiple products or models, the lines can be dedicated (a single model is assigned to a line) or flexible (several models can be assigned to a line). In the latter case, it is important to know the setup costs, the changeover costs and whether these costs are sequence-dependent.

- **Number of stations:** this refers to whether the total number of stations (workers or machines) and the number of stations in each line is fixed or must be determined (in this case, the value can or cannot be bounded).

- **Stochastic issues:** this refers to the stochastic issues that have been considered in the problem definition. These include absenteeism, learning, turnover, failures, operation or task times and setup or changeover times.
- **Evaluation criteria**: indicates the criteria or combination of criteria that the authors use to evaluate the solutions to the problem.

- **Approach**: indicates the approach or technique that is used to solve the problem (heuristic method, simulation, mathematical programming, etc.).
Table 1. Summary: problem/characteristics/evaluation criterion/approach

<table>
<thead>
<tr>
<th>Reference</th>
<th>Problem/characteristics/evaluation criterion/approach</th>
</tr>
</thead>
</table>
| Lehman (1969)                 | To assign models to assembly lines.  
Flexible lines. Fixed number of lines.  
Cost: balance delay, sequence delay and learning cost.  
Heuristic procedure. |
| Wyman & Moberly (1971)        | To compare different configurations.  
Parameters: queue capacity, failure and repair rates, and number of stages.  
Production rate.  
Simulation. |
| Globerson & Tamir (1980)      | To examine the influence of the number of parallel lines on human behaviour (learning, absenteeism and turnover) and therefore on the system response time and percentage of work completed.  
Fixed number of workers. The number of stations is the same for all lines. It is assumed that the entire job can be divided into tasks in such a way that they can be assigned to the desired number of stations. The model is designed for an insurance policy issuing process.  
System response time.  
Simulation. |
| Chakravarty & Shtub (1988)    | To decide upon the number of identical parallel lines to be installed.  
Fixed number of workers. The number of workers is the same for all lines. It is assumed that the entire job can be divided into tasks in such a way that they can be assigned to the desired number of stations.  
System response time.  
Analytical model to estimate the system response time from the number of lines. The model links the number of lines to absenteeism, turnover and learning and these parameters to the system response time. |
| Ahmadi et al. (1992)          | Dynamic allocation problem: to assign models to lines.  
The number of lines is fixed. The problem is viewed as one of scheduling multiple types of jobs on parallel machines.  
A function of changeover and penalty costs (costs incurred by not meeting all of the requirements).  
Three heuristic procedures. |
| Süer & Dagli (1994)           | To determine the number of lines, the number of resources (stations) assigned to each model and the assignment of models to lines for each period in a time horizon. The resources available for the time horizon cannot be exceeded.  
The number of lines must be determined. It is assumed that the entire job can be divided into tasks in such a way that they can be assigned to the desired number of stations.  
Average system response time and makespan (once the assignment of models to lines is known, this is calculated by using six different schedule generation schemes).  
Heuristic procedure. |
<p>| Pinnoi &amp; Welhelm (1997)       | To determine the number of lines and to solve the assembly line balancing problem. |</p>
<table>
<thead>
<tr>
<th>Author</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Süer (1998)</td>
<td>Each model is assigned to either a dedicated or a flexible line (no more than one flexible line is allowed). Cost: fixed plus variable costs of machines and stations plus fixed costs of installing lines plus setup or changeover costs for the flexible line. MILP (models are not solved). To determine the number of parallel lines and the number of workers to be assigned to each station (parallel stations). Single product. Number of required workers greater than the number of tasks. There may be identical lines. Total number of workers. Heuristic procedure (three-phase procedure including an IP and an MILP model).</td>
</tr>
<tr>
<td>Miltenburg (1998)</td>
<td>Fixed number of U-lines. Each line dedicated to one product. Three types of stations (regular, crossover and multiline). There can be one station connecting two adjacent U-lines (multiline stations). The location of the U-lines is fixed. Total number of stations. Dynamic Programming algorithm.</td>
</tr>
<tr>
<td>Sparling (1998)</td>
<td>Fixed number of U-lines. Each line dedicated to one product. Three types of stations (regular, crossover and multiline). The number of multiline stations associated with any U-line cannot exceed two and a multiline station can include tasks from at most k U-lines. The location of the U-lines may not be fixed. Total number of stations. Heuristic algorithm.</td>
</tr>
<tr>
<td>Gökçen et al. (2006)</td>
<td>To solve the assembly line balancing problem. Fixed number of lines. Each line dedicated to one product. Some stations can be shared by two lines (multiline stations). Total number of stations (workers). Heuristic procedures and a mathematical programming model.</td>
</tr>
<tr>
<td>Lan (2007)</td>
<td>To determine the configuration of parallel lines (selected configurations and number of parallel machines at each station), the production rate and the operation time of each line to comply with the quantity and due date of a single order. Static case (single order). Single model. The set of feasible configurations is fixed in advance. Profit (cost of machines and inventory holding costs). A two-phase progressive algorithm that obtains the optimal solution.</td>
</tr>
<tr>
<td>Chiang et al. (2007)</td>
<td>To determine the number of U-lines and to solve the assembly line balancing problem with multiple independent identical or different U-lines. Number of U-lines to be determined and lower and upper bounded. Independent lines. All lines dedicated to the same product. Two types of stations (regular and crossover). No location constraints. Total number of stations (with a given cycle time). Cycle time (with a given number of stations). Cost of lines and stations (with a given cycle time). Mixed Integer Linear Programming. Branch and Bound for the case that incorporates the effect of line configuration and task assignment on equipment requirements.</td>
</tr>
<tr>
<td>Paper</td>
<td>Problem</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lehman (1969)</td>
<td>Assigning models to lines</td>
</tr>
<tr>
<td>Wyman &amp; Moberly (1971)</td>
<td>Evaluating a system</td>
</tr>
<tr>
<td>Globerson &amp; Tamir (1980)</td>
<td>Evaluating a system with different numbers of lines</td>
</tr>
<tr>
<td>Chakravarty &amp; Shhtub (1988)</td>
<td>Determining the number of (identical) lines</td>
</tr>
<tr>
<td>Ahmadi et al. (1992)</td>
<td>Dynamically allocating models to lines (scheduling problem)</td>
</tr>
<tr>
<td>Süer &amp; Dagli (1994)</td>
<td>Determining the number of lines and stations and assigning models to lines</td>
</tr>
<tr>
<td>Pinnoi &amp; Welhelm (1997)</td>
<td>Determining the number of lines and assigning models to lines</td>
</tr>
<tr>
<td>Süer (1998)</td>
<td>1) balancing the line 2) determining parallel stations 3) determining parallel lines</td>
</tr>
<tr>
<td>Gökçen et al. (2006)</td>
<td>Balancing lines (assigning stations to lines and tasks to stations). Multiline</td>
</tr>
<tr>
<td>Study</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lan (2007)</td>
<td>Selecting line configurations, determining the number of parallel machines for each station, determining the production rate and operation time for each line. Single order.</td>
</tr>
<tr>
<td>Chiang et al. (2007)</td>
<td>Determining the number of U-lines and balancing the U-lines (assigning stations to U-lines and tasks to stations)</td>
</tr>
</tbody>
</table>
7. Conclusions and further research

Production systems based on the use of assembly lines provide high productivity levels. When multiple or parallel lines are installed additional costs may arise (in relation to equipment and space) that can be compensated for by an increase in productivity and quality. In any case, it is clear that in practice conclusions on the advisability of adopting parallel lines will depend on specific characteristics such as failure rates, setup times and costs, turnover, absenteeism, learning, etc.

When a system with multiple or parallel lines is designed and managed it is necessary to determine the number of lines to be installed, the number of stations in each line and the tasks or operations to be assigned to each station. In a multi-model or mixed-model case, it is also necessary to determine the assignment of models to lines and, in the most complex cases, the sequence of models in each of the flexible lines.

Although there is a considerable body of literature on assembly line balancing problems (particularly the simplest case, known as SALBP), parallel lines have not yet been studied in detail. This is unsurprising, considering that even the simplest case is difficult to solve (i.e., it is an NP-hard problem).

Very few papers deal with the complete, or almost complete, multiple or parallel assembly line balancing problem, which consists in determining the number of parallel lines to be installed, assigning stations/workers to lines and assigning tasks to stations (in the multi-model or mixed-model case it is also necessary to assign models to lines).

Moreover, it should be noted that the majority of papers on multiple or parallel lines were published in the mid- or late 1990s, when hardware and software advances were still insufficient to solve the optimization problems that can now be solved. Therefore, some of the proposed procedures would produce better results or could deal with problems of a realistic size if they were applied now. The lack of metaheuristic procedures for solving this kind of problem should also be borne in mind.

When considering a multiple or parallel line production system, there is such a wide range of possibilities (particularly when the lines are not identical) that, even if it were possible to model the balancing problem using an MILP model (which would be extremely difficult), the probability of solving it for realistic cases within a reasonable period would still be extremely low. The following
lines of research can be proposed from an analysis of the literature and from the work explained in this study.

- Given the difficulty of modelling and solving the problem, it would be appropriate to design an iterative procedure in which a problem consisting in balancing and evaluating a system with a fixed number of lines (whether some of these are identical should be known) is solved at each iteration:

  ▪ One of the best procedures designed for the SALBP or GALBP can be used to balance each line according to its specific characteristics (most of these procedures are referenced by Becker and Scholl, 2006 and Scholl and Becker, 2006).

  ▪ The evaluation of the total cost will depend on the particular characteristics and on the way in which certain stochastic issues (failures, learning, absenteeism, turnover) are considered; the possibilities range from the simplest case, in which only mean values are considered, to the use of simulations or analytical models.

  ▪ Once the problem has been solved for a given line configuration (i.e., the number of lines has been determined and the identical lines have been identified) and the corresponding costs have been evaluated, the number of lines is increased by one and the next iteration is performed. It is not clear whether the cost function (in terms of the number of lines) is convex or not; it is possible that the solution might worsen and then improve. How to determine the ending condition for the procedure should be studied in depth.

- A quite immediate work could be to adapt the procedure of Chiang et al. (2007) to solve the MALB problem (i.e., to consider straight lines instead of U-shaped lines). The procedure proposed by Chiang et al. (2007) to solve the MULB problem with lines of varying size could also be the basis for designing a procedure to solve more general problems. Indeed, they propose to incorporate stochastic processing times and to consider product families with major and minor setup times. Other stochastic issues (such as failures, learning or absenteeism) could also be considered.

- Another research line, following the works done by Miltenburg (1998), Sparling (1998) and Gökçen et al. (2006), would be to consider in the balancing the possibility of having some stations being shared by different assembly lines (i.e., multiline stations).
Acknowledgements

The author would like to thank Dr. Rafael Pastor and Dr. Albert Corominas for their useful comments on this paper. The author is also grateful to the anonymous referees who also provided helpful comments on this paper.

This paper was supported by project DPI2004-03472 of the Spanish Ministry of Science and Technology and was co-financed by the ERDF.

References


