Production Planning of Mixed-Model Assembly Lines: Overview and Extensions

Nils Boysen, Malte Fliedner, Armin Scholl

06/2007
Production Planning of Mixed-Model Assembly Lines: Overview and Extensions

Nils Boysen\textsuperscript{a}, Malte Fliedner\textsuperscript{a}, Armin Scholl\textsuperscript{b,\ast}

\textsuperscript{a}Universität Hamburg, Institut für Industrielles Management, Von-Melle-Park 5, D-20146 Hamburg, \{boysen,fliedner\}@econ.uni-hamburg.de

\textsuperscript{b}Friedrich-Schiller-Universität Jena, Lehrstuhl für Betriebswirtschaftliche Entscheidungsanalyse, Carl-Zeiß-Straße 3, D-07743 Jena, a.scholl@wiwi.uni-jena.de

\ast Corresponding author, phone +49 3641 / 943171.

Abstract
Mixed-model assembly lines are of great practical relevance and are widely used in a range of industries, such as the final assembly of the automotive and electronics industries. Prior research mainly selected and discussed isolated problems rather than considering the whole planning process. In this article mixed-model production planning is decomposed into five steps: initial configuration of the line, master scheduling, reconfiguration planning, sequencing and resequencing. The paper reviews and discusses all relevant planning steps and proposes general planning instruments as well as formalized decision models for those steps, which have not been thoroughly investigated in the literature thus far.

Keywords: Mixed-model assembly lines; Production planning
1 Introduction

A paced assembly line is a flow-oriented production system, which employs some kind of material transportation system (typically a conveyor belt) to transfer workpieces successively to its productive units (stations) at a given rate, such that the total duration over all operations at a station is limited by a cycle time.

Since the times of Henry Ford and the famous Model-T, assembly line production systems have been employed in many industries and are nowadays widely used in the mass production of standardized commodities. As a consequence of the increasing individualization of consumer products in many industries within the recent years, plenty effort has been spent in order to increase the flexibility and versatility of assembly lines, such that the benefits resulting from the high degree of specialization of labor and its associated learning effects (Shhtub and Dar-El, 1989; Scholl, 1999, p. 2) can also be exploited in the assembly of low volume, highly diversifiable products. The use of advanced production technologies, such as universal machinery with automated tool-swaps (Dolgui et al., 2006), nowadays allows the manufacture of different variants of a common base product on the same line in subsequent production cycles without noticeable setup times or costs (lot size of one). These, so called, mixed-model assembly lines (Wild, 1972, p. 46) are widely employed in assembly-to-order production systems (Mather, 1989) and enable modern production strategies such as mass-customization (Pine, 1999). Important practical fields of application can be found in the electric industry (Bolat, 1997) and the final assembly of cars (Boysen et al., 2006b), the latter of which deals with an especially dramatic diversity summing up to $10^{32}$ different car models on the same assembly line (Meyr, 2004).

The high practical relevance of mixed-model assembly is also reflected by the vast amount of academic work in this field. With only few exceptions, the majority of the numerous research papers treats either one of the following two planning problems:

(i) The so called assembly line balancing constitutes a long- to mid-term planning problem, which seeks to group the total amount of assembly operations and assign them along with the required resources to the productive units (stations) of the assembly line (c.f. Baybars, 1986; Becker and Scholl, 2006; Scholl and Becker, 2006; Boysen et al., 2006a).

(ii) The short term sequencing problem of mixed-model assembly lines assigns all jobs of the given production plan (also referred to as model-mix) to the production cycles in the planning horizon (c.f. Wester and Kilbridge, 1964; Thomopoulos, 1967; Tsai, 1995; Scholl et al., 1998; Boysen et al., 2007a).

Although both problems are without doubt essential tasks of mixed-model production planning, they are by far insufficient to guide and control the complexity of mixed-model production systems. In fact, a successive approach which merely considers and solves the aforementioned two problems separately ignores the following important planning aspects:
(i) Nearly all research papers consider the first time installation of the assembly line (Falkenauer, 2005). This case certainly represents a critical decision problem, as the fundamental design of the production system and the associated expensive machinery are long-time investments being not readily reversible. However, the exact layout of the line and especially the task assignment is in practice subject to a continuous control and improvement process and is thus often modified in reaction to a dynamic environment (Scholl, 1999, p. 109; Boysen et al., 2006b). A complete planning approach needs to appropriately integrate this reconfiguration process.

(ii) The long- to mid-term configuration and the short-term sequencing of a mixed-model line are in practice linked by an additional planning step, which has hardly been covered by research so far. Before the actual jobs of a planning period are sequenced, a master scheduling problem needs to be solved, which coordinates all incoming orders and their respective due dates and generates the individual production plans for each planning period (Hindi and Ploszajski, 1994; Bolat, 2003).

(iii) Once the sequence of models is determined and the actual production process has started, disturbances, like last minute orders, machine breakdowns or defective material, can impair the processing of the sequence and necessitate further adjustments. As part of the short-term production control process, a resequencing problem needs to be taken into account (Bolat, 2003).

With regard to the aforementioned aspects, the remainder of the paper will develop and discuss central elements of a conceptual framework for mixed-model production planning and propose planning approaches for selected problems which have not been thoroughly discussed in the literature thus far. For this purpose the paper is structured as follows: Section 2 will give a brief overview of the total planning approach, whose main steps are then described in detail in the subsequent Sections 3 to 5. Thereby open research questions are identified which are summarized in Section 6.

2 Mixed-Model Production Planning

Comprehensive production planning systems for mixed-model assembly lines have hardly been studied by research up to now. However, there are a couple of papers dealing with simultaneous assembly line balancing and sequencing (McMullen and Frazier, 1998; Kim et al. 2000a+b, 2006; Miltenburg, 2002; Sawik, 2002; Bock et al., 2006). The main problem of these approaches certainly stems from the vastly different time frames of the two decisions. At the time of the balancing decision the actual production plans for the individual shifts are typically unknown and can hardly be anticipated in detail. Merely average or expected model or option demand rates can be determined with sufficient accuracy. It seems generally more meaningful to exploit the additional short-term information whenever production sequences are determined and thus to decompose the total decision problem into separate yet coordinated planning steps.

To the best knowledge of the authors the only academic works in this field were carried out by Decker (1993, p. 13), Scholl (1999, p. 106), Meyr (2004) and Boysen (2005,
None of the papers however provides a complete planning approach which covers configuration and reconfiguration planning, master scheduling, sequencing and resequencing as is displayed in Figure 1.1

Although the planning of a production system in principle starts with the product design, as a large deal of future costs are already determined at this phase (Swift et al., 2004), the actual production planning of the assembly system begins with the initial configuration and installation of the productive units. This configuration planning also decides on all resources which are required in the production process. The numerous mathematical models and solution procedures suggested by the academic literature in order to support this planning step are subsumed under the term assembly line balancing. As assembly systems typically require a long life-time before the high investment costs are amortized, a major challenge at this stage is the appropriate quantification of relevant influencing factors, such as future sales, costs or process times. The various uncertainties often make readjustments to the running system inevitable. Furthermore, unforeseeable changes in demand or the availability of new technologies might continuously necessitate a reconfiguration of the line. This is especially true in the case of mixed-model lines, where different models often utilize the available capacity in very different intensities. Whenever the demand structure shifts to another model-mix, the line balance might need to be modified.

In light of this, a tactical master planning, which collects all of the required information and either adjusts capacity demands or available capacities accordingly, becomes even more essential. An integral part of this master planning phase is the solution of a master scheduling problem, which repetitively assigns the overall set of orders to planning periods (shifts) thereby determining the production plans for the periods. If reconfiguration of the assembly line is supposed to be a viable instrument to adjust capacities, then it is also usually in this phase that the reconfiguration problem needs to be considered, as the rebalancing of the line typically takes a considerable amount of time and can regularly not be “spontaneously” decided upon on a “shift-by-shift” basis.

Once the production plans are determined and prior to the actual start of the assembly, a feasible production sequence is to be determined with an appropriate sequencing
approach. As the production plans are determined by master scheduling, it might be further desirable that the master schedule already anticipates the subordinate sequencing objective. This emphasizes the central role of master scheduling and its interface function between the long- and short-term planning problems and underlines the dire need for more academic work in this field.

As unforeseen disturbances can always render the retrieved production sequence infeasible or at least strongly suboptimal, an efficient production planning and control system should consider ways to modify a predetermined sequence even during execution. The resulting resequencing problem has also hardly been discussed thus far, probably because the stochastic nature of its occurrence seems to make universal approaches ineffective. Thus, flexible policies need to be developed which respond to sudden changes appropriately.

The identified planning steps and their respective challenges with regard to mixed-model production are characterized in detail in the following sections. We principally assume that each problem is solved independently in a successive manner, so that the solutions of a higher level problem are passed on to the lower level in form of an instruction, thereby reducing its decision space. Due to the central role of master scheduling we will further propose mechanisms which integrate the objectives of line balancing and sequencing into a comprehensive master scheduling approach. By doing so, the decision steps are also connected in an anticipatory manner to account for aspects of lower-level steps in higher-level ones. This is a focal issue of integrating different decision problems with different horizon and information conditions into an overall planning hierarchy (see Scholl, 1999, pp. 106-112; Boysen, 2005, pp. 141-150).

3 Balancing Mixed-Model Assembly Lines

3.1 General Remarks On Mixed-Model Assembly Line Balancing

The core problem of assembly line balancing is to assign the specified assembly operations along with their required resources to the stations of the assembly line (Boysen et al., 2006a). Typically, various technical or organizational restrictions need to be observed in order to retrieve a feasible sequence, such as that the work content assigned to a station may not exceed the cycle time.

Since the review of Baybars (1986) this core problem is referred to as simple assembly line balancing problem (SALBP). More advanced models and solution procedures for SALBP are provided by Erel und Sarin (1998), Scholl (1999), Rekiek et al. (2002b) as well as Scholl and Becker (2006). In addition to that, another field of research has been established, which deals with the ascertainment, modeling and solution of more practice-relevant balancing problems under the term general assembly line balancing (GALBP), in which numerous extensions such as parallel stations, stochastic processing times, setup times, u-shaped line designs and processing alternatives have been considered. A recent overview on solutions approaches is provided by Becker and Scholl (2006), a comprehensive classification of the various extensions can be found in Boysen et al. (2006a+b).
The traditional ALBP was limited to a single-product assembly system and up to now the vast majority of research papers consider this case. In order to make use of the numerous powerful solution procedures developed under this assumption for mixed-model lines, it is often proposed to unify the precedence graphs of all models to one overall joint graph (Thomopoulos, 1970; Macaskill, 1972; van Zante-de Fokkert and de Kok, 1997; a different approach is employed by Bukchin and Rabinowitch, 2005).

Similar tasks which employ the same resources are represented by a single node in the joint graph, so that an assignment to different stations is prohibited (Thomopoulos, 1967). This unification can however lead to cycles in the precedence graph (the same task is a predecessor and a successor of another task), which need to be removed in order to generate a feasible task sequence (Ahmadi and Wurgaft, 1994). Finally, the joint task times are calculated on the basis of the model-mix by weighting the respective task times with the estimated probability of the model occurrence.

Due to the recent trends of ever-increasing model varieties, reliable model-mix prognoses become more and more difficult. This combinatorial complexity usually makes model-based demand prognoses impossible. Instead only overall option occurrences can be reliably estimated (Röder und Tibken, 2006). This change in information, however, necessitates modifications to the generation of joint precedence graphs. Tasks need to be mapped to options instead of models, so that average task times can be calculated on the basis of expected option occurrences. Due to the specific structure of mixed-model assemble-to-order production systems, the determination of joint task times \( \bar{t}_i \) will then be rather simple for the majority of tasks. Let the total task set be denoted with \( V \) and the set of options which necessitate a task \( i \) to be carried out by \( O_i \), then \( V \) can be split up into three disjoint subsets \( V^A \cup V^B \cup V^C = V \) to which a task \( i \) is assigned with regard to the number and interaction of options in set \( O_i \) as follows:

1. **Common tasks:** The tasks \( i \in V^A \) have to be carried out on any model independent of the options required, i.e. \( O_i = O \). Additionally, their task times \( t_i \) are identical for all option combinations.

Then, the joint task times are simply given by \( \bar{t}_i = t_i \forall i \in V^A \).

2. **Single-option tasks:** The subset \( V^B \) contains all tasks which can be assigned to a single option occurrence exclusively. This includes all tasks which are required by a single zero/one option (\(|O_i| = 1\)), such as the air conditioning which may be present or not. Additionally, all tasks are covered which are required by multiple options (\(|O_i| > 1\)) provided that these options cannot occur in the same model.

In the automobile industry, this is, e.g., the case when sunroofs are installed. The task “mounting of grommet” has to be performed (maybe at different processing times) irrespective of the exact type of sunroof, be it electric or manual. Although this task is thus required by multiple options, it can be assigned exclusively to an option in any possible model, as the options electric and manual sunroof are mutually exclusive and can never occur together in the same car.
For such a single-option task $i$ the joint processing time $\bar{t}_i$ is equal to the weighted average over all option-specific task times $t_{io}$ in proportion to the occurrence probability $p_o$ of the respective option $o$.

$$
\bar{t}_i = \sum_{o \in O_i} p_o \cdot t_{io} \quad \forall i \in V^B
$$

3. **Multiple-option tasks**: The remaining subset $V^C$ includes any task $i$, whose set of options ($|O_i| > 1$) contains at least two options which can occur or not independently of each other. Such tasks are, e.g., inevitable whenever electrical components are installed in the door of a car. If electrical exterior mirrors and/or power windows are chosen, the power supply has to be made accessible by a wiring harness. This installation is thus a shared task, which becomes necessary whenever either one of the door components is chosen separately or if both options occur jointly. Moreover, task times may diverge between any of the possible option combinations. Only for tasks in this set the occurrence probabilities of all possible option combinations need to be estimated along with their respective processing times, so that the joint processing time can be calculated as the weighted average of processing times over all option combinations.

An in-depth discussion on that topic is provided by Boysen et al. (2006c). With these modifications the mixed-model assembly line balancing problem can be appropriately reduced to the single-product case. Further structural aspects of the mathematical model thereby heavily depend on whether the line is initially constructed or whether an existing line is reconfigured.

### 3.2 Initial Configuration

Whenever an assembly system is initially configured, the future productive units are not yet specified. At the time of the planning decision, stations are mere abstract entities to which the assembly operations can be assigned. Typically, the necessary resources are not yet purchased, only the required features of products and models are - more or less - certain. Likewise, the production process is not yet finally determined. At automobile manufacturer Volkswagen for instance, the configuration planning begins about 37 months prior to the actual start of production, so that the tasks of the precedence graph are not yet fully determined.

The production process can in principle be specified in two ways: Either the processing modes of all operations (tasks) are determined upfront, before a line balancing model is applied on the basis of the resulting precedence graph or the selection between processing alternatives is integrated in the model and performed simultaneously with the task assignment. In the latter case, all (or a sufficient subset of) processing modes need(s) to be retrieved together with the associated task times and costs. As the latter case promises a better overall solution, it can be generally recommended whenever an assembly line is constructed, provided that the increased computational complexity can be handled.
As processing alternatives usually vary in investment costs, the most common surrogate objectives for designing efficient assembly lines - like minimizing the cycle time or number of stations - become insufficient. Instead the relevant cost factors need to be estimated and explicitly modeled. As in the early construction phase also demand and sales figures are uncertain it might even be recommendable to directly consider the trade-off between revenues, which are heavily influenced by production quantities and thus the cycle time, and investment costs in a profit oriented objective function (Rosenblatt and Carlson, 1985; Martin, 1994; Boysen and Fliedner, 2006), so that different price and sales scenarios can be used to evaluate the payback period of the system. Due to the high degree of specification of labour learning effects are often of significant importance especially considering the long lifetime of assembly production systems. If they are not properly accounted for, considerable excess capacities threaten especially in the later phases of the product life cycle. Despite the various problems regarding the proper estimation of learning effects, configuration planning should aim at considering its associated influences (Boucher, 1987; Chakravarty, 1988; Cohen et al., 2006; Sotskov et al, 2006).

3.3 Reconfiguration

In practice, the reconfiguration of an existing line is of at least the same importance as the configuration of a new production system (Falkenauer, 2005). A reconfiguration can become necessary whenever the expected model-mix changes substantially, new production technologies are made available or the production rate needs to be adjusted. In an existing line, stations have an identity, which is reflected by already assigned resources and a fixed spatial position. As stations are already constructed, the common objective of station minimization is only meaningful, if the productive units can actually be dissolved at will. To react on changes in demand volume and/or structure, the cycle time (production rate) can be adjusted by reassigning tasks accordingly. However, the production rate (cycle time) is sometimes externally determined on the basis of fixed production orders and/or reliable sales forecasts. As a consequence, both the number of stations and the cycle time may be rather hard restrictions which limit the feasible solution space. Then, it seems more reasonable to equally distribute the work content as an alternative objective (Agnetis et al., 1995; Merengo et al., 1999; Rekiek et al., 2001, 2002a), as smooth station loads promise a reduction of waste due to work overloads.

Especially whenever heavy machinery has already been installed at a station, the tasks that require this machine may be subject to assignment restrictions and can thus not be reassigned (Boysen et al., 2006b). Often the machinery is not completely immobile, but instead subject to moving costs (Gamberini et al., 2006), which need to be considered in the reconfiguration. In this context, also space restrictions are often critical and further impede a free reassignment of tasks and resources (Kilbridge and Wester, 1961; Sawik, 2002; Bautista and Pereira, 2006).

Similar phenomena apply to the operators of a station. Typically, they possess specific qualifications in order to perform their operations, so that a change in work content
causes additional training costs. Line balancing should thus seek to consider the qualifications to avoid unnecessary training or long start-up phases after the reconfiguration. In accordance with the aforementioned considerations and their relevance in the production system on hand, an appropriate line balancing problem is to be formulated and solved. While there are already numerous powerful solution procedures available for the initial configuration phase (see Boysen et al., 2006b), reconfiguration planning has not been considered accordingly and should be studied more intensively in the future.

Once the capacities of the assembly system are determined, they are passed on to the next planning level, so that master scheduling can determine the production plans on this basis.

4 Master Scheduling

Master scheduling decides on the type and amount of product models to be produced in the planning horizon and assigns them to planning periods (e.g. shifts) (Meyr, 2004). In the automobile industry for instance, the total production volume of the forthcoming months is distributed over the available working days and shifts (Monden, 1998; Decker, 1993). Thus far, this planning problem has hardly been considered by research (with the exception of Hindi and Ploszajski, 1994; Bolat, 2003, who only determine the production plan for the nearest production shift). Typically, production orders are associated with due dates, so that the primary objective of master scheduling is to find a feasible assignment that minimizes holding costs and penalties for late delivery, respectively. This requires an observance of the total capacity or total number of production cycles and the availability of material (Decker, 1993, p. 7; Bolat, 1997; Scholl, 1999, p. 111).

4.1 Master Scheduling with given Capacity and Demand

We start with a basic model that assumes the line’s capacity to be fixed. The planning horizon is divided into $T$ ($t = 1, \ldots, T$) periods to which $N$ ($i = 1, \ldots, N$) production orders are to be assigned. Each order has a due date $L_i$, which was, for instance, contractually agreed upon with the customer. If the order is produced early, variable inventory holding costs $l_i$ accumulate per period. Late deliveries cause penalty costs $s_i$, either due to contractual regulations, discounts, loss of goodwill or anticipated lost sales (Delleart, 1989, p. 15). The exact deviation costs $c_{it}$ of an order $i$ assigned to period $t$ can be determined upfront. It is often assumed that costs increase in a linear manner over time (Baker and Scudder, 1990; Lee and Kim, 1998), so that equation (2) can be employed:

$$c_{it} = \begin{cases} l_i \cdot (L_i - t), & \text{if } t \leq L_i \\ s_i \cdot (t - L_i), & \text{if } t > L_i \end{cases} \quad \forall i = 1, \ldots, N; t = 1, \ldots, T$$

The following basic mixed-integer programming model for master scheduling (MS-B) is proposed to minimize deviation costs (3)-(7):
The objective function (3) considers deviation costs for all orders \( i \) assigned to periods \( t \) \((x_{it} = 1\)). Restrictions (4) make sure that each order is assigned to exactly one planning period. The sum of assigned orders may not exceed the number of available production cycles \( P \) for each period (5). For simplicity it is assumed, that each period has the same length \( D \) so that the number of available cycles can be calculated by dividing \( D \) by the constant cycle time \( c \). An 8-hour shift \((D = 8 \cdot 3600 = 28,800 \text{ seconds})\) and a cycle time of \( c = 60 \text{ seconds} \) would thus result to \( P = 480 \) production cycles. Finally, restrictions (6) ensure that the available amounts of material are not exceeded. Material coefficients \( b_{iq} \) represent the amount of a certain material \( q \) an order \( i \) requires out of the set of materials \( Q \). The maximum available quantity of a material of type \( q \) in period \( t \) is often predetermined in master agreements with the suppliers and given by \( B_{qt} \).

Master scheduling should be carried out repeatedly on rolling horizons (e.g. Kimms, 1998), so that only the production dates of orders assigned to the first periods are truly fixed, while the remaining orders can be reassigned at later points in time to account for changes in the available information base. The actual number of fixed periods mainly depends on the lead time of preceding production levels, the outputs of which are required for the final assembly.

4.2 Master Scheduling with Demand and Capacity Adjustments

The basic model MS-B merely allows a time-based adjustment of orders and due-dates. In principle, there are, however, several further alternative reactions, which can be integrated in the planning approach and either affect capacity demand or available capacities.

**Demand Adjustments:** The capacity demand can be, for instance, regulated by rejecting customer orders. This can be easily modeled by adding a “virtual” planning period, to which a sufficiently high number of production cycles \( P \) and a sufficient quantity of material \( B_{qt} \) are assigned. If an order is assigned to this period, the associated costs
should comprise the loss in sales (contribution margin of the order) as well as potential further loss of goodwill caused by the rejection.

In addition to that, due dates can be influenced by entering into a negotiation process with the customer, often referred to as order-promising. By systematically identifying and stipulating due dates with regard to available capacity, bottlenecks in production can be avoided (Capable-to-Promise, e.g. Stadtler, 2005). In the mathematical model only the cost factors of those orders which are not yet fixed on behalf of the customer need to be modified. For those planning periods where a delivery is acceptable, deviation costs $c_{it}$ are set to zero, periods which are not acceptable receive sufficiently high costs, so that an assignment is prohibited. The date $t$ to which the order $i$ is assigned ($x_{it} = 1$) in the optimal solution is, hence, the ideal date that should be promised to the customer.

**Capacity Adjustments:** Further adjustments can be made with regard to the available capacity. In principle, three types of modification can be distinguished:

(i) **Reconfiguration:** The line can be reconfigured in order to adjust the cycle time according to capacity demands. This might require a reassignment of tasks as well as a reallocation of resources (alternative machinery, change in team size) at one or more stations.

(ii) **Time adjustment:** The lengths of the production periods are changed (short-time work or additional shifts).

(iii) **Subcontracting:** Orders can be subcontracted to third-party suppliers.

All three aspects are discussed in the following:

(i) **Reconfiguration:** The basic model assumes a fixed line balance which in turn determines the number of production cycles per period. As was argued in Section 3.3, assembly lines are, however, often reconfigured. In order to account for this, master scheduling could be coupled with assembly line balancing and planned simultaneously. The resulting models could, however, hardly be solved to optimality for problems of real-world size. Instead, it seems more recommendable to determine alternative line balances upfront for relevant cycle times and pass this information to the master scheduling problem. Once a set $A$ of alternative line balances has been determined, master scheduling can additionally decide on the actual line-balance for each planning period, for instance in order to react to demand peaks. For this purpose, setup costs need to be considered, which are incurred by a change from line balance $a$ to line balance $\alpha$ (see Section 3.3) as well as the operating costs of the selected line balance $a$, which may differ due to increased wage cost for additional operators for instance. The modified model MS-R is provided in the following:
(MS-R) Minimize $z_2 = \sum_{i=1}^{N} \sum_{t=1}^{T} c_{it} \cdot x_{it} + \sum_{t=1}^{T} \sum_{a \in A} y_{at} \cdot K_a^B + \sum_{a,a \in A \atop a \neq a} r_{taa} \cdot K_{aaa}$

subject to (4), (6), (7) and

$$\sum_{a \in A} y_{at} = 1 \quad \forall t = 1, ..., T$$

$$\sum_{i=1}^{N} x_{it} \leq \sum_{a \in A} P_a \cdot y_{at} \quad \forall t = 1, ..., T; \text{ with } P_a = \left\lfloor \frac{D}{c_a} \right\rfloor$$

$$r_{taa} \geq y_{a,t-1} + y_{at} - 1 \quad \forall a, \alpha \in A; t = 1, ..., T$$

$$y_{at}, r_{taa} \in \{0, 1\} \quad \forall a, \alpha \in A; t = 1, ..., T$$

Operating ($K_a^B$) and setup costs ($K_{aaa}$) are taken up in the new objective function (8). The binary variables $y_{at}$ and $r_{taa}$ store, which line balance $a$ is selected in period $t$ ($y_{at} = 1$) and whether a line balance $a$ was changed to $\alpha$ ($r_{taa} = 1$) in period $t$, respectively. The additional equations (9) ensure that exactly one balance is chosen per period. This implies, that line balances can realistically be changed only in between two planning periods. Restrictions (10) are modified, so that the available number of production cycles $P_a$ depends on the cycle time $c_a$ of the selected line balance $a$. The further restrictions (11), together with the objective function, control that the setup variable ($r_{taa}$) is set to 1, if the line balance was changed from $a$ to $\alpha$ between two adjacent planning periods $t-1$ and $t$. This formulation requires the initialization of variables $y_{at}$ to reflect the actual line balance at the decision point.

With these modifications, the reconfiguration of the assembly system can be systematically controlled by the master scheduling model. The set of alternatives $A$ is to be extended, whenever the structure of the model-mix changes significantly or new production technologies are to be employed.

(ii) Adjustment of production time: In an assembly system, the length of the production period can mainly be altered by introducing or canceling entire production shifts. This type of adjustment is accounted for in model MS-S:

(MS-S) Minimize $z_3 = \sum_{i=1}^{N} \sum_{t=1}^{T} c_{it} \cdot x_{it} + \sum_{t=1}^{T} z_t \cdot K^S$

subject to (4), (6), (7) and

$$\sum_{i=1}^{N} x_{it} \leq P \cdot z_t \quad \forall t = 1, ..., T; \text{ with } P = \left\lfloor \frac{D}{c} \right\rfloor$$

$$z_t \in \{0, 1, ..., S_{\text{max}}\} \quad \forall t = 1, ..., T$$
A planning period $t$ is subdivided into an arbitrary number of time intervals of equal length (shifts). The integer variables $z_t$ specify the number of shifts operated per period $t$. The objective function (13) considers the costs $K^S$ which are incurred by any additional shift. The number of production cycles per period is calculated by the number of production cycles per shift $P$ and the number of allotted shifts $z_t$ (14). Restrictions (15) limit the number of shifts per period by a pre-specified maximum value $S_{max}^2$.

Notice that the model also may consider an adjustment based on enlarging or shortening fixed production shifts by defining the shift lengths appropriately. If it is in a period $t$, e.g., possible to operate a single shift with a duration of 6, 7, 8, 9, or 10 hours, the value set of the variable $z_t$ has to be restricted to $\{6, ..., 10\}$.

(iii) Subcontracting of production orders can again be considered via a “virtual” period (Lee and Kim, 1998). The assignment costs of these periods are set according to the price of the subcontractor. This type of adjustment is, however, assumed to be of minor importance in mixed-model production environments, such as the automobile industry.

Obviously, the measures described above separately can be combined arbitrarily to define a flexible assembly system which is able to react on demand changes in the most efficient manner.

Judging on the complexity of the proposed models, we find that already the basic model is NP-hard, as it is a special version of the generalized assignment problem (see Martello and Toth, 1990, p. 190). As a consequence, powerful heuristic solution procedures need to be developed to further practical applications. This is especially necessary if an order-promising is to be implemented as this might require a repetitive real-time solution.

Once the production plans are determined, they are forwarded to the part suppliers to coordinate their production and distribution planning (Decker, 1993, p. 7) and further passed on to the next planning level, where each order is assigned to exactly one production cycle.

5 Sequencing of Mixed-Model Lines

5.1 Initial Production Sequence

Although mixed-model assembly systems technically allow facultative production sequences without incurring setup times or costs, the actual sequence does very well influence the demand for production capacities and required material. As a consequence, the determination of production sequences becomes an important planning problem in this context. The problem consists of finding an assignment of each order of the production plan to a production cycle. In total, two different objectives have been established in the literature, which are observed by three classes of optimization approaches in a different manner (Boysen et al., 2007a):

(i) Modern mixed-model assembly systems are often supplied just-in-time or even just-in-sequence (Meyr, 2004) either by preceding in-house production levels and/or
third-party logistics providers. As the just-in-time concept relies on a minimization of inventory, the sequence of production orders directly determines the sequence of part consumption. In order to avoid demand peaks at preceding production levels, an even spread of material demand over time is seen as an essential prerequisite. Sequencing approaches which aim at smoothing material demand rates are referred to as level scheduling and were inspired by the famous Toyota production system (Monden, 1998). Various different model formulations and solution approaches are provided by Miltenburg (1989), Inman and Bulfin (1991, 1992), and McMullen (1998, 2001), overviews can be found in Kubiak (1993) and Boysen et al. (2006d).

(ii) The other body of literature does not focus on the demand rates of material, but rather on the avoidance of capacity overloads at the productive units. As different product variants consist of different optional parts, they also require different operation times at the stations. If several work-intensive models follow each other at the same station, an overload can occur, which needs to be compensated by some costly reaction, such as the employment of utility workers or line stoppage (Wild, 1972, p. 164). The so called mixed-model sequencing minimizes these overloads by an exact job scheduling considering processing times, station lengths and line movement (Wester and Kilbridge, 1964; Thomopoulos, 1967; Bard et al., 1992, 1994; Tsai, 1995; Scholl et al., 1998; Boysen et al., 2007a).

(iii) Car sequencing, the other capacity-oriented approach, does not rely on a detailed time schedule, but employs simple sequencing rules of the form \( H_0 : N_0 \) in order to control the production sequence. A sequencing rule of 1:3 for the option “sunroof”, for instance, enforces that only one job out of three can be equipped with a sunroof if a capacity overload is to be prevented. Solution procedures for the car sequencing problem can be found in Drezl and Kimms (2001), Gravel et al. (2005), Drezl et al. (2006), Gagné et al. (2005) and Fliedner and Boysen (2006).

So far, there are hardly any recommendations in the literature which suggest a better fit of one of the models given a certain situation. The choice between capacity- and material-oriented approaches should in principle be based on the incurred costs. As, however, especially the cost effects of a more or less even material demand are difficult to quantify, the actual model employed seems rather based on an overall manufacturing philosophy than on real costs. While Asian car manufacturers, such as Toyota and Hyundai seem to favor material-oriented approaches (Monden, 1998; Duplaga et al., 1996), western manufacturers like Renault employ capacity-oriented approaches (Gagné et al., 2005). A reasonable compromise can be reached by making use of combined approaches which consider both objectives simultaneously (Bard et al., 1994; Drezl and Kimms, 2001a+b; Kotani et al., 2004; Drezl et al., 2006). The necessity of completely leveling demand rates as promoted by level scheduling nevertheless becomes doubtful whenever parts need to be transported and issued in discrete amounts at discrete points in time and is more reasonably replaced by a lot-wise observance (Boysen et al., 2007b).

Once the production sequence is determined it is forwarded to the material suppliers to trigger the delivery of required parts and allow a sorting of just-in-sequence material.
Finally, the physical production process is started.

5.2 Integrating Master Scheduling and Sequencing

As elaborated above, the master scheduling and sequencing problem are highly interdependent. The master schedule determines the production plan per period and thereby directly determines which production orders are to be sequenced. The quality of sequencing solutions heavily depends on the selected production orders, so that the results of the sequencing problem can likewise serve as a performance measure to evaluate the superior master schedule. As a consequence, it seems desirable to anticipate the subordinate sequencing objectives within master scheduling appropriately. The exact form of anticipation thereby depends on the actual sequencing approach considered:

(i) If **mixed-model sequencing** is employed, the master scheduling model MS-B can be extended by condition (16), in order to ensure that at each station \( k \) out of the set \( M \) of stations and in each period \( t \) the sum over all operation times \( t_{ik} \) may not exceed the estimated available capacity \( D \). If \( \alpha = 1 \) is chosen, it is made sure that the cycle time is at least on average observed over all models (which represents the cycle time restriction for joint precedence graphs).

\[
\sum_{i=1}^{N} t_{ik} \cdot x_{it} \leq \alpha \cdot D \quad \forall \, k \in M; \, t = 1, \ldots, T \quad \text{with} \quad \alpha > 0 \tag{16}
\]

The smaller value is chose for \( \alpha \), the stronger the objectives of mixed-model sequencing are anticipated and the more capacity demands in each period are reduced. In times of demand peaks (for instance due to sales promotions) it might be required to set \( \alpha > 1 \), which leads to unavoidable capacity violations which cannot be fully compensated by sequence planning, but rather require a temporary increase in available capacity (see Section 4).

(ii) If the **car sequencing** approach is chosen, master scheduling can anticipate the sequencing objective by controlling the demand for option occurrences. The more occurrences of an option \( o \) need to be assembled in the planning period, the higher is the risk of a \( H_o : N_o \) rule violation. The basic model MS-B can for instance be extended by the following restriction (17), which limits the number of option occurrences per planning period:

\[
\sum_{i=1}^{N} d_{io} \cdot x_{it} \leq \alpha \cdot \frac{H_o}{N_o} \cdot P \quad \forall \, o \in O; \, t = 1, \ldots, T \tag{17}
\]

Demand coefficients \( d_{io} \) denote whether order \( i \) requires option \( o \) \((d_{io} = 1)\) or not \((d_{io} = 0)\). The assigned number of option occurrences may not exceed the number of available production cycles \((P = \lfloor \frac{D}{c} \rfloor)\) with regard to the sequencing rule. The weight \( \alpha \) can once again be used to control the extent of the anticipation.
(iii) **Level scheduling** seeks to evenly distribute a given material demand over time. This can, however, always be realized irrespective of the actual model-mix determined by master scheduling (except from those unavoidable irregularities due to the division of integer demand values and the sequence length). As a consequence it does not seem meaningful to attempt an anticipation of the subsequent sequencing objective in such a setting. However, the underlying just-in-time philosophy can be considered by likewise smoothing the demand rates over all periods of the planning horizon. In fact, in a variety of settings this seems even more essential than the leveling within shifts (see Boysen et al., 2007a):

\[
\sum_{i=1}^{N} b_{iq} \cdot x_{it} \leq \alpha \cdot \frac{\sum_{i=1}^{N} b_{iq}}{P} \quad \forall q \in Q; t = 1, ..., T
\]  

Restrictions (18) limit the demand of a part \( q \in Q \) in a period \( t \) to the ratio between the total demand for this part and the number of planning periods \( \frac{N}{P} \) (with \( P = \lfloor D_c \rfloor \)) required to process all orders. The extent of allowed deviations is controlled by weight \( \alpha \geq 1 \).

Any of the three extensions requires a determination of the weight \( \alpha \), which controls the extent to which the subordinate sequencing objectives are considered within master scheduling. The determination should, in principle, be based on the trade-off between estimated costs of due date deviation in master scheduling and those of capacity violations and material demand variations, respectively, in the sequence planning. Especially the costs of potential capacity violations and demand variations are difficult to anticipate, so that further studies on this trade-off are required.

### 5.3 Resequencing

During the course of production, unforeseen disturbances, such as machine breakdowns, defective jobs or material or last-minute orders might necessitate modifications to the original sequence. These short-term alterations are referred to as resequencing (Bolat, 2003). Thus far, only few academic works explicitly treat this problem. This might be due to the fact that the degrees of freedom for modifications are typically small and often determined by technical aspects of the assembly system, so that it becomes more difficult to gain general insights. Especially in mixed-model production a change in sequence is subject to a large variety of additional constraints, as, for instance, staff planning and material requirements planning were carried out on the basis of the original sequence.

Whenever the product variety is high, even the effects of a particular disturbance might be difficult to anticipate, so that highly specified reaction plans become impractical. In automobile assembly, a missing option or module, which could not be installed due to some kind of disturbance, can very well impair subsequent operations on the same workpiece considerably. However, as the number of tasks to be carried out and the number of possible option combinations is large, it would be impossible to anticipate all of these relationships upfront and instruct the workers accordingly. It is thus common
practice to further process the defective or incomplete workpiece as customary and later send it to a special off-line repair station, where specialists exchange defective parts or add missing ones. Especially in this or similar scenarios, a timely resequencing could effectively prevent a costly off-line repair. In this context, notice that up to half of the productive staff is employed in off-line repair areas at major German car manufacturers.

Among the few identified approaches, in principle, two types of resequencing can be distinguished: physical and “virtual” resequencing.

(i) In order to allow a physical resequencing some sort of resequencing buffer is required, which usually takes the form of horizontally or vertically aligned parallel line segments and enable to feed the production orders to the successive line segment in a different sequence than they were received. Figure 2 illustrates the layout of a resequencing buffer.

The problem of appropriately dimensioning a sorting storage to allow a restoration of a modified sequence is treated by Inman (2003), while Ding and Sun (2004) study the effectiveness of different sorting strategies.

(ii) “Virtual” resequencing leaves the physical sequence of jobs unaffected, but instead changes the assignment between workpiece and production order. On a mixed-model assembly line, the different variants are successively specified by the installation of different optional features at the productive units. As a consequence, the assignment between workpiece and order can be altered even after the physical production has begun (Inman und Schmeling, 2003). Assumed that a defective machine impairs the installation of a specific option to order A, this order can be swapped with another not yet processed order B which does not require this specific option, provided that it also requires exactly those product options which have already been installed to A as is displayed in Figure 3.

In the example of Figure 3, machine 2 breaks down at the beginning of production cycle 3. As a consequence option 2 cannot be installed to order 2. However, there exists an order 4 which neither needs option 1 nor 2 and can thus take the place of order 2. If the machine can now be repaired before option 2 occurs again, the production can continue without an observed disturbance. Defective or missing material can in principle be compensated in the same way.

In general, it can be stated that “virtual” resequencing is associated with less investment cost as only a sufficient shop floor control system needs to be provided, which is able to
Figure 3: “Virtual” Resequencing by order swapping

<table>
<thead>
<tr>
<th>Planning Task</th>
<th>Lead Time</th>
<th>Time Frame</th>
<th>Fixing</th>
<th>Resources</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>2-4 years</td>
<td>-</td>
<td>until reconfiguration</td>
<td>stations</td>
<td>joint graph</td>
</tr>
<tr>
<td>Configuration</td>
<td>1-4 weeks</td>
<td>4-8 weeks</td>
<td>shifts/ days</td>
<td>assembly line</td>
<td>orders</td>
</tr>
<tr>
<td>Master Scheduling</td>
<td>structural changes in extent of model-mix reconfiguration</td>
<td>depends on until next reconfiguration</td>
<td>based on recent model-mix</td>
<td>stations</td>
<td>joint graph</td>
</tr>
<tr>
<td>Reconfiguration</td>
<td>1-3 days</td>
<td>1 day/ cycle</td>
<td>planning horizon</td>
<td>stations</td>
<td>models</td>
</tr>
<tr>
<td>Sequencing</td>
<td>adhoc</td>
<td>remaining cycles</td>
<td>planning horizon</td>
<td>stations</td>
<td>orders</td>
</tr>
</tbody>
</table>

Table 1: Aggregation in Hierarchical Mixed-Model Production Planning

perform a reassignment in time and communicate it to the operators. However, “virtual” resequencing tends to be the less effective the later disturbances occur in the production process, as the increased specification of workpieces makes the identification of suitable swaps difficult. Physical resequencing on the contrary can be employed independent of its position in the process, but requires the installation of additional line segments and handling equipments.

Both types of resequencing become more problematic, whenever material is supplied just-in-time, as a change in production sequence triggers additional material movements. Hence, the advantages of a short-term alteration need to be weighed against the costs of a modification. In any way, it can be concluded that this problem still offers a range of challenging questions for future research.

6 Conclusion

In this work a successive production planning process as it is typical for most mixed-model assembly systems is established and discussed. In addition to that, central master scheduling models are introduced, which are suited to serve as an interface between the
long- to mid-term assembly line balancing and the short-term sequencing problem, and
general resequencing techniques are proposed to respond to unforeseen disturbances in
the course of production.

The identified planning steps are eventually summarized in Table 1 together with
their typical lead times, i.e., the time spans for which planning precedes realization, and
planning horizons. The columns Periods, Resources and Products denote the respective
aggregation levels at which the relevant influencing factors are usually comprised, while Fixing
indicates which parts of the actual solution is considered fixed in a rolling horizons
approach.

We see a wide range of further research challenges related to mixed-model production
planning. First and foremost the discussion revealed that research activities are espe-
cially required for the individual problems of reconfiguration planning, master scheduling
and resequencing. Very rewarding, but at the same time certainly the more challenging
from a conceptional point of view, might however turn out further investigations on
the appropriate coordination between the individual planning steps. As simultaneous
multi-criteria approaches seem to be only applicable under very special conditions, innova-
tive anticipation and feedback techniques which coordinate the decomposed planning
problems should lead to the development of more efficient and flexible planning systems
applicable to a vast range of practical settings.

Notes

1 Related planning problems, such as maintenance planning of machines, calculation of safety stocks
for just-in-time delivery and staff deployment which can also be subsumed under production planning
in a wider sense are not in the scope of this paper.

2 To keep the model simple it is assumed that deviation costs $c_{it}$ do not depend on the shift in which
a model $i$ is produced and that any additional shift is equal in length and incurs constant costs of $K^5$.
The model can however easily extended to cover the more general cases, so that the extensions are not
provided.

3 The weight $\alpha$ thereby determines whether the resulting car sequencing problem can be solved without
rule violations. Lower bound computations can be employed in order to set $\alpha$ appropriately (see Fliedner
and Boysen, 2006).

4 Note that resequencing can also be employed to decouple multiple production levels. In the auto-
mobile industry, resequencing buffers are, for instance, employed in between the paint shop and the final
assembly, so that the sequence fed to the paint shop can be modified before the final assembly starts.
By means of this, the total sequencing problem is decomposed into a paint shop sequencing problem,
where batches of the same color are sought (Lustig and Puget, 2001; Speckermann et al., 2004), and an
assembly sequencing problem, which focuses on capacity violations (Inman und Schmeling, 2003).

19
References


