1992

Heuristic methods for the mixed-model assembly line balancing and sequencing

Carlos Joäao Lâeger de Lima Fernandes
Lehigh University

AUTHOR:

Fernandes, Carlos Joao

TITLE:

Heuristic Methods for the Mixed-Model Assembly Line Balancing and Sequencing

DATE: October 11, 1992

HEURISTIC METHODS FOR THE MIXED-MODEL ASSEMBLY LINE BALANCING AND SEQUENCING

by

Carlos João Léger de Lima Fernandes

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Manufacturing Systems Engineering

Lehigh University

October, 1992

CERTIFICATE OF APPROVAL

This thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science.

Sep 29/92
Date

Dr. Mikell P. Groover
Thesis Advisor

Dr. Keith Gardiner
Director, MSE program

Dr. David B. Williams
Chairman, Materials Science & Engineering Dept.

Acknowledgments

I would like to thank my advisor, Professor Mikell P. Groover for his patience and insightful remarks. It was a privilege to have had the opportunity of working with such a valuable man. I wish to thank JNICT (Junta Nacional de Investigação Cientifica e Tecnológica) for making it financially possible for me to enroll in the Master of Science program in Manufacturing Systems Engineering at Lehigh. Also, I can not forget Professors Paulo T. de Castro and Rui Guimarães from the Universidade do Porto, Portugal, who proposed for me to come to Lehigh, and the role played by FLAD (Fundação Luso-Americana para o Desenvolvimento) as well as the American responsibles for the interchange program between Portuguese universities and Lehigh.

Finally, I would like to thank specially my family and my friends for their encouragement, support and help through the good and less good moments of my stay at Lehigh.

Table of Contents

List of Tables

Table | Page

List of Figures

List of Symbols

Symbol	Definition
d	= balance delay
DM(i,l)	= displacement of operator i when starting work on the l^{th} unit
DMAX(i)	= furthest displacement of operator i in the downstream direction
DMIN(i)	= furthest displacement of operator i in the upstream direction
DP(i,l)	= displacement of operator i when completing work on the l^{th} unit
G	= number of labor groups
g	= subscript for labor groups, g = 1, 2,..., G
i	= subscript for workstations, i = 1, 2,..., n
j	= subscript for model, j = 1, 2,..., m
K_j	= number of work elements required to assemble one unit of model j
k	= subscript for work element k, k = 1, 2,..., K
L	= line length (meters)
l	= launch sequence identification, l = 1, 2,..., Q
L_i	= length of workstation i (meters)
L_w	= upstream walking distance (meters)
$L(d)_i$	= downstream allowance distance (meters)
$L(u)_i$	= upstream allowance distance (meters)
m	= number of models
m_{s*}	= number of models in the model set s^*

MPS	= Minimum Part Set
MPS_j	= Minimum Part Set for model j
n	= number of workstations on the line
n^*	= theoretical number of workstations
n_g	= number of workstations in labor group g
nT	= number of minutes available on the line during time T
$O_{i,i+1}$	= overlap between stations i and i+1
p_k	= set of all work elements preceding element k
Q	= total number of units to be produced
Q_j	= quantity of model j to be produced
S	= set of models remaining to be assigned
S_j	= set of units of model j remaining to be assigned
S_p	= fixed center-to-center distance between successive units (meters)
S_{s*}	= similarity index for all elements in set s^* ($S_{s*} \in [0,1]$)
s^*	= model set
s_{s*k}	= measure of the utilization for element k over all models in s^* ($s_s{}^*{}_k \in [0,1]$)
T	= minutes during the shift or other period to be scheduled
$T_{e_{jk}}$	= work element time for element k for model j
T_{il}	= service time at station i, for the unit launched at position l (minutes)
T_{max}	= maximum station time (on a single-model assembly line)
T_r	= repositioning time (min)

T_{si} = time required to produce one unit at station i (on a single-model line)

$T_{s_{ij}}$ = service time for model j at station i (minutes)

$\overline{T}_{si}$ = average time per unit at station i

T_{wc_j} = total work content time for model j

TT_H = maximum desired station time

TT_k = total time per shift to perform work element k on all units

TT_L = minimum desired station time

TT_{max} = maximum station time (per scheduled period)

TT_{si} = station time = total time required to assemble Q units at station i

TT_{wc} = total work content = sum of the work time done during period T

TT_{wc_j} = total time required to assemble all units of model j

$\overline{TT}_k$ = average time to perform task k in all Q units

U_{s*} = similarity index for all elements in set s^*

u_{s*k} = measure of the utilization for element k over all models in s^*

V_c = conveyor speed (m/min)

V_o = operator upstream walking speed (m/min)

w = total number of workers in the assembly line

w_i = number of workers at station i

w_k = weight of work element k

γ = fixed rate launch interval (minutes)

ϵ = balance efficiency

σ_k^2 = variance of work element k time

τ_i = station passage time (min)

$\tau(d)_i$ = time required to walk the downstream allowance distance (min)

$\tau(u)_i$ = time required to walk the upstream allowance distance (min)

Abstract

Two major problems are encountered in planning and operating a mixed-model production line: line balancing and sequencing of products into the line. Line Balancing involves allocation of work elements to workstations on the line in such a way that all stations have an equal amount of work to perform. This would result in smoother production and reduced workstation idle time. Once the line balance is done, an adequate model sequence must be determined. The ideal sequence of products into the line is such that the idle time resulting from an imperfect balance will not increase. Part of this thesis was a review of the previous work done in mixed-model assembly lines. Up to now, it seems that heuristic methods to solve the line balancing and sequencing for mixed-model lines are still the best option. Even small problems result in such a great number of constraints that it is not practical (sometimes even impossible) to solve them through the use of optimum algorithms. Also, several factors contribute to render the solutions obtained by optimum algorithms less than the actual optimum. Such factors include the variability in work element times. Heuristic methods for the line balancing of single-model assembly lines can be extended to mixed-model lines through a slight adaptation. One of these methods, the Largest Set Rule, is extended in this research in order to deal with the mixed-model line case. Extensions to deal with work element time variability are also suggested.

A heuristic method for the sequencing problem is proposed. This method, given the line balancing solution, attempts to have each workstation occupied an amount of time

that is the closest possible to the station average time. The method proved to yield very acceptable results, and because of its simplicity and easy implementation in a computer, appears to be a good option for solving the sequencing problem. This thesis focused on assembly (production) lines of the moving conveyor, products fixed type. For the sequence determined through the proposed method, there will be an optimal line length (and workstation lengths) such that the line will be free of certain inefficiencies such as work congestion and work deficiency. The proposed method produces fairly good results for lines in which the station length is not imposed, or if this is not the case, if the imposed limits are not too different from the optimal station limits.

A comparison study with several other methods (Time Spread and Kilbridge and Wester methods) revealed that the proposed method performed repeatedly better. The comparison study also indicated that a bowl allocation of workloads (heavier loads at stations in either end and lighter in the middle) seems to improve the throughput slightly in relation to balanced lines. When possible a line composed of open stations (the operator is free to cross the station boundaries) should be used. This results in shorter lines and smaller throughput times. Concurrent work (two adjacent operators working on the same unit) also tends to reduce throughput time. For open stations the best launching interval appears to be variable launching rate, and for closed stations fixed launching rate is best.

1. Introduction

1.1. Brief Historical Perspective of Assembly Lines.

The process of bringing together two or more component parts in order to form a new entity is known as assembly. In an assembly process parts are successively added to an assembly (or sub-assembly) until the finished product is completed. A product being assembled is often designated as a job. Every product manufactured that is composed of more than one component will require assembly operations. The assembly operation can be completely automated (e.g. packet of matches), or if the product is small with few components or if the required quantities of product are very small the total of assembly operations is likely to be executed at individual workstations. The same happens for large products such as aircrafts, ships, etc., where the product is fixed in a location and workers will move from product to product performing the work that has to be done (Dar-El 1986). The most common assembly line is the flow-line where the product to be assembled moves successively from one workstation to the next down the line, having work being done in each workstation.

Henry Ford was the principal contributor to the assembly process for large quantities of products (mass production). The Ford model T automobile was the first product to be mass produced. Ford realized that if the assembly process, which had traditionally been performed by individual operators, was broken into individual tasks, distributed over separate operators working at assembly stations spaced along the line, the total assembly time could be reduced and the quality of the product would increase. By dividing the work among several operators, each operator would work on a set of tasks of limited content, rather than having to perform all tasks. After a learning period, the operator would become specialized in the specific set of tasks, and better work quality and an increase in the speed of work would result.

Ford applied certain operating principles to his production lines that resulted in a great improvement in efficiency of the work and gave birth to flow-line technology. Placing tools and workers in the sequence of operations so that each part had to travel the least distance, using a material handling system that took the parts from one workstation to the next in the sequence, and launching the parts into the line at spaced intervals resulted in increased production rate, improvement in quality, and reduction in production cost.

The first assembly line that used these principles was the assembly of a flywheel magneto, at the Highland Park Plant, in 1913 (Boothroyd et al. 1982). The improvements obtained in the production of the flywheel magneto persuaded Ford to include this type of assembly process in the production of the Ford Model T automobile. It was not practical to have the assembly process and the production process together in the same line, and a separate line, specially dedicated to the assembly of the automobiles, was created. These principles of assembly have been carried over until today.

Today, as in Ford's time, the objective in assembly is to achieve high quality and low production cost. In many industries, there have been attempts to replace human operators by automatic assembly processes. Human operators are kept for the tasks which are impractical or uneconomical to automate. However, it is realized that the automation of assembly processes results in less flexibility in the production system, due to the fact that the automated equipment is usually special purpose, very expensive, and may need relatively long setup times (Dar-El 1986), whereas human operators are more flexible. A worker can easily change the nature of work he has been doing. After a period of learning, the operator is prepared to perform the operations efficiently. A machine does not have the built-in flexibility that allows frequent production changes.

Table 1.1 shows that in 1967 in the United States, the percentage of the total labor force involved in the assembly process varies from 20% to 60%. The assembly costs are often more than 50% of the total manufacturing costs. Since 1967, many assembly operations have been automated or partially automated, so these labor percentages are probably somewhat lower today in most industries.

These numbers reveal that the assembly line is a very important aspect of manufacturing. In today's world, where the competition is severe, if competitiveness is to be maintained, it will be necessary to have efficient assembly lines. An efficient assembly line is likely to result in savings in the assembly costs, thus allowing these savings to be used for other purposes.

During the line design process, the designer should try to provide the line with enough flexibility to cope with possible future changes at a minimal cost. When dealing with a new product it is very common that changes in product design, manufacturing processes, tooling, fixtures, and work methods may be necessary. If the line is not designed with the appropriate flexibility, the process of changing the line to adapt to new situations may be very time consuming and costly.

Table 1.1: Percentage of Production Workers Involved in Assembly

Industry	% of workers involved in assembly
Motor Vehicles	45.6
Aircraft	25.6
Telephone and Telegraph	58.9
Farm Machinery	20.1
Household refrigerators and Freezers	32.0
Typewriters	35.9
Household Cooking Equipment	38.1
Motorcycles, Bicycles, and parts	26.3

Source: Boothroyd et al. (1982) - data from 1967.

Currently it is observed that customer-demanded changes are frequent and that there is the need for several different products or for varieties of the same product. The classical flow-line committed to the assembly of a single product (or nearly identical products), producing mass quantities of it, has given way to single flow-line that produces a variety of different products. An example of the great variety of different products is found in Monden (1983). In a Toyota Motor Corp. factory there were several final assembly lines - the Corona line, the Crown line, etc. Each of these final assembly lines produced a great number of different (but related) models. For example, at the

Corona line 3,000 to 4,000 kinds of Coronas were assembled. The differentiation between these models lies in the different combination of engines, transmissions, accelerators, number of doors, color, tires, etc. It would be totally impractical to have a single assembly line producing each different type of Corona model, and therefore the different models are assembled on the same Corona line.

1.2. Classification of Assembly Lines.

According to the number of different models to be produced on the same line, assembly lines can be divided into three categories: (1) single-model assembly lines, in which only one type of model is produced, (2) batch-model assembly lines, where two or more models are produced in batches, and (3) mixed-model assembly lines, where two or more models are produced simultaneously.

The single-model line is used when the demand for a specific product is high enough to justify the dedication of an assembly line to the production of a single product. The batch-model assembly line is used when two or more models are to be produced. Each model is produced in batches, and therefore the line is committed to the production of one model at a time. Batch-model assembly is more flexible than the previous type of assembly line because a greater variety of products is possible. When a batch of a type of model is being produced the line is basically functioning as single-model line. The models produced are kept in finished goods inventory (Prenting and Thomopoulos 1974). The batch size and sequencing of batches into the line is usually done by criteria such as economical order quantity, minimization of changeover costs, etc. If the batch size is very small, this type of line approaches the case of the mixed-model line (a mixed-model line can be viewed as being a batch-model line where the batch size is one). If the batch size is large, the batch-model line approaches the case of a single model-line (Groover 1987).

1.3. The Mixed-Model Assembly Line (MMAL).

Our attention in the present study will focus on the mixed-model assembly line. As in the batch-model case, this line is used when two or more models are to be assembled. However, contrasting with the batch-model line, in a mixed-model line various models are being produced simultaneously. Models can be intermixed in any arbitrary order without the need for setting up the line, making this kind of assembly system the most flexible of the three. In this type of line, as in the previous one, the difference in the models is such that it is practical to have one single line dedicated to their production. Also contrasting with the batch-model line, in a mixed-model line, line changeovers are not needed, or if they are, they are not an aspect of major concern. A single-model line is a particular case of the MMAL were there is only one model to be produced.

The mixed-model assembly line is the most flexible assembly line of the three because of its capability to produce in any order any type of model. It is possible to achieve a continuous flow of each model, and the finished good inventories are kept low.

This peculiarity enables the use of such kind of systems in a Just-In-Time environment, where the necessary products are produced in the necessary quantities at the necessary time.

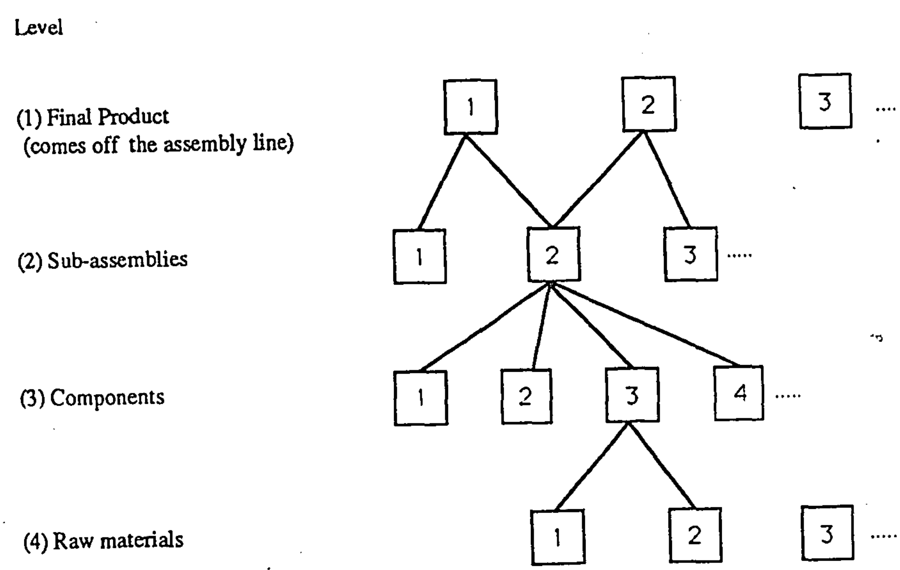
The mixed-model assembly line can be viewed as being a part of a Flexible Manufacturing System, because "speed, quality and production rates similar to mass-production systems can be achieved together with the possibility of producing a diversified number of different models, without the need for line changeovers, keeping low finished goods inventories and producing a continuous flow of products" (Bard et al. 1992).

There are two major problems in mixed-model assembly lines: (1) line balancing, which is to assign work elements to stations in such a way that the station workloads will be evenly distributed, and (2) sequencing of products, which is the order that products are fed onto the line. The goals in solving these problems are to achieve a uniform rate of production for each product (having a continuous flow of each model is one of the main objectives of a mixed-model assembly line) and to smooth (equalize) the workloads among the stations.

The concept of mixed-model assembly lines (MMAL) arises when in a single flow line several different products within a family are assembled. This enables the line to meet the diversified demand of the customers, keeping low finished goods inventories. Small lot sizes and the ability to quickly reconfigure the line are the common norm in most high-tech industries (Bard et al. 1992), and this may be achieved with the MMAL.

Often the assembly line is the final stage of a larger production system. It is designated as the top level in a multi-level production system (Fig. 1.1); each level requirement will trigger production in the preceding level (Miltenburg and Sinnamon 1989). The scheduling of the assembly line will determine the production schedule at the preceding levels.

Figure 1.1: Example of a Product Structure.



To control the assembly line is essential if an appropriate functioning of the other production levels is to be attained. In a multi-level production system, incorrect control of the assembly line may negatively affect the entire plant (Okamura and Yamashina, 1979). Several factors influence the design and operation of an assembly line. As indicated, the two major problems in mixed-model assembly lines are line balancing and sequencing of models into the line.

Suppose that it is possible for the workers at each workstation to receive the unit in which they are going to accomplish assembly work as soon as they are ready for it. In such a situation the entire line would be working only the necessary time to accomplish all operations on all units (the total work content) and the required output would be completed in the minimum possible time. However, a worker may not receive a job immediately when he is ready to work on it, because that job may still be in the previous workstation. If each workstation had the same operation time for each different unit it would be possible for the operators to be working continuously (without stop) and each worker would have the same total work load. In a mixed-model line, because different models will most probably have different times at each station, there is the possibility for workers not to be occupied all the time and for different workloads to exist at different stations. If this happens, the result will be an unevenness in the interval between products coming off the line, an increase in the throughput time and the production output may not be achieved during the shift time. Steps to minimize this are taken. These are line balancing and, once the balance is done, sequencing the products.

Line balancing consists of assigning work elements to workstations on the line. It is desirable to have station workloads distributed evenly. This will enable idle time to be minimized and throughput to be maximized.

Sequencing of models involves the determination of the order in which different models are launched onto the line. When there is only one model, the order in which units are launched onto the line is obviously not important. However, if the number of models is greater than one, launching order becomes important. If the sequence of models is not appropriately determined, idle time increases (even if the line balance was done properly) or units will leave the workstations uncompleted. For example, the launch of successive units of a model with very high workload could result in forcing workers out of their stations to such an extent that they would not be able to catch up and would be constantly out of their stations. In such a situation it is very possible that they would not be able to complete some of the units, and the line would produce incomplete items. Obviously this is undesirable.

The usual procedure is to balance the line first and, given the line balancing solution, the products are sequenced.

Line balancing and sequencing are not only aspects of concern in the design phase, but also during operation. In cases where the line is already designed and the product-mix and/or the production output changes it may be necessary to rebalance the line and to determine again the sequence of products.

Other aspects of the line are determinant at the design phase. Such aspects are the physical configuration of the workstations (length, station boundaries -closed, open, hybrid), use of a paced line versus an unpaced line, conveyor speed, space between units, assignment of models to lines, use of buffer stocks, products removable from the conveyor or fixed to the conveyor, etc. Most of these aspects cannot be considered independently of the others. In the majority of cases an overall design philosophy is required because the most efficient line is not necessarily the result of the best design of each component part of the line. Each part interacts with others and therefore should be viewed as being a part of the assembly line as a whole.

The assignment of tasks to stations will influence the minimum time that a given product is available at each station, because it will influence the station length. This is usually known as the Tolerance Time or Station Passage Time. It is the time required for a product to travel through a station (this definition, obviously, is only applied to flow-lines that move at a constant speed, not intermittently). The station passage time is directly related to the station physical limits. The sequencing of models will also influence the station limits. Note that for a non-continuous flow-line the job being assembled will be stationary at the station and therefore the station physical limits are not so relevant. An example of how the line balance influences the station passage time can be seen in the case of a balance solution that results in a service time greater than the assigned station passage time. If incomplete items are to be avoided this station passage time would have to be increased. This is done by either increasing the station length, decreasing the conveyor speed, or both.

Another problem that arises when designing the line is to decide how many different models will be produced on the line. If the models differ too much from each other it may not be practical to produce them together in the same line. If models are too dissimilar, i.e. the assembly tasks for each model are very different, then each workstation on the line will require a great number of different tools and the operators would need to have a large scope of skills in order to work on all models. Although such skills can be learned after a learning period, the need for several different tools in each station may not be desirable. In order to minimize the costs of tooling special attention should be given to avoiding duplicate tooling at different workstations. Very dissimilar models may also result in considerable setup times and therefore the efficiency of the line is reduced.

1.4. Topics Covered in this Research.

This research was essentially directed at the line balancing problem and sequencing of manual mixed-model assembly lines, where the units to be assembled are transported in a moving conveyor system and cannot be removed from the conveyor. The mechanical problems of designing the line are not addressed here.

The complex flow of materials that characterizes an assembly line is assumed to be ideal, which means that the component parts needed during the assembly process are delivered at the right time, at the right place and in the right quantities. It is also assumed that the line operates under ideal conditions, which means that there are no station breakdowns (stations are assumed 100% reliable - this may be true in manual assembly lines because workers are less likely to break down in the reliability sense), no defective component parts, no defective sub-assemblies, etc.

Many of the conclusions that are drawn for manual assembly lines can easily be extended to automated assembly lines and to fabrication lines.

1.5. Research Problem.

Part of this research includes a literature survey of what has been done in balancing and sequencing for mixed-model assembly lines. Line balancing methods for the single-model case are extended to the mixed-model case. An heuristic method for the sequencing problem is presented and discussed. This heuristic uses the same principles of Toyota's Goal-Chasing method. In order to evaluate the effectiveness of the method, a comparison with other sequencing methods is made. A bowl allocation of workloads to stations was compared with two balanced lines for a given problem and conclusions were drawn. Different launching rates are compared for the same conditions and conclusions are developed.

1.6. How the paper is organized.

In the next chapter the terminology of mixed-model assembly lines is introduced. In Chapter 3, line inefficiencies are described. Chapter 4 presents a method for assigning models to the same assembly line. Chapter 5 is dedicated to the line balancing problem. In Chapter 6 the sequencing problem and a method to solve it are presented. A review of previous work in line balancing and sequencing of mixed-model is presented in Chapter 7. A comparative analysis of different methods for sequencing and evaluating an unbalanced line are presented in Chapter 8, and finally in Chapter 9 the conclusions and possibilities for further research are presented.

2. Terminology of Mixed-Model Assembly Lines

Different configurations of assembly lines can be found. This regards the physical configuration of the line as well as the operating conditions (e.g. the use a fixed launching rate, etc.).

2.1. Product Models.

Consider a situation where m models are available and where j (j = 1, 2,..., m) is used to identify each model. Let T be the shift time (minutes during shift to be scheduled), Q_j the quantity of model j to be produced and Q the total number of units to be produced during period T. Hence, the output Q is given by

$$Q = \sum_{j=1}^{m} Q_j \tag{2.1}$$

2.2. Workstations.

Workstations are locations where a given amount of work is performed. The flow line consists of a series of workstations. In a manual assembly line, a workstation consists of workers and may be equipped with tools. The majority of previous work in mixed-model assembly lines (Kilbridge and Wester 1961, Prenting and Thomopoulos 1967, Okamura and Yamashina 1979, etc.) assumes that each workstation is manned by a single operator. In this research the same is assumed.

The number of stations on the line is represented by n and subscript i (i = 1, 2,..., n) is used to identify each workstation.

Workstations can be classified as open or closed. The definition has to do with the type of station boundaries. This definition only has significance in flow lines where there is a continuously moving conveyor and the workers are required to walk back and forth working on the units as they move past. It is usual to symbolize the station boundaries with the symbols (,), [,], denoting, respectively, station open to the left, open to the right, closed to the left, and closed to the right.

The number of workers w_i at each workstation is assumed to be one. Therefore, $w_i = 1$. Station i is said to be the operator's i home station. The total number of workers in the assembly line is given by w, where

$$w = \sum_{i=1}^{n} w_i \tag{2.2}$$

2.2.1. Closed Stations.

In this type of workstation, it is impossible or undesirable for workers to cross the station boundaries. Examples of such stations are locations where the work cannot be accomplished outside the station limits such as spray paint booths, heating chambers, etc. On this type of station, the amount of time required to complete the assembly work must be respected; otherwise the product will leave the station incomplete. These stations are symbolized by [i], i being the station number.

2.2.2. Open Stations.

In this type of workstation the workers are free to cross the station boundaries. Usually the distance that the worker can walk from his station boundaries is limited due to reasons such as limited range of powered tools (Dar-El 1978), etc. The extent of the distance that a worker can walk away from his home station may be restricted or unrestricted. Unrestricted means that the worker is able to move without limits to another location. If adjacent workstations have a region that is common to both, the stations are said to overlap.

Sometimes the workload in a station is so heavy that the worker is forced to cross the downstream station boundary in order to accomplish his work on the product; at other times the workload is so light that the worker walks across the upstream station boundary in order to start work on the next product that has not yet entered the station boundaries. This is not possible with closed stations so the worker will be unable to finish the work on the unit or will be idle waiting for the next unit to arrive. Open stations also result in shorter line lengths than closed stations and, therefore, production cost is likely to be less.

The extent to which operators may cross their station boundaries is conditioned by the requirement that adjacent operators will not interfere with each other's work. Open stations are represented by (i).

2.2.3. Hybrid Stations.

It is possible for assembly line stations to be combinations of the two previous types. For example, a station may be closed to the left and open to the right. This means that the worker is able to cross the downstream boundary of the station, but not the upstream boundary. An obvious example of such station is the first station in an assembly line, where upstream to this station there is nothing and therefore, the station is closed to the left. Obviously an open station is open to the left and to the right, and a closed station is closed to the left and to the right.

2.3. The Transfer System.

The product being assembled may be transferred between stations in the assembly line by a conveyor. The most common is to have a conveyor moving at constant speed. The conveyor speed is designated by V_c (m/min). This is said to be a continuous flow-line. Another type of transfer system is the synchronous conveyor, where every product remains at each station and then abruptly moves to the next, all parts moving at the same time. In this type of line the operators work under paced conditions. However, the use of asynchronous conveyance systems (a product only moves to the next station down the line when the operator has completed the work) is increasing in certain industries (such as computer manufacturing, etc.), because they provide more flexible production systems. It is possible to have a non-mechanical line where jobs are transferred manually between stations by the operators. Non-mechanical and asynchronous lines are unpaced lines and are usually provided with buffer stocks between stations.

Dar-El (1978) classifies the products to be executed on the assembly line as Products Fixed and Products Movable. This definition is based on whether or not the product can be moved independent of the conveyor movement. When it is possible to remove the product from the conveyor or the product is stationary in relation to the moving conveyor the product is designated as Product Movable. If this is not possible the product is designated as Product Fixed. An example of such lines is found in industries where the product is too heavy or too large to be removed from the conveyor (e.g. the automobile industry). In a Product Fixed assembly line, the job to be assembled cannot be removed from its position in the conveyor; however it may be possible to rotate it.

Typically Product Fixed lines do not have buffer stocks that allow products to accumulate between stations with the consequence that operators work under paced conditions. Product Movable are asynchronous systems where buffers are allowed and work is done under conditions that are not so rigidly paced.

In the assembly of large products, such as aircraft, ships, machine tools, etc., the product remains stationary at a given location and the "stations" move from product to product, i.e. when workers at a given station finish their work, they move to another product and are replaced by the workers from the previous station (this type of system is not considered to be a flow-line). Typically the Product Fixed assembly lines are dedicated to the production of heavy products (e.g., automobiles, heavy appliances), and the Product Movable assembly lines to lighter products that can easily be removed from the conveyor (e.g., small appliances, electronic assemblies).

2.4. The Launching System.

With respect to the launching period, there are two possible modes of introducing units into the assembly line: Fixed Launching Rate (FLR) and Variable Launching Rate (VLR). In Fixed Launching Rate, units are introduced into the line separated by a constant time interval. With Variable Launching Rate, the time interval separating two consecutive launches is equal to the first station time of the last unit launched. The fixed launching interval is given by γ. The fixed launching interval is achieved by maintaining a fixed distance in the conveyor between two consecutive units, given by S_p. The time interval γ is also known as the Production Cycle Time and it is the time between successive units coming off the line. It can be seen that if the production requirement is to be achieved γ has to be greater or at least equal to the theoretical production rate. If not, there will be a shortage in production.

For a FLR system if the line works under unpaced conditions the rate at which the products come off the line will be different from the FLR[1]. In this situation the feed of products to stations may become constant only for the first station on the line, and the line will be actually working as if the launching system used was the VLR.

Let γ = fixed launch rate interval (minutes/part); S_p = fixed interval between successive jobs (meters/part). Then

$$S_p = V_c \gamma \quad (m/\text{min})(\text{min}/part) \tag{2.3}$$

2.5. Inventory Buffer Storage.

These are commonly used on manual assembly lines because they allow the

[1]Under paced work conditions, the units are fed into the line at a rate equal to the FLR. Each workstation has available an amount of time to work on the units equal to the FLR; the time interval between parts coming off the line is equal to the FLR. If the line is working under unpaced conditions, some workstations may be working on a product an amount of time different than the FLR and, consequently, the time interval between parts coming off the line will be different from the FLR.

smoothing of work flow (Groover 1987) and reduction of the effects of task time variability (Monden 1983).

The launching discipline becomes irrelevant when buffer storage is allowed. For systems with buffer stocks between adjacent workstations the workers are not confined to paced work conditions. In a station provided with buffer storage, if a job arrives before the worker has completed the work on the previous unit, the arriving unit will be held in the buffer until the operator is free to start work on it. This enables the worker to complete the work on all units. Considerations such as the optimal capacity of a buffer are out of the scope of this research and will not be assessed.

Unpaced lines are likely to eliminate the production of incomplete items but may result in an increase in the throughput time and a corresponding reduction in production rate.

2.6. Minimum Rational Work Element.

Minimum rational work elements are the smallest economic subdivisions of the work required to assemble a product. Below this minimum, assembly work cannot be divided rationally. For example (Kilbridge and Wester 1961), a minimum rational element may include the following motion pattern: reach for a tool, grasp it, move it into position, perform a single task, return the tool. This work element is considered indivisible because it cannot be split over two work stations without creating unnecessary work in the form of extra handling.

Let $T_{e_{jk}}$ = work element time for element k on model j (minutes); k = subscript for work element k, k = 1, 2,..., K; K_j = number of work elements required to assemble one unit of model j.

The sum of all required task times to accomplish one finished unit of model j is known as the total work content time for model j and is represented by T_{wc_j}. The total assembly time needed to complete all units of model j is given by

$$TT_{wc_j} = Q_j T_{wc_j} \tag{2.4}$$

The total assembly time needed to accomplish all operations on all Q units is known as the total work content time and is given by

TT_{wc} = total work content = sum of total assembly time required during period T

$$TT_{wc} = \sum_{j=1}^{m} Q_j T_{wc_j} \tag{2.5}$$

The total time required to perform element k in all units is given by

$$TT_k = \sum_{j=1}^{m} Q_j T_{e_{jk}} \tag{2.6}$$

2.7. Precedence Constraints.

Also known as "technological sequencing requirements" (Groover 1987), precedence constraints are the reason why work elements must comply to a certain sequencing order. An illustration of a precedence relation is in the assembly of a small electric appliance; a switch must be mounted onto the motor bracket before the cover of the appliance can be attached (Groover 1987). Because work elements are subjected to precedence constraints the sequence in which the assembly work can be accomplished is restricted.

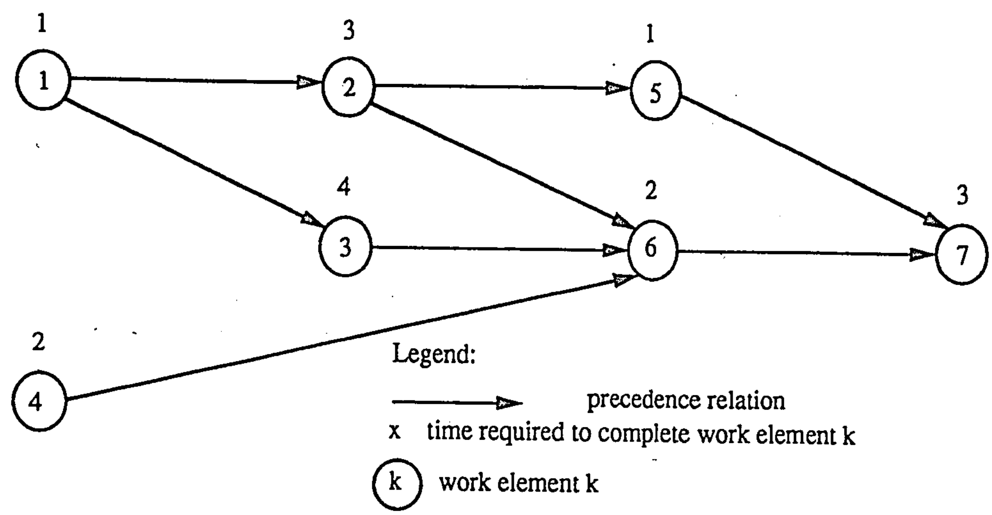
Work elements may be subjected to other types of constraints. Zoning constraints means that a task may have to be placed near other tasks, preferably at the same workstation - positive zoning; or that the task may have to be distanced - negative zoning - this case happens when tasks may interfere with one another. Sometimes it may be required that some tasks from one model be performed at the same station as certain tasks from other models. These constraints are known as locational constraints and arise usually when some work elements require specialized skills or equipment (Villa 1981).

Another type of constraint is related to the position of workstations and is called a position constraint. This type of constraint is found in the assembly of large products (e.g., automobiles) where the product dimensions are such that one operator cannot perform work on all sides. In situations like this one, operators are located on the two sides of the assembly line (Groover 1987).

2.8. The Precedence Diagram.

The precedence diagram is a graphical representation of the precedence constraints among work elements. It is composed of nodes which symbolize the work elements and arrows which indicate the order in which elements must be performed. The sequence in which work elements are performed progresses from left to right; the elements at the left of the diagram must be done first (Groover 1987). Usually work element times are shown above each node. An example of a precedence diagram for a single-model assembly line is illustrated in Figure 2.1.

Figure 2.1: Example of a Precedence Diagram for a Single-Model Assembly Line.



In a mixed-model assembly line the precedence diagram includes the precedence relations for each model. Above each node there are indicated the total times required to perform that element on all units (i.e., TT_k). An example of a mixed-model precedence diagram is shown in Figure 5.3.

2.9. Station Service Time.

$T_{s_{ij}}$ is defined as the service time for model j at station i, which means $T_{s_{ij}}$ is the time to assemble model j at station i. Hence

$$T_{s_{ij}} = \sum_{k \in i} T_{e_{jk}} \tag{2.7}$$

where $T_{e_{jk}}$ are the work elements assigned to station i for model j.

2.10. Station Time.

The total time per shift required at station i to assemble the Q required units will be designated station time and is given by

$$TT_{si} = \sum_{k \in i} \sum_{j=1}^{m} Q_j T_{e_{jk}} \tag{2.8}$$

It can be seen that the total work content time must be equal to the sum of the station times, i.e.,

$$TT_{wc} = \sum_{i=1}^{n} TT_{si} = \sum_{k=1}^{K_j} \sum_{j=1}^{m} Q_j T_{e_{jk}} \tag{2.9}$$

2.11. Repositioning Time, Operator Walking Speed and Operator Upstream Walking Distance.

When the operator has finished the work on a unit, he has to walk in the upstream direction until he reaches the next unit and starts working. Let V_o be the operator walking speed when operator is walking upstream. Once the operator is moving parallel to the conveyor, accomplishing assembly work on the unit, the speed of the operator walking downstream is the speed of the conveyor.

In a continuous flow line when an operator completes work on a given job and walks upstream to the next job, the time interval between the moment he left the current job until the moment he reaches the next job is the repositioning time T_r.

$$T_r = \frac{S_p}{V_c + V_o} \tag{2.10}$$

The operator upstream walking distance L_w is given by

$$L_w = V_o T_r = \frac{V_o S_p}{V_c + V_o} = \frac{V_c V_o \gamma}{V_c + V_o} \tag{2.11}$$

Because the speed of the operator walking upstream is greater than the speed of the conveyor it is usually considered that the time required for an operator to walk between two consecutive units can be neglected (Dar-El and Cother 1975, Kao 1981), i.e. the repositioning time is neglected.

2.12. Station Dimensions.

Let L_i represent the length of station i. $L(u)_i$ and $L(d)_i$ are respectively the maximum distance that an operator can move past the upstream and downstream station limits (the lengths are in meters and V_c in m/min).

2.13. Tolerance time or Station Passage Time.

As defined in the previous chapter the station passage time is the time an operator has available to work on a job from the moment that job reaches his station limits until the moment it passes the downstream boundary. In this thesis the notation given by Kilbridge and Wester (1963) for the station passage time will be used which is τ_i. The station passage time is a function of the station length and the conveyor speed. Hence,

$$\tau_i = \frac{L_i}{V_c} \tag{2.12}$$

Kilbridge and Wester (1963) showed that for a continuous flow-line with non-overlapping stations the station passage time must be equal to or greater than the maximum service time for that station, i.e. $\tau_i \geq \max\left(T_{s_{ij}}\right)$, for j = 1,..., m and i = 1,..., n[2]. However, if operator idle time is to be avoided then

$$\tau_i = \max\left(T_{s_{ij}}\right) \tag{2.13}$$

The upstream and downstream allowance times respectively $\tau(u)_i$ and $\tau(d)_i$ are a function of the upstream and downstream allowance distances $L(u)_i$ and $L(d)_i$ and of the conveyor speed. Hence

$$\tau(u)_i = \frac{L(u)_i}{V_c} \tag{2.14}$$

$$\tau(d)_i = \frac{L(d)_i}{V_c} \tag{2.15}$$

In Product Fixed assembly lines where the launching interval and the station lengths are fixed, the station passage time can be altered by an appropriate choice of conveyor speed and spacing between units. The effect of task time variability may be reduced by the appropriate choice of item spacing and conveyor speed.

2.14. Early Start and Late Start Schedule.

Early and late start schedule are defined in relation to the position where the operator in each station receives the first job in the sequence. An early start means that an operator at a given station starts to work on the first job of the sequence as soon as it enters his station limits. The operator is positioned next to the upstream limit. Late

[2]For a closed station system, with no buffer storage allowed, $L(u)_i=0$ and $L(d)_i=0$, and therefore, τ_i is the time separating the start of work on a given unit by consecutive stations i and i+1.

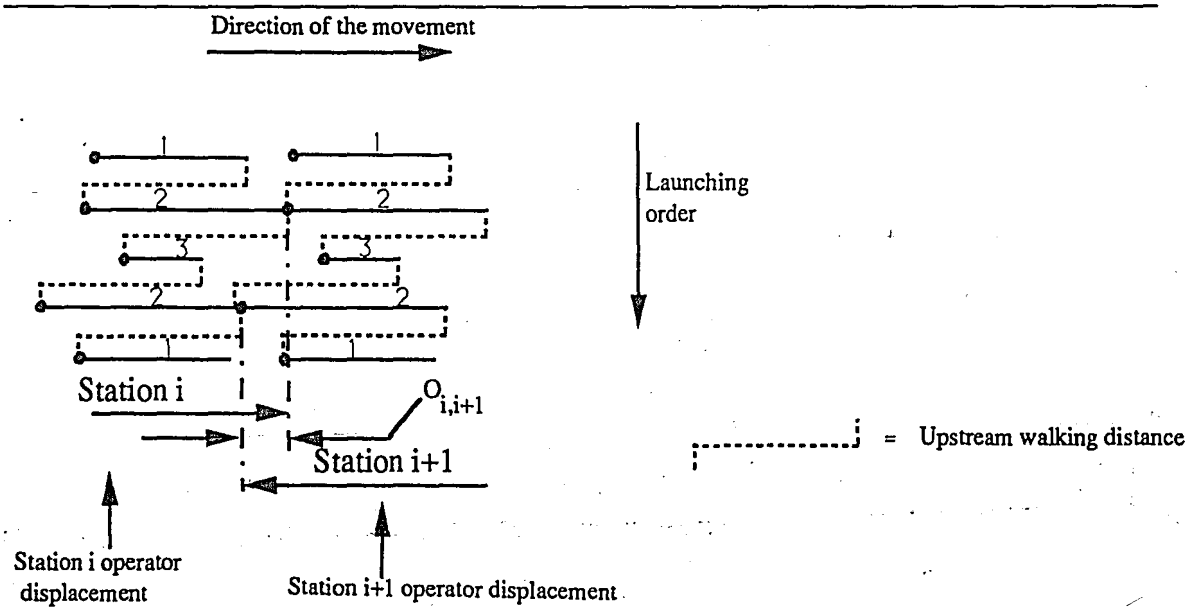
start means the operator does not start work on the first job immediately when that job enters his station limits.

2.15. Concurrent Work and Station Overlap.

When two operators at adjacent workstations are allowed to work simultaneously in the same unit, it is said that they are working concurrently. Operators should not interfere with each other while working concurrently.

It is said that two stations overlap when the operators of each station are allowed to work in an area that is common to both stations. Overlapping between stations i and i+1 is given by the maximum downstream position for station i, minus the maximum upstream position for station i+1. The case is illustrated in Figure 2.2, where $O_{i,i+1}$ represents the overlapping between stations i and i+1, i.e. the region common to both.

Figure 2.2: Example of Station Overlap.



2.16. Minimum Part Set.

The total number of units (Q) to be assembled in the planned schedule is the sum of the number of units of each individual model to be assembled (Q_j). The total production requirement Q, or equivalently the total part set can be represented by a vector of integers $Q=(Q_1, Q_2, \ldots, Q_m)$. If q is the common divisor of the number of units of each model then the vector of integers MPS = $(Q_1/q, Q_2/q, \ldots, Q_m/q)$ is the smallest partition of the total part set having the same proportion as the production target Q. MPS stands for Minimum Part Set, and can be viewed equivalently as

$$MPS = \sum_{j=1}^{m} \frac{Q_j}{q} \qquad \textbf{(2.16)}$$

2.17. Lines with Several Labor Groups.

When the assembly line includes several disjointed but related areas or labor groups[3], each area is considered independently of the others. The minimum number of stations required per labor group g (g = 1, 2,..., G) is given by

$$n_g = \min\ intg \geq \frac{TT_{wcg}}{T} \qquad \textbf{(2.17)}$$

where n_g is the number of stations for labor group g, TT_{wcg} is the amount of time required to complete all operations in all units for a given production schedule, i.e. the total work content of labor group g. The total number of stations in the assembly line is

[3]E.g., an automobile assembly plant is often divided into several separated areas that are related to each other. Examples of such areas are the Body Assembly, Paint Shop and Final Assembly. These areas are related in the sense that an automobile being assembled progresses sequentially through them.

$$\sum_{g=1}^{G} n_g \tag{2.18}$$

2.18. Task Time Variability.

When the assembly work is performed by human operators it is inevitable that there will be some variability in task times. Task times are usually considered deterministic but they are in fact stochastic. In automated assembly processes it is possible to achieve virtually deterministic assembly times.

2.19. The Multi-Function Worker.

The multi-function worker is one who is capable of performing a large scope of different tasks. Traditionally workers are limited to a certain number of tasks, which they master. The cross-training of operators enables transfer to other workstations where they may be more useful. This is particularly beneficial in workstations where operators are overloaded with work and cannot finish it; if another workstation is under-occupied, the operators at this station may be able to travel to the overloaded station and help the overloaded operators.

The utility worker is a multi-function worker in the sense that he is able to perform a large variety of tasks. This type of worker "floats" (he is not assigned to a particular workstation) in the assembly line and works on the stations that have fallen behind, i.e. on the stations that due to a overload of work are not able to complete the work. Cross-training of operators helps to avoid boredom and increase job motivation.

3. Assembly Line Inefficiencies

Four types of inefficiencies can be defined in mixed-model assembly lines (Thomopoulos, 1974; Macaskill, 1973). Some of these inefficiencies only make sense in a moving conveyor line.

3.1. Idle Time.

Idle time can occur when the operator is kept idle waiting for work to enter the limits where he is allowed to work. The operator is available to work, but is restrained from working.

3.2. Work Congestion.

Congestion occurs when the assembly work is done beyond the station downstream limit, in the downstream allowance region. When jobs flow through a station faster than the operator can complete them, the operator is forced to pass his downstream station boundary in order to complete the work. This type of inefficiency only happens in stations with an open boundary to the right. If the station is closed to the right, then the unit leaves the station incomplete.

3.3. Work Deficiency

Work deficiency occurs when the assembly work is done before the upstream limit, in the upstream allowance region. When jobs flow so slowly that the operator is able to complete the work on the current job before the next one has entered his station limits, and if the worker wants to avoid becoming idle, he has to cross his upstream station boundary and start work on the next unit. This type of inefficiency only occurs on stations with an open boundary to the left.

3.4. Utility Work.

This occurs when the worker is not able to complete the work within his working limits, and the job leaves the downstream limit of the working area unfinished. In these situations a utility worker can be assigned to the station to assist the operator so that the unit can be finished or the unfinished work will be completed in a station further down the line.

3.5. Comments.

The account of these inefficiencies is a measure of the assembly line inefficiency. Some other measures are sometimes used, such as the Home Time defined by Sumichrast et al. (1992) as the percentage of time that workers are at their home station (working and idle). The overall assembly line length is sometimes a measure of the assembly line inefficiency because the greater the line length the greater is likely to be the production cost.

Work deficiency, work congestion, home time, and utility work make sense only in continuous flow-lines.

Minimum throughput time is equivalent to maximum throughput (production rate), and therefore, measures that seek to minimize throughput time should be taken.

It is highly undesirable to produce incomplete items at any station. An incomplete job in a station may preclude that job being worked in the following stations. These stations will become idle and a sharp decrease in throughput may occur, with the inherent risk of not achieving the required output!

Work congestion and work deficiency are not critical inefficiencies because they only affect production times, not idle time.

Utility work is not desirable because it means at least one extra worker, which will increase production cost. If it is possible to complete the unfinished jobs in a station further down the line, the passage of incomplete jobs to the following stations may not affect the work on those jobs. Nevertheless, a utility station is an extra station on the line and an increase in the production cost.

In general, zero work congestion is a sufficient condition to avoid incomplete items. To avoid incomplete items could be achieved by a situation of zero utility work. Utility work is directly related to the station downstream allowance limit and this limit may be very difficult to determine, whereas the station length may be more easily determined. Therefore, to design a system that results in zero utility work is quite complicated. To design a system with no work congestion will be easier, and this is the reason why it seems more appropriate to minimize work congestion. In extreme cases of work congestion the operator may not be able to catch up with the arriving items and the line may need to be stopped. Avoiding work congestion will preclude this situation from happening.

The two major priorities seem to be minimization of throughput time and avoidance of work congestion. The methodology presented in Chapter 6 for determining the station lengths will avoid work congestion and, therefore, the priority will be to minimize throughput time and idle time. It will also be shown that a certain amount of idle time may be useful to lessen the effect of task time variability and it probably helps worker morale.

The type of inefficiency that should be minimized is dependent on the configuration of the line and on other aspects that may not be very easy to identify. Each particular situation will require a different objective and, consequently, what is a priority for a certain system may not be a priority for another type of system. For example, in an asynchronous system, the terms work congestion or work deficiency do not apply.

4. Assigning Models to the Assembly Line

In order to produce on the same line two or more models, the nature of work to be performed on the different models must have at least some similarities. Otherwise it would not make sense to use the same line for the different models because it would result in unreasonably large station work load. The problem of allocating models to lines was first studied by Lehman (1969 - on Buxey et al. 1973) who developed a heuristic to assign groups of models for lines based on minimizing the costs associated with balance delay, idle time due to the sequence used and operator learning.

Another possible methodology to assign models to a line is to allocate to the same line the models that have the greatest similarity from an assembly point of view.

Thomopoulos developed a measure, called Similarity Index, that evaluates the similarity of work element tasks between two or more models (Prenting and Thomopoulos 1974). This measure can be used for allocating models to the same line. For example, if it is desired that each assembly line produces simultaneously two or more different models, the models assigned to each line will be such as the similarity index is maximized for a given combination. A similarity index of zero means that the models have absolutely no similarities, whereas an index of one means that the models are identical. The common situation is to allow for element times to vary among elements and models, i.e. some tasks may have different times for different models, including zero time, which means that the task is not performed in that model.

To explain the Similarity Index, consider m different models that are to be assembled. If $s_1, s_2, \ldots, s_m$ represents any combination of m models, then these m models can be grouped into a set $(s_1, s_2, \ldots, s_c)$. For $2 \leq c \leq m$ a set is called a model set and is designated by s^*. The number of models in a model set is designated by m_{s^*}.

Let t_{s^*k} be the sum of the times for task k over a model set s^*, i.e.,

$$t_{s^*k} = \sum_{s^*} T_{e_{jk}} \tag{4.1}$$

Assuming that all models in model set s^* have the maximum time defined for element k ($k = 1, 2, \ldots, K$), let T_{s^*k} be the sum of the times for task k over model set s^*, defined by

$$T_{s^*k} = m_{s^*} \max\left[T_{e_{jk}}\ (j \in s^*)\right] \tag{4.2}$$

for $k = 1, 2, \ldots, K$, where K is the maximum number of tasks for set s^*.

A measure of the utilization for element k over all models in s^* is given by u_{s^*k}, where

$$u_{s^*k} = \frac{t_{s^*k}}{T_{s^*k}} \tag{4.3}$$

defined only when $T_{s^*k} \neq 0$.

In order to have the utilization index taking values from zero to unity, u_{s^*k} is transformed into s_{s^*k}. Hence,

$$s_{s^*k} = \frac{u_{s^*k} - \frac{1}{m_{s^*}}}{1 - \frac{1}{m_{s^*}}} \tag{4.4}$$

and $$0 \le s_{s*k} \le 1 \quad (k = 1, 2, \ldots, K) \tag{4.5}$$

The variable s_{s*k} is called the similarity index for task k in set s^*. When $s_{s*k} = 1$, all the task times for task k in set s^* are the same and therefore, the association is maximized. For $s_{s*k} = 0$ only one model in set s^* requires task k, and the association is minimum.

Defining U_{s*k} as the similarity index for all tasks in set s^* (accounts with the weighted average of all u_{s*k}), then

$$U_{s^*} = \frac{\sum_{k=1}^{K} t_{s^*k}}{\sum_{k=1}^{K} T_{s^*k}} \tag{4.6}$$

Similarly, the weighted average of all s_{s*k} is given by

$$S_{s^*} = \frac{U_{s^*} - \frac{1}{m_{s^*}}}{1 - \frac{1}{m_{s^*}}} \tag{4.7}$$

where $0 \le S_{s*} \le 1$. S_{s*} is the similarity index for all elements in set s^*.

When $S_{s*} = 1$, all task times in set s^* are identical, and the maximum association of models prevails[1]. If $S_{s*} = 0$, each task is performed in only one model of the set, and therefore, the minimum association is found[2].

In a mixed-model assembly line S_{s*} will take a value that is typically greater than zero. In order to assign models to different lines all the possible sets should be considered, and the assignments will be such as the similarity index S_{s*} is maximized. For more than one assembly line, Thomopoulos suggests that the assignment should be the one that results in the highest average similarity index.

The similarity among units produced on the line can also be measured. To this purpose, in order to account for the possible different quantities required for each model, a modification of the S_{s*} is needed. For more details on the similarity index refer to Prenting and Thomopoulos 1974.

This similarity index among units produced may be used in batch-model lines. A criterion to sequence batches down the line, may be to determine the sequence that maximizes the sum of similarity indices (Prentice and Thomopoulos 1974).

Other similarity indices can be generated using the following considerations; for example, an index based on component parts used in the different models. It may be useful to assign to the line the models that have the highest association among component

---

[1]$S_{s*} = 1$ means that all models require the same tasks and that the time required to perform a particular task is the same for all models. In this case the models are analogous.

[2]In such case the models in set s^* are totally dissimilar and if possible should not be produced on the same assembly line.

parts. This to help keep the quantity of each part used by the assembly line closest to constant. This also would reduce the sources required for these parts. The index may be obtained by using the parts associated with each model instead of work elements.

The similarity index is not the only criteria for assigning models to lines. Other factors may enter the decision, and they usually do; for example, it may be impractical for a certain company, to create an assembly line dedicated exclusively to a certain product and therefore this product will be produced in the existing assembly line together with other products.

5. Line Balancing

In order to produce a finished product a set of assembly tasks have to be performed. The total amount of work needed to perform the assembly of a final product is divided into individual tasks and assigned to successive stations along the line. It is both fair and efficient to apportion equal amounts of assembly work to the workstations. The process of assigning as evenly as possible the assembly work among the workstations is known as Line Balancing. For each product there is a certain number of tasks required to complete a finished item. These tasks can be grouped in many ways and still rationally produce the finished item. Grouping the tasks is a combinatorial problem. For a product with K work elements there are K! possible sequences of elements. However, not all sequences are feasible ones. The sequence of processing steps may be restricted. Work elements are subjected to precedence constraints which means that some tasks cannot be performed before others. Other types of constraints (mentioned in section 2.6) may also be present in the line balancing solution.

It is desirable to have a smooth production, i.e. to have parts coming off the line evenly spaced. This is possible if the workstation processing times are approximately equal.

Line balancing is an attempt to group the assembly tasks so the total time required at each workstation is as close to the same as possible. Different acceptable groupings of tasks can result in different amounts of assembly line nonproductive time and even alter the number of workstations required for a desired production output. Kilbridge and

Wester (1961) estimated that "industry can waste four to ten percent of operator time on assembly lines through unequal work assignments". If the station times are equal a perfect balance is achieved. If this does not happen, and this is the common situation, the slowest station will set the overall production rate of the line. Whereas line balancing is relatively easy to achieve in a single-model assembly line, the problem becomes complicated in mixed-model assembly lines because there are several models, each model requiring different times at different workstations.

The line balancing problem for single-model lines is defined as:

> to assign work elements to stations in such a manner that all precedence constraints are respected and the minimization of the total amount of idle time or equivalently the minimization of the number of workstations is achieved. The process time for each station must not exceed the cycle time.

The problem of line balancing was first studied by Bryton (Moodie and Young 1965) in 1954, for single-model assembly lines. Salveson is usually credited as being the first who formulated line balancing as a linear programming problem. Since that time several solutions have been presented in the literature.

Thomopoulos (Prenting and Thomopoulos 1974) appears to be the first to have explicitly studied the problem of line balancing for mixed-model assembly lines. He presented a method of assigning tasks to stations that assures that similar tasks are assigned to the same workstation or the same group of workstations. The methodology is the same as the one used for single-model line balancing problems, but instead of using cycle time to limit the workstation time, it uses the total time desired to assemble the required number of products. Also, together with the production requirements a mixed-model precedence diagram is used. Above each node is written the total time per schedule to perform that element (i.e. the time to perform that element on all units that require it).

This methodology is sometimes called aggregated task-group balancing, or simply task-group balancing. The repetition of a given task will be assigned to the same station. This has the result that each task is assigned to only one station and, consequently, no other operator(s) need to have the skills necessary to perform that work element. Therefore the time and cost of learning is reduced and the general efficiency of work is improved (Prenting and Thomopoulos 1974)[1]. Because different work elements require different tools, different skills, etc., it should be provided that similar work elements are performed in the same station or group of stations.

When the products to be assembled on the line are of a similar nature, i.e. when the work on each product involves similar elements performed in a similar order, independent line balance[2] produces fairly satisfactory results. In these circumstances the workers would have to perform the same type of work operations independently of the type of model that is being produced. If the models to be produced are dissimilar,

---

[1]This opinion is currently contradicted. Although it may be desirable to assign similar work elements to the same group of stations because of the use of similar tools, configuration of workstations, etc., it is thought that workers should be trained to accomplish a variety of tasks. This is the concept of the multi-function worker. Further development is given in page 26.

[2]Independent balance means to balance the line for each product considering that the line produces only that one type of product (i.e., the line is functioning as a single-model line).

independent line balance is likely to result in dissimilar work elements being allocated to each station (Wild 1972).

5.1. Mathematical Formulation of the Line Balancing Problem for Mixed-Model Assembly Lines.

In mixed-model assembly lines the shift time T is a basis of reference, the same way as the cycle time is a basis for reference in single-model assembly lines.

For the scheduled time period to assemble the desired output we have: $T_{e_{jk}}$ = is the time to perform task k on model j; n = number of workstations on the line; i = subscript for the workstations; TT_{si} = total time per station during the scheduled period T; and TT_k = total time per shift required to perform task k on all units. $T_{e_{jk}}$ depends on the job complexity, tools available, fixtures, operator skill, etc.

T_{cj} = cycle time for model j. It is the time between successive units coming off the line. The cycle time for model j is the maximum station time for model j. Note that this definition, valid for single-model assembly lines, does not make much sense in mixed-model lines where the time between successive units of model j may not be equal to T_{cj}.

The theoretical cycle time, defined as the maximum time that a unit should spend at a work station, can be written as

$$T_c \leq \frac{E}{R_p} \tag{5.1}$$

were T_c is the theoretical cycle time, E is the line efficiency of the assembly line (in this text 100% efficiency will be assumed, i.e. there are no downtimes, etc.), and R_p is the required production rate (R_p is the number of units of all models to be produced over the time available to produce them). E = 100% is a fairly acceptable simplification for the case of manual assembly lines - where mechanical malfunctions are less likely than in automated lines.

For a production requirement of Q units per shift T, assuming no system breakdowns (i.e., E = 100%), the theoretical rate of production $R_c = R_p$ is given by $\frac{Q}{T}$ or equivalently the theoretical cycle time is given by

$$T_c = \frac{T}{Q} \tag{5.2}$$

Let T_{wc_j} be the total work content for model j, i.e. the assembly time required to produce one unit of model j. Hence,

$$T_{wc_j} = \sum_{k=1}^{K_j} T_{e_{jk}} = \sum_{i=1}^{n} T_{s_{ij}} \tag{5.3}$$

where $T_{s_{ij}}$ is the station service time for model j at station i[3]. The total time required to produce all units of model j is designated by TT_{wc_j}, where

$$TT_{wc_j} = Q_j T_{wc_j} \tag{5.4}$$

The total assembly time required to perform all units in the scheduled period is known as the total work content (TT_{wc}), and is defined as:

---

[3]The line balancing solution will determine these times.

$$TT_{wc} = \sum_{j=1}^{m} \sum_{k=1}^{K_j} Q_j T_{e_{jk}} = \sum_{j=1}^{m} TT_{wc_j} \tag{5.5}$$

TT_{wc} represents the total amount of work that is to be accomplished on the line during the scheduled period T.

The total time required to perform element k in all units is

$$TT_k = \sum_{j=1}^{m} Q_j T_{e_{jk}} \tag{5.6}$$

The theoretical minimum number of workstations n^* is given by,

$$n^* = \min \textit{ integer} \geq \frac{\sum_{j=1}^{m} \sum_{k=1}^{K_j} Q_j T_{e_{jk}}}{T} = \frac{TT_{wc}}{T} \tag{5.7}$$

If n is the known number of stations, the total service time per station TT_{si} will be

$$TT_{si} \leq T \tag{5.8}$$

In a situation of perfect balance all TT_{si} are equal. If all $TT_{si} = T$, then we have 100% efficient use of the scheduled period T.

The total time required to perform all tasks in all models is equal to the sum of the station times. Hence,

$$\sum_{i=1}^{n} TT_{si} = \sum_{j=1}^{m} \sum_{k=1}^{K_j} Q_j T_{e_{jk}} = TT_{wc} \tag{5.9}$$

The line balancing problem for mixed-model assembly lines can now be stated as:

> to assign work elements to workstations in such a manner that all constraints (precedence and others) are respected and the minimization of

the total amount of idle time or equivalently the minimization of the number of workstations is achieved. The total assembly time per workstation must not exceed the shift time, otherwise the required output cannot be achieved. Therefore the objective is to

$$minimize \sum_{i=1}^{n} (T - TT_{si}) \tag{5.10}$$

where $T \geq TT_{si}$, $i = 1, 2, \ldots, n$.

Minimizing Eq. 5.10 is equivalent to minimizing the number of stations or the shift time or the product of the two, depending on what is held constant. Note that

$$\sum_{i=1}^{n} (T - TT_{si}) = nT - \sum_{i=1}^{n} TT_{si} = nT - constant \tag{5.11}$$

then,

$$\begin{aligned} \min\ (nT - constant) &= \min\ (nT) - constant \\ &= T[\min\ (n)] - constant \\ &= n[\min\ (T)] - constant \end{aligned} \tag{5.12}$$

It should be noted that $\sum_{i=1}^{n} TT_{si} = TT_{wc}$ is a constant of the problem. Therefore, line balancing for mixed-model assembly lines is achieved by finding the assignments of tasks to workstations that minimizes the product of n by T. If one of these variables is fixed, the line balance reduces to minimizing T (case where n is fixed) or minimizing n (case where T is fixed). If both n and T are fixed, the amount of idle time will be the same, independently of the balancing solution.

Sometimes, instead of using Eq. 5.10, the station times are bounded by a minimum and a maximum station time, i.e.,

$$TT_L \le TT_{si} \le TT_H \tag{5.13}$$

where TT_L and TT_H are respectively the lower and higher desirable station times (total times per period T). The station time may be below TT_L because of precedence constraints or other type of restrictions.

Methods that solve the line balancing problem for single-model lines may be used for solving the mixed-model balance. The line balance for single-model lines is based on the cycle time and on the work element times - assign work elements to stations such that the sum of the work element times in each station is less than or equal to the cycle time. In mixed-model lines, the cycle time is replaced by the shift time, and the task times are replaced by the total task time per shift (i.e, the time required to perform a work element in all units). Hence, instead of T_c, T is used, and instead of T_{ek}[4], TT_k is used (as mentioned before, this procedure is commonly known as task-group balancing). The use of total task times rather than the task time per model (TT_k rather than $T_{e_{jk}}$) will result in assigning each work element to only one station.

The combined precedence diagram for the product-mix should be constructed. This diagram combines the precedence diagrams of the m different products.

5.2. Measures for the Efficiency of the Balance Solution.

Measures that evaluate the efficiency of the balance solution can be computed. These measures are the Balance Delay, Balance Efficiency and the Smoothness Index.

---

[4]On a single-model assembly line T_{ek} is the time to perform task k.

5.2.1. Balance Delay.

The balance delay measures the line inefficiency that results from idle time due to imperfect assignment of work elements to workstations. The balance delay may be defined as a percentage or a decimal fraction. For a given schedule T the balance delay d is given by

$$d = \frac{nT - \sum_{i=1}^{n} TT_{si}}{nT} \tag{5.14}$$

For a perfect balance[5] (i.e., evenly distributed station times) the balance delay is zero, and therefore $nT = \sum_{i=1}^{n} TT_{si}$.

Sometimes instead of using d the Balance Loss is used. This is defined as $nT - \sum_{j=1}^{m} Q_j T_{wc_j}$, which is equivalent to

$$nT - \sum_{i=1}^{n} TT_{si} \tag{5.15}$$

Values for balance delay ranging between 5 to 10% are usually considered acceptable.

5.2.2. Balance Efficiency.

Sometimes instead of balance delay balance efficiency is used, which is the counterpart of d. Balance efficiency ϵ is given by

[5]It is impossible to achieve a perfect balance in manual assembly lines because of work element and station times variability.

$$\epsilon = \frac{\sum_{i=1}^{n} TT_{si}}{nT} \cdot 100\% \tag{5.16}$$

5.2.3. Smoothness Index.

Moodie and Young (1965) defined a measure for single-model assembly lines, which they called Smoothness Index (S.I.). It indicates the smoothness of a given balance. For this type of assembly lines, the smoothness index is defined as

$$S.I. = \sqrt{\sum_{i=1}^{n} (T_{\max} - T_{s_i})^2} \tag{5.17}$$

where T_{max} is the maximum station time required to assemble one unit (of the model produced), and T_{s_i} is the time required to produce one unit at station i. A perfect balance would result in a smoothness index of 0 (in a situation of perfect balance $T_{max} = T_{s_i}$). We modified the smoothness index in order to extend this definition for the case of mixed-model assembly lines. The smoothness index becomes the square root of the sum of squares of station times deviations from the maximum station time per period T (TT_{smax}). These station times are total times per shift. Hence,

$$S.I. = \sqrt{\sum_{i=1}^{n} (TT_{s\max} - TT_{si})^2} \tag{5.18}$$

for i = 1, 2,..., n.

For a perfect balance $TT_{smax} = TT_{si}$, i = 1, 2, ..., n, and therefore the smoothness index would be equal to zero.

5.3. Heuristics for Solving the Line Balancing Problem.

Several authors (Kilbridge and Wester 1961, Prenting and Thomopoulos 1974, Macaskill 1972) argue that the size and complexity of some problems make the use of algorithms yielding optimal solutions impractical, and because of that, they defend the use of heuristics. The use of the most popular heuristics is found to yield near optimal solutions. Mastor (1970) investigated the effectiveness of several line balancing methods for single-model assembly lines. The measure of effectiveness used was the maximum output rate obtained for a line with a specified number of stations. The speed of computation was used as a measure of the cost of computation. Among the line balancing techniques evaluated, the method proposed by Held, Karp and Sharesian (1963) consistently achieved the best results. However, this method required a greater computing time than the Comsoal method developed by Arcus (1966). Sophisticated methods for line balancing are available. The effectiveness of the solution obtained can be increased with the use of those sophisticated techniques. In general, simpler methods such as the Largest Set Rule or the Ranked Positional Weight are likely to be adequate and less costly for the majority of situations.

Mastor (1970) alludes to the "Station-to-Work-Element-Ratio" (or simply the station-task-ratio) which is the total number of stations divided by the total number of work elements; when for a given line balancing problem the number of workstations increases, the station-task-ratio increases. Mastor argues that as the average number of work elements per workstation decreases, there are fewer combinations of work elements that may be assigned to a workstation. As a consequence, each station may not be using all the time available to that particular station. In such a situation, a greater amount of idle time results.

Macaskill (1972) argues that the station-task-ratio is lower[6] for mixed-model assembly lines than for single-model assembly lines because there are multiple models in the MMAL, each having its own unique elements; and therefore the balance effectiveness tends to be higher in mixed-model assembly lines than in single model lines. This permits the conclusion that the use of less sophisticated methods will not degrade the efficiency of the line balancing solution. Also, in mixed-model problems, the sequencing of products may diminish or even negate apparent advantages that are obtained by the use of sophisticated balancing methods.

Macaskill (1972) evaluated the performance of a computer program that used the R.P.W. (Ranked Positional Weight) technique and concluded that the task-group balancing (i.e., the balance is based on total task times) obtains quickly balanced solutions in which each task is always assigned to the same station. It was also concluded that due to the fact that mixed-model problems have lower station/task ratios than single-model problems, quite unrefined balance methods will often result in balances of acceptable efficiency. The results of the computer simulation showed that the balance of large-scale mixed-model problems, using task-grouping methods, results in relatively

[6]A lower station-task-ratio means that the average number of tasks per station is higher.

small computer time and the storage requirement is not excessive.

Methods based on heuristics (therefore, they do not guaranty an optimal solution) that are very often used for line balancing are the Largest-Candidate Rule and the Ranked Positional Weight method. These two, according to Macaskill (1972) do not differ in relation to the resultant balance efficiency and the computational time is approximately the same. However, in his study, Macaskill used the R.P.W. technique because "it allows the order of assignments to be changed easily by making arbitrary changes in the positional weight values".

Sometimes, although a solution obtained by the two above mentioned methods is satisfactory, it may not be appropriate from a material-handling viewpoint. It may happen that tasks that belong to different subassemblies are assigned to the same station (We-Min Chow 1990). To feed different subassemblies to the same station may be impractical and a source of problems.

Another balancing method, the Largest Set Rule attempts to assign tasks in such a way that each station will, when possible, be assigned tasks that belong to only one subassembly. In relation to computer effort, the Largest Set Rule does not differ from the two above mentioned methods. It has the advantage of attempting to assign tasks that belong to different sub-assemblies to different workstations. Balance efficiency seems to be at least as good as the efficiency achieved by the other two methods. In some cases it even performed better (We-Min Chow 1990). This was also verified with the example presented in section 5.3.2.

5.3.1. Example of a Line Balancing Heuristic: The Largest Set Rule.

We-Min Chow (1990) presents the Largest Set Rule algorithm for single-model assembly lines. The following is an adaptation to the mixed-model assembly line problem that was developed in the current study. The MMAL Largest Set Rule algorithm proceeds as follows:

1. For each work element k calculate the total task time per shift:

   $$TT_k = \sum_{j=1}^{m} Q_j T_{e_{jk}}$$

2. The weight factor w_k for each work element k is defined by

   $w_k = \sum_{k' \in p_k} TT_{k'} + TT_k$, where p_k = set of all work elements preceding work element k in the precedence diagram.

3. Let S be the set that is composed off all work elements and i = 0.

4. i = i+1. Let T_i be an intermediate variable in the calculation procedure. T_i = T (T is the shift time).

5. Calculate the weight factor for each element in S. Find the work element $\bar{k}$ with largest weight factor less or equal to T_i, i.e. $\bar{k}$: $w_{\bar{k}} = \max (w_k) \leq T_i$. If there are no work elements satisfying these conditions go to step 7.

6. Assign work element k and all its precedents in S to station i. Delete from S the work elements just assigned. Reduce T_i by w_k.

$S = \varnothing$? Yes. Then the solution is found.

No. Go to step 5.

7. If $T_i = T$, then stop. The shift time is too small and no feasible solution exists. If $T_i < T$ then a new workstation should be added to the line. Go to step 4.

Note: It may not be possible to solve the line balancing problem when the shift time is too small. One example of this is when the number of stations is imposed. If the shift time is too small it will be impossible to keep to the given number of stations without resulting in a station time greater than the shift time. One possible way of dealing with too small shift times is to use parallel stations.

5.3.2. Application of the Largest Set Rule to a Line Balancing Problem.

The following example was developed for a Lehigh class. Consider a mixed-model assembly line where two similar models (models 1 and 2) of a product are to be produced. The line is composed of 4 stations and the shift time is 60 minutes. The production requirements for the shift time are 7 units of model 1 and 5 units of model 2. The elements, work element times and precedence constraints are given respectively for model 1 and model 2 in tables 5.1 and 5.2 below. Task times are assumed deterministic.

Table 5.1: Elements and Precedence Constraints for Model 1

Element	Element Time	Immediate Predecessor(s)
1	1	-
2	3	1
3	4	1
4	2	-
5	1	2
6	2	2, 3, 4
7	3	5, 6

Table 5.2: Elements and Precedence Constraints for Model 2

Element	Element Time	Immediate Predecessor(s)
1	1	-
2	3	1
3	4	1, 8
4	2	8
6	2	2, 3, 4
7	3	6, 9
8	4	-
9	2	4

The precedence diagram for each model is illustrated in Figures 5.1 and 5.2.

In mixed-model assembly lines, as mentioned, the line balancing is done using total task times rather than individual task times. For the scheduled period time T, the time required to accomplish work element k on all units is given by

$$TT_k = \sum_{j=1}^{m} T_{e_{jk}}$$

which for this example becomes

$$TT_k = 7T_{e_{1k}} + 5T_{e_{2k}}$$

For example, the time required to perform element 6 on all units required for the scheduled period T=60 minutes is $TT_6 = 7 \cdot 2 + 5 \cdot 2 = 24$.

Figure 5.1: Precedence Diagram for Model 1

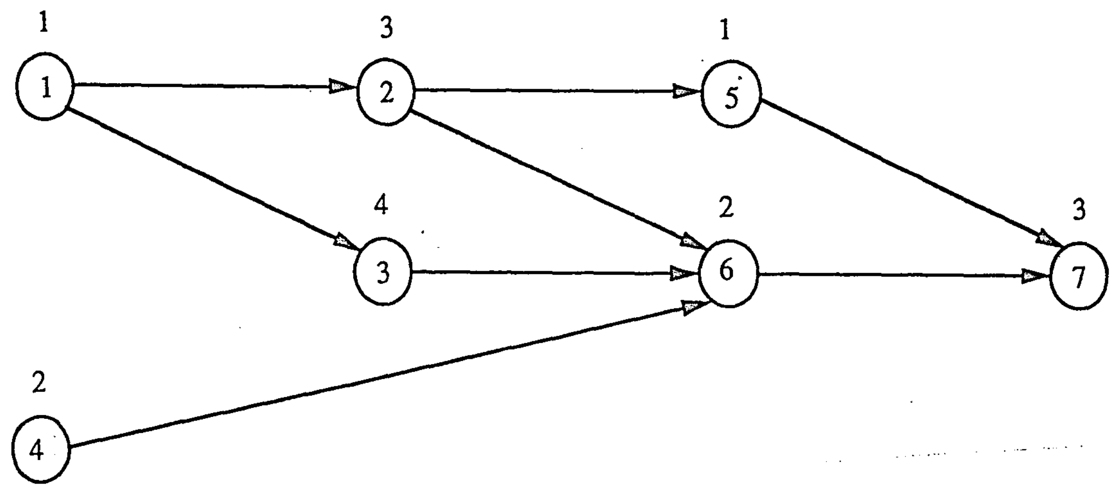


Figure 5.2: Precedence Diagram for Model 2

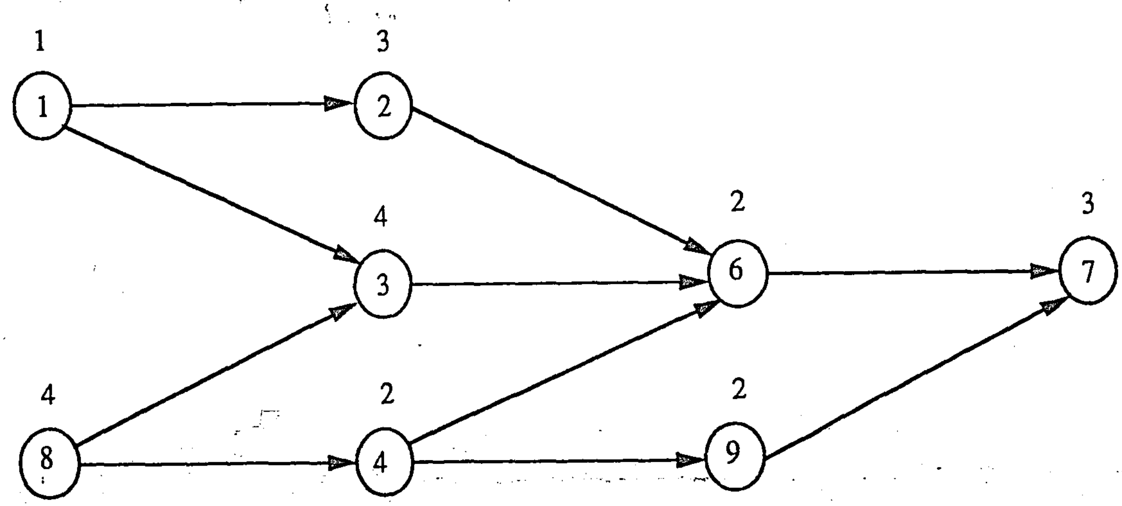
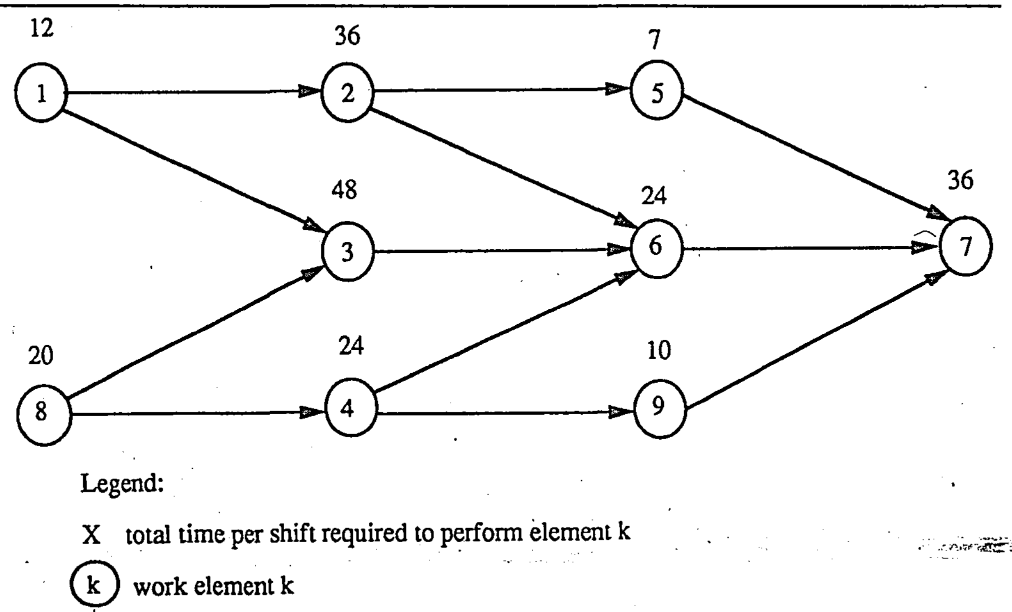


Table 5.3: Element Times per Model and per Shift

	Time/Unit		Time/Shift		
	Model		Model		Total Time
Elements	1	2	1	2	per Shift
1	1	1	7	5	12
2	3	3	21	15	36
3	4	4	28	20	48
4	2	2	14	10	24
5	1	0	7	0	7
6	2	2	14	10	24
7	3	3	21	15	36
8	0	4	0	20	20
9	0	2	0	10	10
Totals	16	21	112	105	217

The precedence diagram (based on total element times) for the model-mix is presented in Figure 5.3.

Figure 5.3: Precedence Diagram for the Model-Mix.



Line Balancing using the Largest Set Rule:

1. For each work element calculate the total task time per shift. They are presented in the last column of Table 5.3.

3. S = Set of all work elements = {1, 2, 3, 4, 5, 6, 7, 8, 9}.

   i = 0.

4. i = i+1 = 1. T_1 = 60.

5. Determination of the weight factors:

Table 5.4: Weight Factor for Elements in S (i = 1)

Elements	w_k
1	12
2	48
3	80
4	44
5	55
6	164
7	217
8	20
9	54

The element with largest weight factor $\leq$ 60 it is element 5. Therefore, 5 and its precedents (1 and 2) are assigned to the first station;

$T_1 - 55 = 5 \Rightarrow T_1 = 5 < 60 \therefore$ a new workstation has to be added to the line.

Station 1:	Element	Time
	1	12
	2	36
	5	7
	TT_{s1} =	55

∘ $i = 2$; $T_2 = 60$; $S = \{3, 4, 6, 7, 8, 9\}$. The calculation of the weight factors in S is:

Table 5.5: Weight Factor for Elements in S (i = 2)

Elements	w_k
1	-
2	-
3	68
4	44
5	-
6	116
7	162
8	20
9	54

By the same procedure as above, the elements assigned to station 2 are {8, 4, 9} $\Rightarrow S = \{3, 6, 7\}$

Station 2:	Element	Time
	8	20
	4	24
	9	10
	TT_{s2} =	54

∘ i = 3 results in

Station 3:	Element	Time
	3	48
	TT_{s3} =	54

Therefore S = {6, 7}.

∘ i = 4

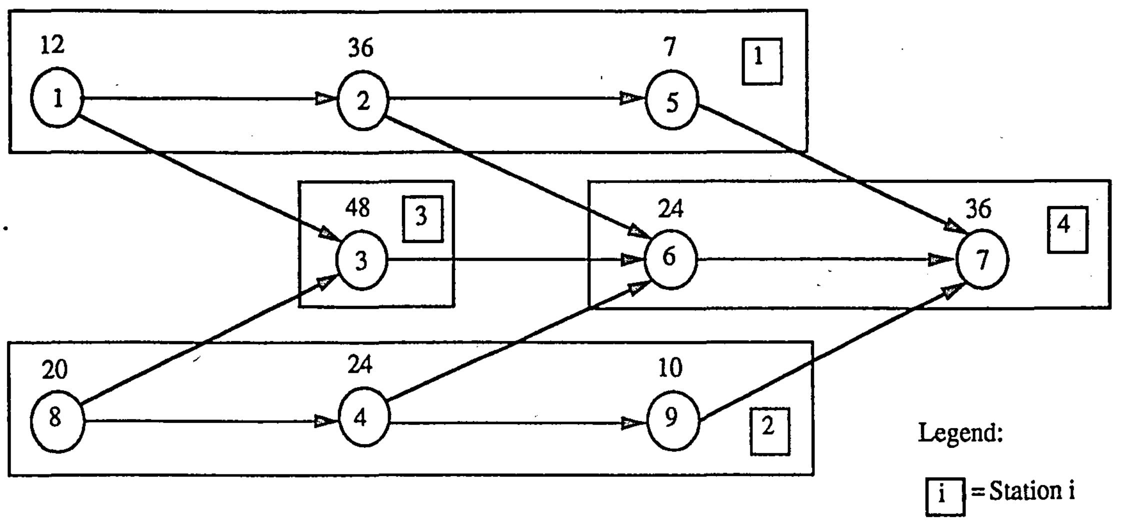
Station 4:	Element	Time
	6	24
	7	36
	TT_{s4} =	54

The balance using the largest set rule resulted in the following assignments of work elements to workstations:

Table 5.6: Line Balancing Solution for the Largest Set Rule (LSR)

Station i	Station time (TT_{si})	$T_{s_{ij}}$ A	$T_{s_{ij}}$ B
1	55	5	4
2	54	2	8
3	48	4	4
4	60	5	5

Figure 5.4: Assignment of Elements to Stations (LSR Method)



The Smoothness Index, as defined in eq. 5.18 is calculated to be:

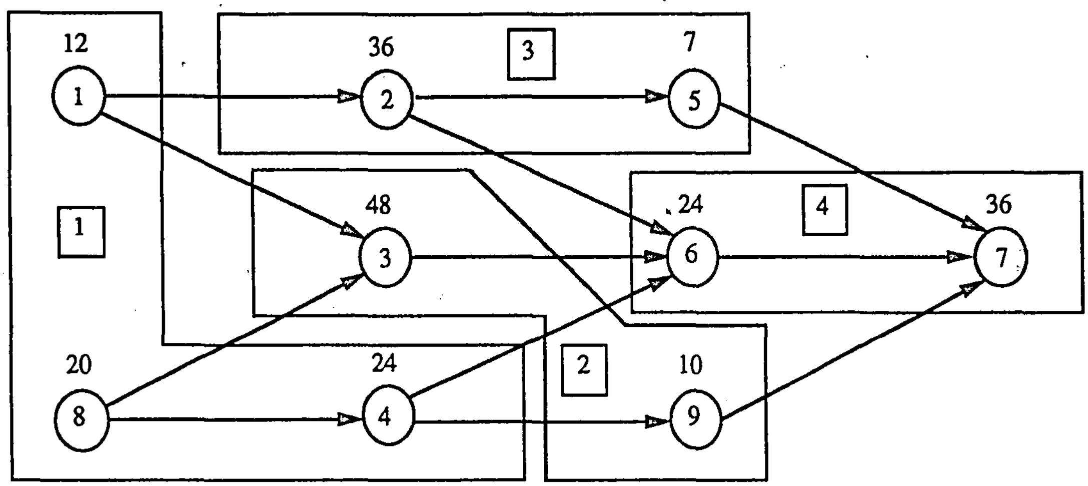
$$S.I. = 14.32$$

The balance solution may have been different if another balancing method was used. For example the R.P.W. would have resulted in:

Table 5.7: Line Balancing Solution for the RPW Method

Station i	Station time (TT_{si})	$T_{s_{ij}}$ A	B
1	56	3	7
2	58	4	6
3	60	5	5
4	43	4	3

Figure 5.5: Assignment of Elements to Stations (RPW Method)



The Smoothness Index for this solution is:

$$S.I. = 17.578.$$

Note that for both solutions the balance delay would be:

$$d = 0.0958, \text{ or equivalently } d = 9.58\%.$$

This could not be otherwise because to minimize idle time is equivalent to one of the following situations: (1) to minimize the number of stations - when the shift time is fixed; (2) to minimize the shift time - when the number of stations is fixed or (3) to minimize the product of the number of stations and the shift time. In this example both the shift time and the number of stations are fixed, therefore the balance delay is the same for the two balancing methods (the resultant total amount of idle time is the same). However, the solution given by the Largest Set Rule seems preferable because the smoothness index is smaller and because tasks belonging to different subassemblies are assigned to different workstations (this should facilitate the delivery of subassemblies, component parts, etc., to the workstations). Note that with the Largest Set Rule tasks 1 and 8, which belong to different sub-assemblies, are assigned to the same station.

5.4. Unavoidable Idle Time at Each Station.

When the balance is not perfect there will be an amount of idle time in each station where TT_{si} is less than T. The idle time during the period T will be:

$$T - TT_{si}, \quad i = 1, 2, \ldots, n \tag{5.19}$$

Normally the station idle time will be zero at the slowest station. The total unavoidable idle time is designated the Balancing Loss.

5.5. Smoothing the Station Assignments.

Usually the line balancing solution provides task assignments to stations that result in different models having different times in each station, even if the station workloads are perfectly balanced. The launch of several units of a given model will result in station idleness. Thomopoulos (Prenting and Thomopoulos 1974) presents a methodology that attempts to smooth the assignments that are achieved by the balancing method. The objective is to equalize the workload for each model over all stations. This results in smoother station assignments, enabling operators to work at a steadier pace and to reduce sensitivity to the sequence (Prenting and Thomopoulos 1974). Let $T_{s_{ij}}$ = the amount of time that a unit of model j is being processed at station i (j = 1, 2,..., m, i = 1, 2,..., n); then the total time during the period T that station i is occupied is given by TT_{si} (defined in section 2.9).

The total time required at station i to produce all Q_j units of model j is

$$Q_j T_{s_{ij}} \tag{5.20}$$

for j = 1, 2,..., m, i = 1, 2,..., n.

The average (or the desired) amount of the total work content for all units of model j is given by

$$\frac{Q_j}{n} \left(\sum_{k=1}^{K_j} T_{e_{jk}}\right) = \frac{TT_{wcj}}{n} \tag{5.21}$$

then the objective of mixed-model line balancing for smoother assignments is to allocate work elements to stations in such a way that precedence constraints are respected, station idle time is minimized and service times are found which minimize

$$\Delta = \sum_{i=1}^{n} \sum_{j=1}^{m} \left| \frac{T_{wc_j}}{n} - Q_j T_{s_{ij}} \right| \tag{5.22}$$

Minimizing Δ tends to smooth or equalize the total work load for each model over all stations. If $\Delta = 0$, then each model will have the same service time at each station.

This is a multi-phase procedure in that there is a first phase where an acceptable assignment of work elements to stations is determined, which is then smoothed in the second phase. This smoothing process is done iteratively. For each step of the iteration Δ is calculated. The iteration process stops when Δ is equal to zero or when a predetermined number of steps has been reached. In the second case the solution will be the one that yields the minimum Δ. In the study performed by Macaskill (1972) it was initially thought that effective measures to smooth the assignments would require a considerable computer effort and an increase in computer time. These measures were not included in his study. The conclusions showed "that unevenness of tasks for a given model tended to reduce overall performance of the line, to increase line length, and to increase sensitivity to the sequence in which the products are launched into the line". However, provided that the product sequence can be controlled, i.e. if the sequence is carefully determined in order to minimize inefficiencies, the effect of uneven assignments appears to be fairly harmless. In another study where simulations for the mixed-model sequencing problem were studied (Macaskill 1973), it was verified that the unevenness in station workloads for a given model did not lessen significantly the performance of the shift performance.

Even if the total work for each model is evenly assigned, if the sequence is not appropriate, inefficiencies such as station idle time will be present.

It was concluded that even though the balance for MMAL using total times will result in uneven station workloads for each model, this method is acceptable for general use and will offer many advantages (such as less calculation effort).

5.6. Concurrent Work.

As mentioned in section 2.15, concurrent work means that operators belonging to adjacent workstations are allowed to work simultaneously on the same unit. Macaskill (1973) argues that concurrent work has great advantages. For example, when concurrent work is not allowed if a worker cannot finish the work on a unit within his station limits he is forced out of the station to complete the work. The worker in the next station cannot start working because the precedent operator is still occupied. Then this operator will start working late and may not be able to finish work within his station limits. This situation is likely to be propagated over the entire assembly line. Allowing for concurrent work will diminish the possibility of situations like the one described here. Therefore, concurrent work is also a potential for decreasing idle time and throughput time. Another benefit of concurrent work is that it reduces the effects of variability in work element times.

5.7. The Multi-Function Worker.

Monden (1983) presents the concept of the multi-function worker, which means that a worker is trained to perform a large scope of tasks. It is believed that through a larger scope of skills, boredom in work is avoided (or at least diminished), motivation is increased and better job satisfaction is achieved. The outcomes are an increase in work efficiency, and therefore an increase in line efficiency. Monden gives an example of an automobile company where workers are trained to operate as many as ten machines. When there is an increase in demand, temporary workers are hired and each worker will handle less than ten machines. This way the utilization of the machine's capacity is fully utilized. These temporary workers need, of course, to be trained. When demand decreases workers are transferred to other sections of the factory or receive new training in other areas. The workers remaining on the production line will handle more machines than before. This enables the factory to be flexible in coping with changes in production. This line of thought can be easily extended to assembly lines. Dar-El and Navidi (1981) described an application of a mixed-model problem where operators are trained to perform a large scope of tasks. For the problem presented there was a workstation (A) were an operator was not 100% occupied. This operator was trained to perform the operations that were done at workstation B. When the worker at station A was not occupied he could help the operator at station B. Workstation B had two operators occupied 57% of the time. Allowing for one worker to travel from his home station to station B reduced the number of workers at this station to one. The remaining operator at station B became 100% occupied, and the work that he could not handle was taken by the worker of station A when the latter was unoccupied.

The main benefit of cross-training workers is that when a worker is not fully utilized at his home station he may travel to another station where there is a worker unable to finish the work on certain units. This enables two workers to be simultaneously working on the same unit (performing different tasks) or on two units.

There is therefore a potential to reduce idle time and consequently to reduce the throughput time.

There are practical restrictions on the distance that a worker may travel away from his home station. Also, workers cannot interfere with each others work. If concurrent work is not well planned, interference between the two operators may happen and the line efficiency will decrease.

To train operators to deal with several tasks has the drawback of increasing the cost of the learning process, but because it has the potential for increasing line efficiency (less idle time, shorter throughput time), it is advisable to use it.

5.8. Parallel Stations.

When the time to accomplish task k on all Q units in shift T (TT_k) is smaller than the shift time, it may be necessary to have parallel stations. Having stations in parallel can multiply the rate of production (e.g. for two stations in parallel, the station cycle time is divided by two and the production rate is multiplied by two).

Buxey et al. (1973) state that in nonmechanical unpaced lines "the use of two identical stations per stage has inherent operating advantages" and that for Products Fixed items in continuous flow-lines the "decoupling of stations is effected by manipulation of conveyor speed and item spacing". A study performed by Wild and Slack (reported in Buxey et al. 1973) indicates that the use of two identical stations per stage has great advantages when the number of stations in the line is large, the buffer storage capacity is low and operator variability is high. These results can be extrapolated to nonmechanical unpaced assembly lines. Sometimes the flow of materials required by parallel stations may prevent their use.

Reiter (1966 - also reported in Buxey et al. 1973) has shown that line balancing can be improved when concurrent work is allowed. Concurrent work can be seen as a sort of paralleling.

5.9. Lines with Several Labor Groups.

To find the solution for the line balancing problem when the assembly line includes several disjointed but related areas or labor groups, each area is considered independently of the others. However, when the models are launched into the line each model will flow from one area to the other. Therefore the sequencing of products must be the same for all areas.

5.10. New Production Runs.

For a change in product-mix, Macaskill (1972) indicates that if the second product-mix does not differ too much from the first, it will often be acceptable to retain the first assignment for the second mix and adjust the production requirement in a way that no station is overloaded. If the second product-mix is very different from the first one, then there is no other choice than to rebalance the line. This may include costly aspects such as hiring and firing workers or another type of solution. The system should have enough built-in flexibility to cope with changes in the product-mix and/or the production level. Operators should be prepared to work longer or shorter periods (including overtime and undertime), and tools should be flexible enough so that each time the line is rebalanced they can be redistributed.

It is very common in the automobile industry to rebalance lines to achieve different output rates (Buffa 1980). Changes in the output will induce changes in the theoretical cycle time. In a continuous flow line a change in the output level can be handled with a change in the conveyor speed. In general, if the new production does not require radical changes in the line, the operators at the workstations will adapt to the new situation. For example, Buxey et al. (1973) states that operators on a manual assembly line will respond to a reduction in the feed interval by reducing their service time.

It is possible that the required production may not be achieved during the shift time and that a residue of r units is left to be completed in the next shift, being the r^{th} unit at the first station on the line. The work on the next shift will include the completion of the residual work plus the required production for that shift. Macaskill (1973) in a computer simulation for mixed-model lines, assumed for balance purposes that the residual work between shifts was approximately the same.

We propose the following to balance a line considering the possibility of having residual units: (1) if r and its distribution can be estimated the balance will be done the same way as for no residuals, but the work to complete the r units will be added to the work required to complete the shift production; (2) if r cannot be estimated then the balance is done normally and the units are sequenced and the residual units are reported. The residual units and their distribution among the line would be used in a new balance and a new sequencing solution would be generated. The residual units are again reported and a new balance is done. This would be an iterative process that would stop when the solution reached a satisfactory level or when a predetermined number of steps was done.

Measures to lessen the possibility of having residual units left at the end of a shift are taken in the sequencing procedure, but sometimes it may be impossible to complete all units within a shift.

5.11. Variability of Work Element Times.

When assembly work is performed by human operators it is almost impossible to have deterministic task times. Previous research (Wild 1971) showed that typically the relationship between output and time of work is a low output in the beginning of the work period, then increasing and finally decreasing towards the end of the shift. Hicks and Young (1962) concluded that task times are randomly distributed and can be approximated by the normal distribution.

Moodie and Young (1965) studied the line balance of a single-model line where task times are assumed normally distributed. The line balancing objective was changed to account for the variability of task times. The station time now should introduce the variability of task times. Therefore,

$$TT_{si} = \sum_{k \in i} \overline{TT}_k + z\sqrt{\sum_{k \in i} \sigma_k^2} \tag{5.23}$$

where $\overline{TT}_k$ is the mean of the observed times when performing task k in Q units(it is the mean time to perform task k on all Q units); $\sum_{k \in i} \overline{TT}_k$ is the sum of the mean times for the work elements allocated to station i; σ_k^2 is the variance of the observed task k times. Hence $\sqrt{\sum_{k \in i} \sigma_k^2}$ is the standard deviation for station i time; z is a constant value that represents the probability of TT_{si} exceeding the shift time T. The values for z can be obtained from a table of the normal distribution. For example, if it is required that in 97.5% of the occasions the workload assigned to the stations is less than the shift time, a factor z = 1.96 is required. This would mean that $\sum_{k \in i} \overline{TT}_k \leq T - 1.96\sqrt{\sum_{k \in i} \sigma_k^2}$, for each station i. Note that $z = \frac{(TT_{si} - T)}{\sigma_i}$ is the transformation parameter of a normal distribution (TT_{si}, σ_i) into the standard normal distribution (0,1).

Modifying the line balancing objective function to account for task time variability will result in

$$Min \sum_{i=1}^{n} \left[T - \sum_{k \in i} \overline{TT}_k - z \sqrt{\sum_{k \in i} \sigma_k^2)} \right] \tag{5.24}$$

We suggest that an heuristic method be used to solve the line balancing problem. The methodology would be the same as for the deterministic case, but instead of using deterministic task times, stochastic times would be used.

Hoffman (1990) in a study on single-model assembly line balancing, suggests that the use of a safety slack (i.e. the use of a cycle time greater than the largest sum of task times at any one station), may be used as a method to deal with variable task times thereby making difficult balances easier to solve. Using a safety slack will result in station lengths slightly larger than the length required for deterministic times.

Arcus (1966) states that if the variance of task times about the mean becomes greater the consequence will be that the worker tends to move out of the station further upstream and downstream, the distances out of the station becoming greater.

Macaskill (1973) recommends that when task time variability is small "a deterministic situation will provide a satisfactory tool for predicting the performance of the assembly line".

A very common procedure is to do a "notational" line balancing (Buxey et al. 1973) which means that the line is balanced considering only the mean station times.

Another assumption about work element times is that they are additive (Groover

1987). However, an operator may save same time if given work element is performed after another specific one. If this happens the time to perform both tasks may be less than the sum of both task times taken individually.

5.12. Measures to Reduce the Effect of Task Time Variability.

Buffer storage may be useful to reduce the effect of variability in task time. Macaskill (1973) concludes that concurrent work will tend to reduce task time variability. Allowing for a certain amount of idle time will also lessen the effect of task time variability. It will be better to provide the line with an amount of slack time in order to provide more flexibility for the workers. If there is not a small amount of idle time and, a worker is having some problems in accomplishing a certain task, the unit may leave the station incomplete. If there is slack time, the worker can use this amount of extra time and the possibility of having incomplete units becomes smaller.

5.13. Comments.

Line balancing can also be applied to automated assembly lines or mechanized fabrication lines. However, it is generally more difficult to subdivide operation times on mechanized processes than in manual production tasks. This makes line balancing for such cases very difficult (even impossible) to achieve. The result will tend to be poor equipment utilization and relatively high costs. In assembly lines, where the work is more likely to be manual, balance is easier to achieve because the work to be performed may be divided into small parts and distributed over several workstations. Because there is usually very little equipment at each station, utilization of equipment may not be of great importance (Buffa 1990). This is not the case for an automated line where work is performed by machines and usually maximum machine utilization is presumed.

The results of previous studies appear to reinforce the opinion that it is acceptable to balance the assembly line with heuristic methods that yield fairly good balances even if there will be unevenness in the station workloads for a given model. In cases where task times vary significantly, such as in manual assembly lines, the use of algorithms that yield optimal solutions (for deterministic task times) may not perform the same way, and can in fact be far from the optimum when variability in task times is present. Furthermore, the complexity of some problems (great number of work elements and of possible models), inhibits the use of such algorithms because of the computer effort required. The speed of calculation and computer storage requirement were aspects of great concern in the 1970's. Due to advances in computer technology these aspects do not seem so relevant as they were then. However large problems still require a considerable amount of effort and may be a reason why heuristic methods seem preferable. The variability in task times introduces a new complication to the problem and gives support for the choice of heuristic methods rather than optimal solution algorithms. Heuristics have proven to achieve near optimal results and are less costly to use. With the trend in computer technology, it is expected that in a near future the use of sophisticated methods will not pose major problems.

The sequencing of products into the line is addressed in the next chapter.

6. The Sequencing Problem

The different models to be produced on a mixed-model assembly line typically require different amounts of work at the different workstations. If several units of a model with high work content are successively launched into the line this will result in overloading of stations during the period that the line is producing these units. If after that several units of a model with low work content are successively launched, the stations will be under-utilized. Those two situations are not desirable because a situation of overload can prevent the worker from completing the work on the units and under-utilization means worker idle time.

It is preferable to feed the line with a sequence of products that will result in a smoothed production, which means that the stations will not be over or underloaded for significant amounts of time, and the interval between successive units coming off the line is approximately constant. An inadequate sequence will tend to increase the idle time of the line; it will add an extra amount of idle time to the balancing loss due to imperfect balancing, increase the effect of unevenness of workloads, and decrease line performance. Previous research (Prenting and Thomopoulos 1974, Dar-El 1978, Macaskill 1973, Bard et al. 1992) demonstrated that an inappropriate sequence will result in a significant reduction in the throughput and increase the assembly line length (resulting in an increase in the production cost).

Suppose that all models have exactly the same station service times. In such a situation the line would be essentially working as a single-model line; the time between successive units coming off the line would be determined by the slowest station and the order in which the units are launched into the line makes no difference. Because each model spends at each station exactly the same amount of time than the other models, the order in which models are fed into the line does not increase the operator idle time. In such lines the main concern is to minimize the idle time due to imperfect balance, and therefore to achieve the best possible line balance.

However, in mixed-model lines, different products have different station service times. In this case, the sequence of models onto the line becomes an issue.

For Q units ($Q = Q_1 + Q_2 + Q_3 + \ldots$) there are $\frac{Q!}{Q_1!\,Q_2!\,Q_3!\ldots}$ possible sequences. For a given line balancing solution, at least one of these sequences is the optimal one. However, it is not practical to search all possible sequences and choose the best one. This would require a considerable calculation effort even for small problems. It is very common to determine the sequence through the use of heuristic methods. Previous research in the mixed-model assembly line sequencing problem (Bolat 1988), suggested that the use of techniques that search for the optimal solution (such as branch and bound algorithms, for example) become inefficient for large problems. The number of units to be scheduled and the restrictions on the problem become such, that these algorithms cannot solve the problem. Similar to line balancing of mixed-model assembly lines, it seems that heuristics are the best option to solve the sequencing problem. Heuristics not only can achieve good solutions, but they are also less complicated than algorithms that search for the optimum, and the cost of using heuristics is likely to be smaller. This chapter is concerned with the use of an acceptable[1] method for determining the sequence in which units are launched into the line. Deterministic assembly times are assumed.

6.1. An Heuristic Method to Solve the Sequencing Problem.

The proposed method employs the same procedure for selecting the units to be launched into the line as the Goal-Chasing method (Monden 1983). The objective is to maintain, for each station i and for each launch, a workload as close as possible to the average station time per unit. Open stations enable the line to become more flexible and can result in a decrease in throughput time and rate of incomplete items. If a unit launched requires more work at a station than station passage time allows, the operator is forced out of his station to complete the work. If the station has a closed boundary, the unit will leave the station uncompleted. If units requiring more work than the station passage time allows are launched successively, the operator will be constantly travelling past the downstream limit and may reach a point where he cannot catch up to arriving jobs. On the other hand, if jobs requiring less work than what is allowed by the station passage time are launched successively, the operator will be idle for the case of closed stations or will have to travel past his upstream station limit in order to start work on the next product. This latter situation may be undesirable because it can cause interference between adjacent workers.

[1]In terms of the quality of the obtained solution and expended calculation effort.

Buxey et al. (1973) state that ideally a model that overloads a station should be followed by one which allows for slack time in that station. This should allow the operator to finish the model with great workload and still have time to complete the next model within his station limits. The method proposed in this thesis is an attempt to do exactly that, i.e. if a unit requires a great amount of work, the next unit to be launched will require a lighter workload and vice-versa (lighter workloads will trigger the launching of units that require heavier workloads). Such method has the result that the operator will be working on each unit an amount of time that is close to an average time per unit. This average time per unit is the station average time, given by:

$$\overline{T}_{si} = \frac{TT_{si}}{Q} \tag{6.1}$$

The reasoning behind this is to have the operators working virtually the same time on each unit launched. This will smooth the workload for each operator, although it may not smooth the workload across all workstations, since this is the problem in line balancing. The method should therefore be applied after the line has been balanced and the obtained balance solution is acceptable, which means that the workload is approximately leveled across all stations.

The basic approach in the method is that the unit that should be launched at the l^{th} position in the sequence is the one that results in the time to produce l units equal to the total station average time to produce the l units. Defining $X_{i,l}$ as the time required at station i to complete the work on l units and $\frac{l \cdot TT_{si}}{Q} = l \cdot \overline{T}_{si}$ as the average time required to complete the assembly work on the l units, the unit launched at position l will be the one that minimizes the difference between $X_{i,l}$ and $l \cdot \overline{T}_{si}$.

In an assembly line there are several workstations; the unit chosen to be launched at position l is the one that results in the minimum difference across all stations (the differences between $X_{i,l}$ and $l \cdot \overline{T}_{si}$ for i = 1 to n are added and the unit that results in the least value for this sum is the one chosen to be launched. The difference between the two mentioned values is squared in order to avoid the inclusion of negative terms in the summation.

6.1.1. Mathematical Model.

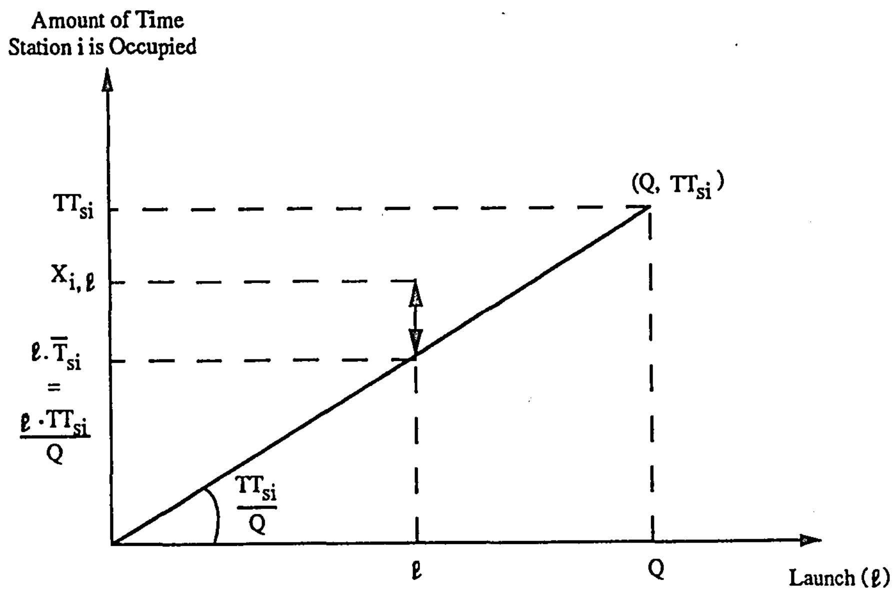
Consider a mixed-model assembly line where a number of different models are to be produced. Let m = number of different models to be produced, where $m = \sum_j$, j = 1, 2, 3,..., m (Q_j is the quantity of model j to be produced); Q = total production requirement $\left(Q = \sum_{j=1}^{m} Q_j\right)$; TT_{si} = amount of time necessary to complete all units to be produced at station i (this value is given by the balance solution); $X_{i,l}$ = amount of time necessary at station i to complete the first l units.

Then TT_{si}/Q = average time available at station i to work on each unit launched; and $\dfrac{l \cdot TT_{si}}{Q} = l \cdot \overline{T}_{si}$ = average time to produce l units at station i.

To keep the average workload at station i as constant as possible it is desirable for $X_{i,l}$ to be as close as possible to the value of $l \cdot \overline{T}_{si}$. The sequence to be chosen is the one that for each launch results in the closest station time to the average station time for each station.

Consider Figure 6.1.

Figure 6.1: Relation Between $X_{i,l}$ and $l \cdot \overline{T}_{si}$



For the points P_l and Q_l, where $P_l = (l \cdot TT_{s1}/Q, l \cdot TT_{s2}/Q, \ldots, l \cdot TT_{sn}/Q)$, (i = 1, 2, ..., n) and $Q_l = (X_{1,l}, X_{2,l}, \ldots, X_{n,l})$, if the goal is to have the workload in each station as constant as possible, for each launch l the point Q_l must be as close as possible to P_l. The distance between two given points is given by

$$D_l = \|P_l - Q_l\| = \sqrt{\sum_{i=1}^{n} \left(\frac{l \cdot TT_{si}}{Q} - X_{i,l} \right)^2} \qquad (6.2)$$

The model chosen to be launched as the l^{th} launch will be the one that minimizes

D_l, or equivalently, the model launched at the l^{th} position in the sequence is the one that

$$\underset{j \in S}{minimizes} \sum_{i=1}^{n} \left(\frac{l \cdot TT_{si}}{Q} - X_{i,l-1} - T_{s_{ij}} \right)^2 = \underset{j \in S}{minimizes} \sum_{i=1}^{n} \left(l \cdot \overline{T}_{si} - X_{i,l-1} - T_{s_{ij}} \right)^2 \quad \textbf{(6.3)}$$

where $T_{s_{ij}}$ = service time (station time) for model j at station i; $X_{i,l-1}$ = necessary assembly time to perform l-1 units at station i; and S = the set of models remaining to be launched. $S = \{S_1, S_2, \ldots, S_j\}$; S_j is the set of units of model j remaining to be assembled.

6.1.2. The Sequencing Algorithm.

The sequencing algorithm can be outlined in the following steps:

1. For l=1, $X_{i,l-1} = 0$ (i = 1, 2,..., n), S = Q.

2. The model chosen to be launched in position l is the one which

$$\underset{j \in S}{minimizes} \sum_{i=1}^{n} \left(l \cdot \overline{T}_{si} - X_{i,l-1} - T_{s_{ij}} \right)^2 \quad \textbf{(6.4)}$$

Remove the unit assigned from S_j: $S_j = S_j - \{1\}$

•Is $S_j = \varnothing$?

• If $S_j = \varnothing$ remove S_j from S: $S = S - S_j$.

•• is $S = \varnothing$?

- Yes, stop. All the units are assigned to a position in the sequence.

- No, go to step 3.

• If $S_j \neq \varnothing$, go to step 3.

3. Calculate $X_{i,l} = X_{i,l-1} + T_{s_{ij}}$, set l=l+1 and go to step 2.

Suppose that at the l^{th} launch, two models have the same value for Eq. 6.3, and therefore, both are candidates to be launched. Suppose also that these are model 1 and 2. Eq. 6.4 will be calculated twice for each model remaining to be launched at position $l+1$, once for model 1 as the l^{th} launch and once for model 2 as the l^{th} launch. The model chosen to be launched in position $l+1$ is the one that results in the least value for eq. 6.4, and its corresponding model (1 or 2) is launched at position l. If for the second level there is a tie again instead of going one level deeper it is suggested that the model to be launched at the l^{th} position is the one that will enable a greater variety in the production (e.g. if there is a consecutive tie between models 1 and 2 at the l^{th} $(l+1)^{th}$ positions, instead of repeating the calculations for one level deeper, if the model launched at l-1 was of type 1 then, at the l^{th} position should be launched a model 2 and vice-versa).

6.1.3. Example.

The line balancing of the problem presented in section 5.3.1.1 resulted in the following station times:

Table 6.1: Workload Distribution

Station i	Station Time (min)
1	55
2	54
3	48
4	60

Table 6.2: Service Times

Model	$T_{s_{ij}}$				Demand
j	1	2	3	4	Q_j
1	5	2	4	5	7
2	4	8	4	5	5

There are 12 units to be launched. Therefore, for $l = 1$

Model 1:

$$\left(1\cdot\frac{55}{12}-0-5\right)^2+\left(1\cdot\frac{54}{12}-0-2\right)^2+\left(1\cdot\frac{48}{12}-0-4\right)^2+\left(1\cdot\frac{60}{12}-0-5\right)^2 = 6.4$$

Model 2:

$$\left(1\cdot\frac{55}{12}-0-4\right)^2+\left(1\cdot\frac{54}{12}-0-8\right)^2+\left(1\cdot\frac{48}{12}-0-4\right)^2+\left(1\cdot\frac{60}{12}-0-5\right)^2 = 12.5$$

Therefore the model which minimizes eq. 6.4 is model 1. The computation of $X_{i,1}$ gave

the following results:

$X_{1,1} = 5$, $X_{2,1} = 2$, $X_{3,1} = 4$, $X_{4,1} = 5$

For the 2nd launch (l=2) eq. 6.4 is calculated again for each model:

Model 1:

$$\left(2\cdot\frac{55}{12}-5-5\right)^2+\left(2\cdot\frac{54}{12}-2-2\right)^2+\left(2\cdot\frac{48}{12}-4-4\right)^2+\left(2\cdot\frac{60}{12}-5-5\right)^2 = 25.9$$

Model 2:

$$\left(2\cdot\frac{55}{12}-5-4\right)^2+\left(2\cdot\frac{54}{12}-2-8\right)^2+\left(2\cdot\frac{48}{12}-4-4\right)^2+\left(2\cdot\frac{60}{12}-5-5\right)^2 = 1.03$$

Thus, model 2 is launched at the second position in the sequence. The time that each station is occupied producing the first two units is

$$X_{1,2} = 9,\ X_{2,2} = 10,\ X_{3,2} = 8,\ X_{4,2} = 10$$

The same calculations are repeated until there is only one type of model left to be scheduled.

The sequencing algorithm was applied to 12 launches which resulted in the following:

Table 6.3: Resultant Sequence for the Example in Section 5.3.2.

Launch (l)	Model	$X_{1,l}$	$X_{2,l}$	$X_{3,l}$	$X_{4,l}$
1	1	5	2	4	5
2	2	9	10	8	10
3	1	14	12	12	15
4	2	18	20	16	20
5	1	23	22	20	25
6	2	27	30	24	30
7	1	32	32	28	35
8	1	37	34	32	40
9	2	41	42	36	45
10	1	46	44	40	50
11	2	50	52	44	55
12	1	55	54	48	60

There was a tie at l=6; thus it made no difference whether model 1 or 2 was launched in that position. For launch l=7 there was again a tie for the sequences 12 and 21 - to launch 1 at l=6 and 2 at l=7 was the same that launch 2 at l=6 and 1 at l=7. A third level evaluation resulted in a tie between the sequences 121 and 211. Instead of going to the fourth level, for position l=6 model 2 was chosen because the previous had been a model 1.

The variation of station times in relation to the average station time can be seen graphically.

Figure 6.2: Distances Between $X_{1,l}$ and $l \cdot \overline{T}_{s1}$ for l Varying from 1 to Q

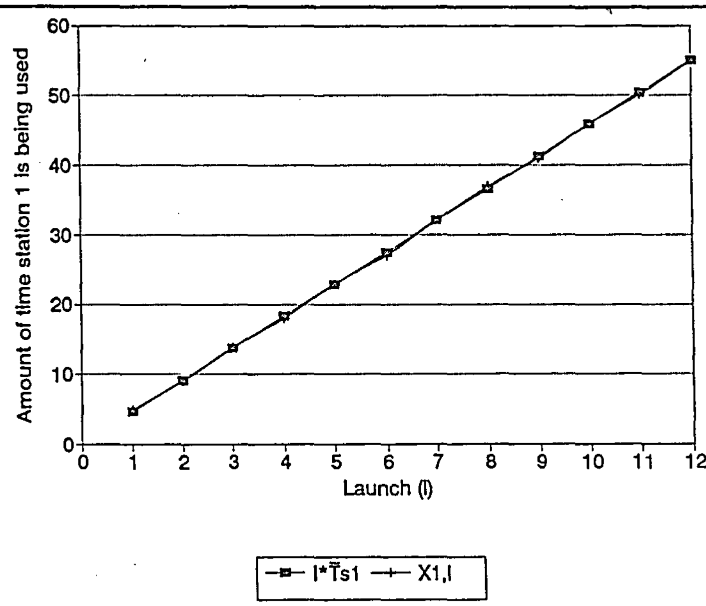


Figure 6.3: Distances Between $X_{2,l}$ and $l \cdot \overline{T}_{s2}$ for l Varying from 1 to Q

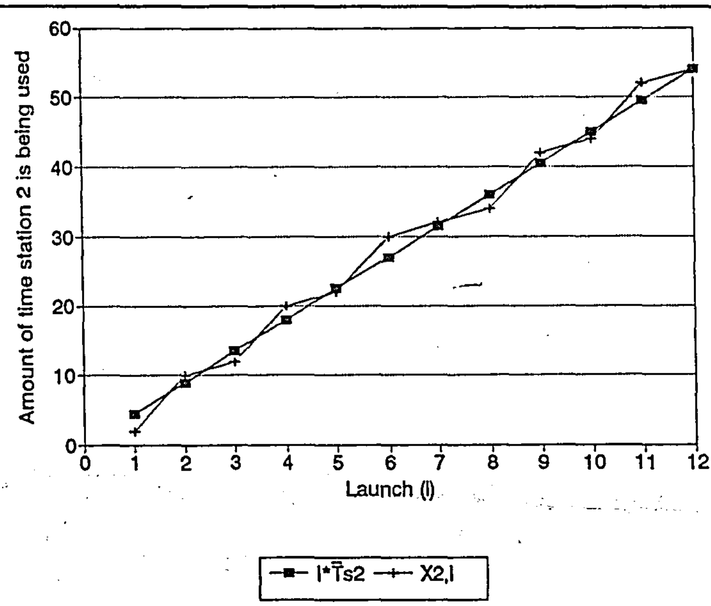


Figure 6.4: Distances Between $X_{3,l}$ and $l \cdot \overline{T}_{s3}$ for l Varying from 1 to Q

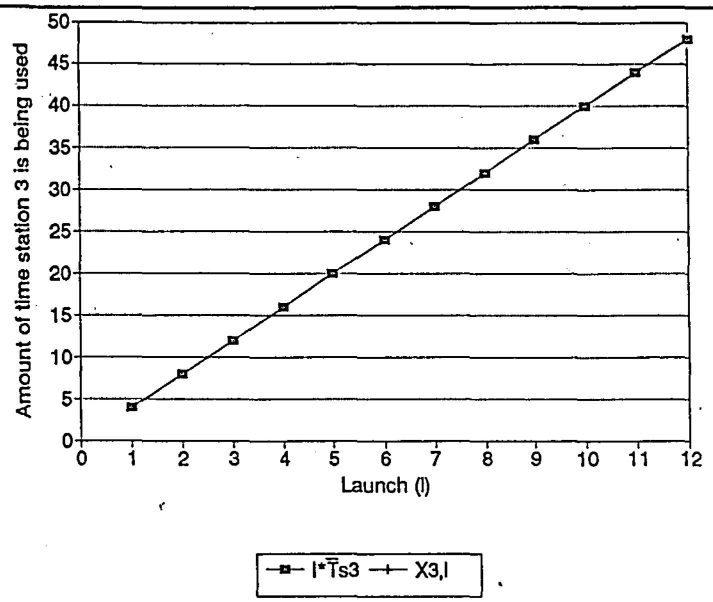
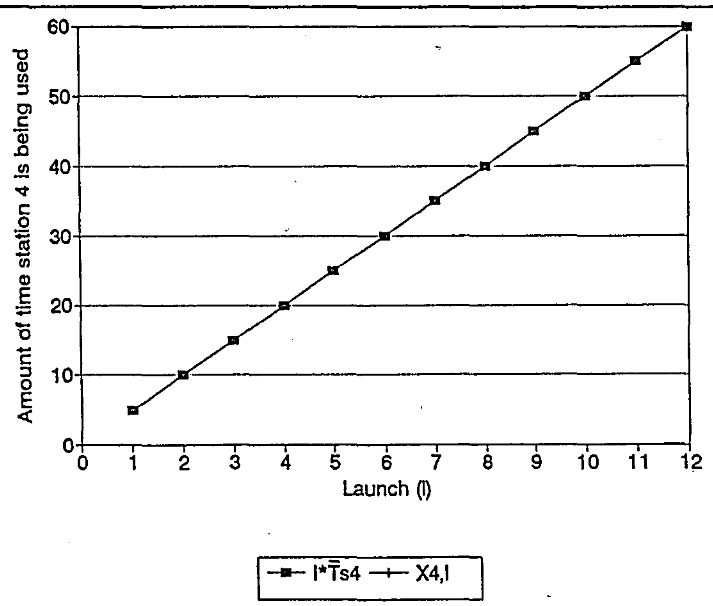


Figure 6.5: Distances Between $X_{4,l}$ and $l \cdot \overline{T}_{s4}$ for l Varying from 1 to Q



6.2. Comments.

Sumichrast et al. (1992) proposed a similar heuristic for solving the sequencing problem, which they called the Time Spread method. The difference with the method proposed here is that instead of using average station times per unit, the proportion of each station time to the total work content $\left(\frac{l \cdot TT_{si}}{TT_{wc}}\right)$ is used. The goal of the Time Spread is to level the station work loads. Compared with other heuristics for mixed-model in just-in-time production systems, namely the Goal-Chasing methods I and II, Miltenburg (1989) algorithm 3 using heuristic 2 and the batch-model sequencing, the Time Spread (TS) performs very well, being surpassed only by Miltenburg's algorithm for the cases when the products had a complex structure and uniform parts usage was the objective. The authors state that "if assembly efficiency, including product quality and worker flexibility, is the objective, then the Time Spread method seems to be slightly preferable (in relation to the other methods tested)". The Time Spread method, on the overall, resulted in less idle time, less number of incomplete units and in the highest percentage of time that workers are at their home station.

Neither uniform parts usage nor considerations about product structure are contemplated in a production system where the assembly line is the only point of concern (this research is focused on the assembly line alone, i.e not being a part of a multi-level production system) and, therefore, the Time Spread method seems to be a good sequencing procedure for mixed-model assembly lines.

For the example, in section 5.3.2 (the 2 model, 4 station problem), the method proposed in section 6.1 performed better than the Time Spread Method. The latter yielded a solution that was not the most efficient, resulting in launching of units into the line in a batch mode. The TS method was also compared with the proposed method for other examples (see chapter 8) and the proposed method repeatedly gave better performance.

Dar-El and Cother (1975) argue that to choose arbitrary station limits and determine the sequence that minimizes inefficiencies for these limits is not a good procedure because, depending on the station limits, a given sequence can become efficient or inefficient. Furthermore they argue that physical station limits and the extent of operator movements may be difficult to define and this makes the determination of station limits before knowing the sequence not very realistic. They proposed a method that minimizes the overall line length when no operator idle time, work congestion or work deficiency are allowed.

The sequencing procedure presented in 6.1 is not dependent on station limits and therefore does not attempt to minimize the inefficiencies (defined in chapter 3) that depend on station limits. Once the sequence is found it is possible to determine the station lengths that the sequence will impose (remember that this research is focused on continuous moving conveyor lines, where operators must move with the line and therefore the station dimensions are important; for stationary systems or lines with buffer storage allowed, the station limits are not so relevant and are determined by restrictions such as the reach of tools, dimension of the models to be assembled, etc.).

The procedure should be: (1) calculate the sequence by the proposed method and (2) for that sequence determine the station lengths. This will result in a line where neither utility work, work congestion or work deficiency exist. Of course, if station limits are imposed, these inefficiencies will most probably happen and, as with the other methods (Prenting and Thomopoulos 1974, Dar-El and Cother 1975, Macaskill 1973), the required output may not be achieved. However, if the product-mix and the required output for the new production do not require lengths too different from the existing ones, the line can work quite efficiently. If this is not the case, then the line would not be able to achieve satisfactory results. To change the time that a unit is available to an operator, the launching interval can always be increased or decreased (for example, Buxey et al. (1973) indicate that operators will respond to a decrease in the launching interval by speeding up their work); also the conveyor speed can be changed to create different station passage times. If none of these things work, then it is probably because the new production run requires station lengths that are substantially different from the existing ones.

6.3. Calculation of the Optimal Fixed Launching Rate.

In order to calculate the optimal launching rate the methodology presented by Kilbridge and Wester (1963) is extended to deal with the general case which is each model having different service times at different workstations (technological constraints, such as zoning constraints, imposed on the problem will result in a situation such as this, even if the total station workloads are perfectly balanced).

For the sequencing problem, Kilbridge and Wester considered two objectives: (1) to avoid station idleness and, (2) to avoid work congestion. Avoiding idle time will assure that stations are always kept busy. Work congestion, as previously mentioned, happens when the operator is forced to walk past his station downstream limit in order to complete the assembly work on the unit. If several units with high work loads are successively launched into the line, the operator will be constantly working out of his station limits and may not be able to catch up. A situation such as this can cause interference between adjacent workers or the unit can proceed down the line incomplete. Overlapping of stations is not allowed.

To develop a model that will permit observation, an arbitrary launching of 3 different models is illustrated in Figure 6.6. Although this example is specific, it permits general conclusions to be drawn. The data for this case is the following:

	$T_{s_{ij}}$		
Station	1	2	3
Model			
1	3	2	1
2	3	3	2
1	2	1	1

Suppose that Q units are to be launched.

6.3.1. Objective I. Avoid station idle time.

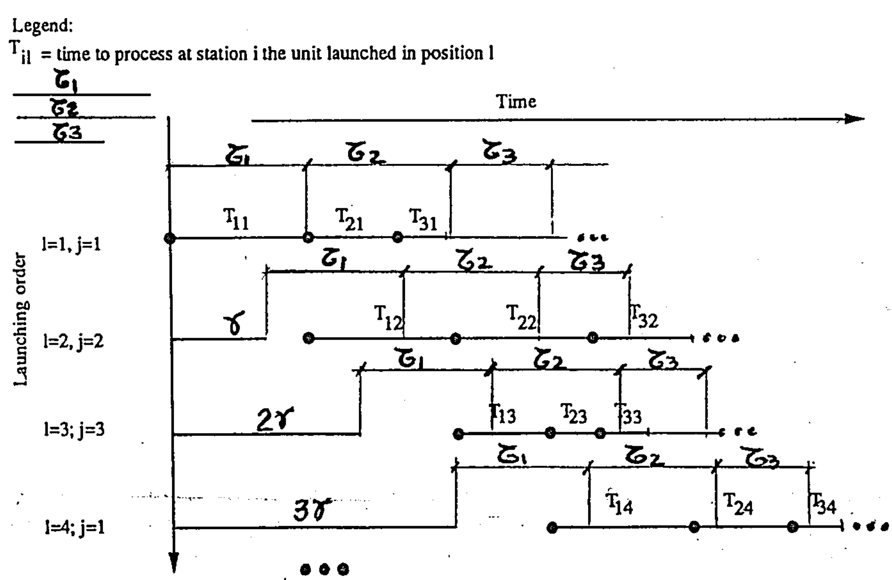
Consider the situation of the Fig. 6.6. τ_i (i = 1, 2,..., n) is the station i passage time. It seems evident, if work congestion is to be avoided, that the station passage time has to be such that all units can be performed within the stations limits. Therefore the station passage time for station i, as defined by Kilbridge and Wester, should be at least equal to the maximum time that station i is working on a given model. τ_i is also the time separating the start of work on a given unit by station i and station i+1 (because overlapping is not allowed). The amount of time required at station i to perform model j (the station service time) is given by $T_{s_{ij}}$.

Hence,

$$\tau_i = \max\left(T_{s_{ij}}\right), \qquad j=1,2,\ldots,m \tag{6.5}$$

γ is the time between two consecutive launches. Let T_{il} = be the required time to assemble at station i the unit launched in position l.

Figure 6.6: Arbitrary Launching of 3 Models



In order to avoid station idle time at station 1, the launch of the second unit has to be done according to

$$\gamma \le T_{11} \tag{6.6}$$

By the same reasoning, the third and following launches have to be such as

$$2\gamma \le T_{11} + T_{12} \tag{6.7}$$

$$3\gamma \le T_{11} + T_{12} + T_{13} \tag{6.8}$$

....

$$Q\gamma \le T_{11} + T_{12} + T_{13} + \ldots + T_{1Q} \tag{6.9}$$

Therefore for Q units to be assembled the fixed launching rate γ has to be

$$\gamma \le \frac{T_{11} + T_{12} + \ldots + T_1 Q}{Q} \tag{6.10}$$

If idle time is to be avoided at station 1, Eq. 6.10 is equivalent to

$$\gamma \le \frac{TT_{s1}}{Q} \tag{6.11}$$

where TT_{s1} is the total assembly time required at station 1 to perform all units (this value is known from the line balance solution).

If $\gamma < TT_{s1}/Q$ the units will be launched too soon and it will cause congestion at station 1. Therefore the optimal value for γ is

$$\gamma = \frac{TT_{s1}}{Q} \tag{6.12}$$

To avoid idle time in station 2 it will be necessary to verify certain conditions.

Let $T_{11} - \gamma = \delta_1$. For the second launch it is necessary that

$$\gamma + \delta_1 + T_{12} \leq T_{11} + T_{21} = \gamma + T_{11} - \gamma + T_{12} \leq T_{11} + T_{21}$$

which is equivalent to

$$T_{12} \leq T_{21} \tag{6.13}$$

$2\gamma + \delta_2 = T_{11} + T_{12}$. For the third launch (l=3)

$$2\gamma + \delta_2 + T_{13} \leq T_{11} + T_{21} + T_{22} =$$

$$= T_{11} + T_{12} + T_{13} \leq T_{11} + T_{21} + T_{22} =$$

$$= T_{12} + T_{13} \leq T_{21} + T_{22} \tag{6.14}$$

....

For the l^{th} launch ($l = 1, 2,\ldots, Q$) it can be shown that to avoid idle time at station 2,

$$\sum_{h=2}^{l} T_{1h} \leq \sum_{h=1}^{l-1} T_{2h} \tag{6.15}$$

Generalizing this reasoning for a line with n stations (i = 1,..., n) it can be shown that to avoid idle time, each launch l ($l = 1, 2,\ldots, Q$) has to satisfy the following:

$$l\gamma = \sum_{h=1}^{l} T_{sh} \text{ , for i = 1} \tag{6.16}$$

$$\sum_{h=2}^{l} T_{i-1h} \leq \sum_{h=1}^{l-1} T_{ih} \text{ , for i = 2,..., n} \tag{6.16}$$

For example, if idle time is to be avoided in the third station, the fourth launch has to be such as $T_{22} + T_{23} + T_{24} \leq T_{31} + T_{32} + T_{33}$.

From the above, it can be seen that for Q launches, $\gamma = \frac{TT_{s1}}{Q}$ satisfies Eq. 6.16 and idle time at station 1 is avoided. On the other stations, idle time is not a function of γ. Then, the launching interval is only dependent on TT_{s1} and Q . Therefore, γ is the average time per unit produced only for the first station.

6.3.2. Objective II. Avoid Work Congestion.

To avoid work congestion at every station a set of equations has to be derived for each station.

<u>Station 1</u>

l=1: $\tau_1 \geq T_{11}$

l=2: $\gamma + \tau_1 \geq T_{11} + T_{12}$

l=3: $2\gamma + \tau_1 \geq T_{11} + T_{12} + T_{13}$

.
.
.

l=Q: $(Q\text{-}1)\gamma + \tau_1 \geq T_{11} + T_{12} + T_{13} + \ldots + T_{1Q}$ **(6.18)**

<u>Station 2</u>

l=1: $\tau_1 + \tau_2 \geq T_{11} + T_{21}$

l=2: $\gamma + \tau_1 + \tau_2 \geq T_{11} + T_{21} + T_{22}$

l=3: $2\gamma + \tau_1 + \tau_2 \geq T_{11} + T_{21} + T_{22} + T_{23}$

.
.
.

l=Q: $(N\text{-}1)\gamma + \tau_1 + \tau_2 \geq T_{11} + T_{21} + T_{22} + T_{23} + T_{2Q}$ **(6.19)**

....

<u>Station n</u>

l=1 $\tau_1 + \tau_2 + \ldots + \tau_M \geq T_{11} + T_{21} + \ldots + T_{n1}$

l=2: $\gamma + \tau_1 + \tau_2 + \ldots + \tau_M \geq T_{11} + T_{21} + \ldots + T_{n1} + T_{n2}$

l=3: $2\gamma + \tau_1 + \tau_2 + \ldots + \tau_M \geq T_{11} + T_{21} + \ldots + T_{n1} + T_{n2} + T_{n3}$

. .

. .

. .

l=Q: $(Q\text{-}1)\gamma + \tau_1 + \tau_2 + \ldots + \tau_n \geq T_{11} + T_{21} + \ldots + T_{n1} + T_{n2} + {}$
$+ T_{n3} + \ldots + T_{nQ}$ **(6.20)**

These equations can be written in a general form. If work congestion and the forcing of operators out of their stations is to be avoided, each launch l (l = 1, 2,..., Q) has to satisfy the inequality

$$(l-1)\gamma + \sum_{h=1}^{i} \tau_h \geq \sum_{h=1}^{i-1} T_{h1} + \sum_{h=1}^{l} T_{ih} \quad , \; i = 1, 2, \ldots, n \qquad \textbf{(6.21)}$$

Each station will have to satisfy these conditions for each launch. It is seen that in order to avoid station idleness and work congestion a set of inequalities has to be satisfied for each launch and for each station. Without a computer to calculate these inequalities, the calculation effort would be considerable because even a simple problem would require a great number of calculations.

The sequencing solution proposed by Kilbridge and Wester (1963) is a particular case of this, which happens when each model has the same time over all stations (each model spends an amount of time at the different stations equal to the model cycle time), although the station times for each model vary.

The purpose of developing these equations was to prove that if idle time and work congestion are to be avoided the fixed launching rate is only a function of the total workload of the first station on the line.

Therefore, the optimal fixed launching rate is given by

$$\gamma_{opt} = \frac{TT_{s1}}{Q} \tag{6.22}$$

6.4. Determination of Station Lengths for the Given Sequence.

Particularly for a line composed of closed stations, the station length is important. If the length is not appropriate, the operator may become idle for long periods or may not be able to complete the work during the time the unit is within his station limits. For open stations, because the operator is free to pass the station boundaries, idle time and incomplete units are less probable. When possible, a line composed of open stations should be chosen. This type of line will provide the maximum efficiency (Dar-El and Cother 1975). In a continuous flow-line the extent of operator movements will determine the required length for the station. For non-continuous lines (asynchronous systems and synchronous systems), the station length is not a point of major concern. In such cases, the station length will be determined by considerations such as type of product to be assembled - e.g. dimensions of product, required tools for the station, ergonomic and other technological constraints.

Previous research has indicated that it is preferable to use an early start schedule (the operator starts to work on the first unit in the sequence as soon as it passes the station upstream limit) rather than a late start schedule because it has the advantage of resulting in smaller overall line lengths. For closed stations (which are the most critical because if a unit cannot be completed within the stations limits the conveyor has to be stopped or the unit will leave the station incomplete), an early start will allow for a certain amount of idle time to cope with variability of task times, and therefore, seems to be particularly useful for the type of line considered[2]. For a line with open stations the choice between an early or late start does not seem to make sense, except perhaps for the first station in the line (which is usually closed to the left), where an early start may result in more idle time than a late start. However, as mentioned, this idle time may be useful to deal with task time variability. Also it will not make much of a difference in terms of introducing idle time and it is likely to result in shorter overall line lengths.

To determine the station lengths for a continuous flow line the notation presented by Dar-El and Cother (1975) will be used. For the beginning of the sequence each

[2]This may seem a contradiction. On the one hand it is said that one of the objectives is to minimize idle time to achieve a decrease in throughput time. On the other hand it is said that it would be useful to have idle time to enable the system to cope with task time variability. In fact, it is desirable to have both things. Minimization of idle time should be built into the line design (in the form of balanced lines), but in lines characterized by great task time variability a certain amount of idle time should be allowed. This amount of idle time will be used as slack time to permit the operator to deal with task time variability. An early start is likely to introduce this helpful amount of slack time.

operator will be at his station upstream boundary, ready to start work on the first unit as soon as it enters the station limits.

Let DM(i,l) be the displacement of operator i when starting work on the l^{th} unit launched; DP(i,l) be the displacement of operator i when completing work on the l^{th} unit launched; DMAX(i) be the furthest displacement of operator i in the downstream direction; DMIN(i) be the furthest displacement of operator i in the upstream direction; L_i is the length of station i; and L is the overall line length.

For the first unit in the sequence, the operator displacement is zero, i.e. DM(i,1) = 0. The upstream walking speed is assumed constant. The displacement of operator i when completing work on the l^{th} unit in the sequence is given by:

$$DP(i,l) = DM(i,l) + T_{s_{ij}} V_c \tag{6.23}$$

and the displacement when starting work on the next unit (the $l+1$ launch) is:

$$DM(i,l+1) = DP(i,l) - L_w \tag{6.24}$$

where L_w is the operator i upstream walking distance given by

$$L_w = V_o T_r = \frac{V_o S_p}{V_c + V_o} = \frac{V_c V_o \gamma}{V_c + V_o} \tag{6.25}$$

Note that the upstream walking distance may not be always constant. In fact, the operator, after completing work on a unit, will walk in the upstream direction a distance equal to w or until he reaches the upstream boundary (where the upstream boundary is closed), whichever occurs first. In the latter case, the operator will be idle waiting for the next unit to reach his station limits.

The furthest displacement of operator i in the upstream direction is:

$$DMIN(i) = Min|DM(i,l),\ l = 1,...,Q| \tag{6.26}$$

The furthest displacement of operator i in the downstream direction is:

$$DMAX(i) = Max|DP(i,l),\ l = 1,...,Q| \tag{6.27}$$

Therefore, the length of each station is given by

$$L_i = DMAX(i) - DMIN(i) \tag{6.28}$$

and the overall line length

$$L = \sum_{i=1}^{n} L_i \tag{6.29}$$

6.4.1. Station Length for Closed Stations.

Because the operator at the beginning of the sequence is at the station boundary and this is the furthest displacement in the upstream direction then DMIN(i) = 0, i = 1,..., n. Therefore, the station length is given by

$$L_i = DMAX(i) - DMIN(i) = DMAX(i) \tag{6.30}$$

6.4.2. Station Length for Open Stations.

In assembly lines with open stations, if the length of the conveyor between two adjacent stations is small, the stations will have a region that is common to both. The determination of the length for this type of stations is not so critical as for closed stations, because operators can pass their station limits to a certain extent so the work is more likely to be finished.

The proposed method provides that if a model overloads a station, the next unit to be launched will underload the station. This gives time to the operator at that station to pass the downstream limit and finish the work on the unit that has an excess of workload. When he starts working on the next unit, this unit has already entered the station limits and is somewhere in the station; however, because this unit has a service time smaller than the previous, it is very likely that the operator will be able to finish it within the station limits. After that he can walk in the upstream direction, maybe pass the station upstream limit and pick the next unit which will have a heavier workload. It is likely that this unit will be finished within the station limits (or if he passes again the downstream limit, the extent of his movement will not be excessive).

Therefore, this sequencing method is an attempt to have the units completed within the station limits, and if this is not possible the extent of the operator movement past his station limits will be small. For closed stations, if the station passage times are very different from the average station times (this can happen for example in cases where the stations lengths are previously fixed), the efficiency of the line suffers a steep decrease.

To calculate the limits for open stations a slight modification has to be made: in open stations, the operator starts to work on the first unit of the sequence as soon as it enters the station, but the station limits are given by the DMAX(i) and DMIN(i). The furthest upstream displacement is still

$$DMIN(i) = Min\,|\,DM(i,l),\ l=1,...Q\,| \qquad \textbf{(6.31)}$$

However, DMIN(i,1) is not zero for $i \geq 2$. For $i = 1$ (the first station on the line), DMIN(1,1) is equal to zero. This point will be the reference to which the displacements are measured.

The overlapping between stations i and i+1 is represented by $O_{i,i+1}$, where

$$O_{i,i+1} = DMAX(i) - DMIN(i+1) \qquad \textbf{(6.32)}$$

The length of the assembly line becomes

$$L = \sum_{i=1}^{n} L_i - \sum_{i=1}^{n-1} O_{i,i+1} \qquad \textbf{(6.33)}$$

6.4.3. Hybrid Stations.

The assembly line can be composed of open and closed stations mixed together or, stations can have one of the boundaries closed and the other open. The determination of lengths for each type of station is done by applying the equations presented in sections 6.4.1 and 6.4.2.

The calculation of station lengths can be easily implemented on a spread sheet. The determination of station lengths by the proposed methodology will result in no work congestion and no work deficiency. Also a unit will not leave the station incomplete.

Each sequence will require a particular station length, therefore, for fixed facilities the performance of the line may be variable.

6.5. Determination of Station Lengths and Operator displacement. Example.

The example used is the one given in section 6.1.3. The stations have open boundaries, the first station is closed to the left and no concurrent work is allowed. To simplify the calculations $V_c = 1$ and V_o is much greater than V_c. Then, for this case $\frac{V_c V_o \gamma}{V_c + V_o} \simeq V_c \gamma = \gamma$, because $\frac{V_o}{V_c + V_o} \simeq 1$, and consequently, the upstream walking distance L_w is numerically equal to the launching interval, i.e. $L_w = \gamma$. Note that because the first station is closed to the left (which means that the operator at the first station may wait for the arrival of the next unit) and there is no concurrent work allowed (which means that the operator will only walk a distance equal to L_w, if that does not result in adjacent operators working on the same unit), the upstream walking distance may not be the same all the time, and different operators may walk different distances. Nevertheless, whichever the case the upstream walking distance will be bounded by L_w.

Table 6.4: Operator Displacements ($\gamma = 4.583$); Stations 1 & 2.

	Station 1		Station 2	
1	DM(1,1)	DP(1,1)	DM(2,1)	DP(2,1)
1	0	5.00	5.00	7.00
2	.42	4.42	4.42	12.42
3	0	5.00	7.83	9.83
4	.42	4.42	5.25	13.25
5	0	5.00	8.67	10.67
6	.42	4.42	6.08	14.08
7	0	5.00	9.50	11.50
8	.42	5.42	6.92	8.92
9	.83	4.83	4.83	12.83
10	.25	5.25	8.25	10.25
11	.67	4.67	5.67	13.76
12	.08	5.08	9.08	11.08

Table 6.5: Operator Displacements (γ = 4.583); Stations 3 & 4.

l	Station 3 DM(3,l)	Station 3 DP(3,l)	Station 4 DM(4,l)	Station 4 DP(4,l)
1	7.00	11.00	11.00	16.00
2	12.42	16.42	16.42	21.42
3	11.83	15.83	16.83	21.83
4	13.25	17.25	17.25	22.25
5	12.67	16.67	17.67	22.67
6	14.08	18.08	18.08	23.08
7	13.50	17.50	18.50	23.50
8	12.92	16.92	18.92	23.92
9	12.83	16.83	19.33	24.33
10	12.25	16.25	19.75	24.75
11	13.67	17.67	20.17	25.17
12	13.08	17.08	20.58	25.58

DMIN(1) = 0
DMAX(1) = 5.42 L_1 = 5.42

$O_{1,2}$ = 1

DMIN(2) = 4.42
DMAX(2) = 14.08 L_2 = 9.62

$O_{2,3}$ = 7.08

DMIN(3) = 7.00
DMAX(3) = 18.08 L_3 = 11.08

$O_{3,4}$ = 7.08

DMIN(4) = 11.00
DMAX(4) = 25.58 L_4 = 14.58

The line length is 25.58 meters.

Figure 6.7: Operator Displacement Diagram for Stations 1 and 2

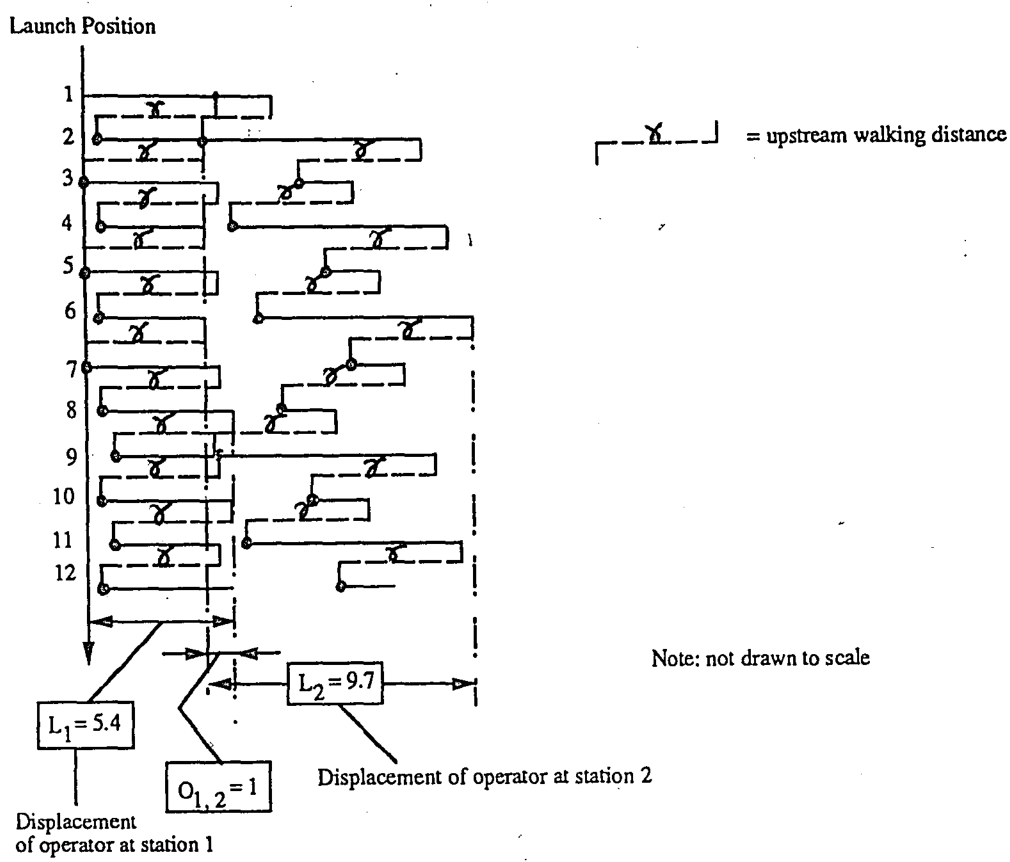
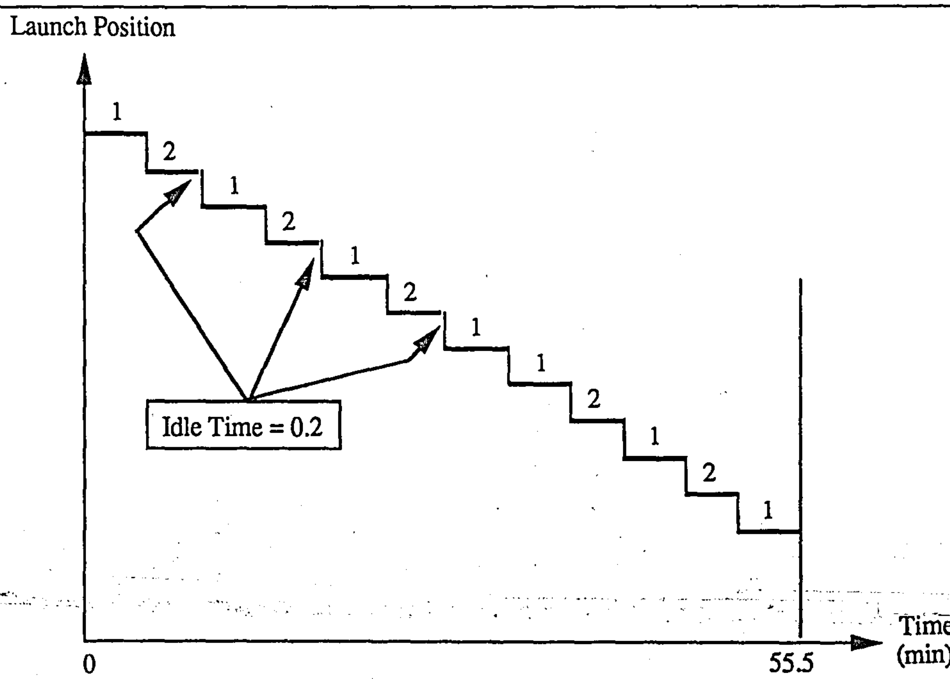


Figure 6.8: Gantt Chart for the Scheduled Period (for Station 1 Only)



6.6. Calculation of FLR when the Assembly Line is Composed of G Labor Groups.

When the assembly line is composed of various labor groups, Thomopoulos (Prenting and Thomopoulos 1974), suggests that the Fixed Launching Rate for the line should be determined by the following procedure:

1) Consider each labor group g as an independent assembly line.

2) For the G "assembly lines" determine the optimal Fixed Launching Rate γ_g.

3) The Fixed Launching Rate for the entire assembly line is given by $\gamma = \max \gamma_g$, for g = 1, 2, ..., G.

Previous literature has suggested that the choice of a particular criterion for the sequence is dependent on physical, technical and ergonomic constraints of each specific situation. For example, in lines where the station length is fixed, and production runs are small, minimizing throughput time would seem appropriate. If the task times are greatly variable then it may be more appropriate to allow for a certain amount of idle/slack time that will allow the operator to cope with task time variability. The proposed method just tries to have each worker accomplishing the same amount of work for the different units launched. For situations where a particular objective is required, such as minimizing throughput time, it may result in a greater throughput time than what would be obtained by a method directed towards minimizing it. The proposed method is very simple, and (as will be shown in Chapter 8) seems to give acceptable solutions. The line designer will have to make a trade-off between the cost of an algorithm that provides a best solution (for the proposed objective) and the use of the proposed method, where a good solution will be found most likely at a lesser cost.

7. Literature Review

The following is a review of some of the previous work in mixed-model assembly lines.

7.1. Line Balancing.

Apparently very little has been published on line balancing for mixed-model assembly lines. This may be due to the fact that methods for single-model assembly line balancing may be used for the mixed-model case by the already mentioned modification.

Several methods for the line balance of single-model lines are found in the literature (e.g. Hegelson and Birnie 1961, Kilbridge and Wester 1961, Held, Karp and Sharesian 1963, Tongue 1965, Mansoor 1973, We-Min Chow 1990, etc.). Among those, there are algorithms that solve the line balance problem yielding the optimal solution, heuristics that give a good solution (reaching very often the optimal solution), and computer programs like COMSOAL, CALB, etc. (these computer programs often yield the optimal solution).

The efficiency of the balancing solution depends on the balancing method used, and usually the optimum solutions require the most computing time (Macaskill 1972). The use of a very sophisticated balancing method, may be costly (in terms of computation effort) and still there is the possibility of not reaching the optimal solution because at the line design phase, the designer has only an estimation of the task times, and actual times may differ from the estimated. In manual lines it is very likely that these times will differ. Other aspects that are not identified at this time may also affect the line balancing. Kilbridge and Wester (1961) acknowledge that the solution given by the heuristic they propose is only a guideline for the line designer and empirical considerations may affect the line balance procedure.

7.1.1. Algorithms Seeking the Optimal Solution.

Villa (1970) extends the Gutjahr and Nemhauser algorithm (which solves the line balancing problem as a shortest route problem) to the case of mixed-model assembly lines, and proposes a branch and bound method for the mixed-model line problem. This branch and bound algorithm is a sequential procedure because it finds the optimal solution by considering stations one at the time. The algorithm will yield the optimal solution if the problem has at least one feasible solution, and will indicate that there is not a feasible solution if the problem does not have a solution. It was proved that the algorithm converges. Extensions for locational constraints, minimum work content constraints and variable cycle times were also included. The assignment of tasks to stations is such that the number of stations is the same for all models, each work element is assigned to exactly one station and the precedence constraints are satisfied. Integer programming is tested with a small problem (2 models, 9 tasks). For such a problem there were a very large number of variables and constraints (respectively, 120 and 60). This reveals that even for small problems the formulation and optimization by linear integer programming can be very difficult.

The proposed branch and bound algorithm required less computational effort than the Gutjahr and Nemhauser algorithm. For the branch and bound method the computational results showed that the number of tasks does not have a great influence on the computer execution time. More relevant to the execution time was the complexity of the precedence constraints (i.e., how tasks relate to each other)[1]. An increase in the number of models seems to result in a linear increase in execution time, the same being verified with the number of feasible assignments to the first station without considering the cycle time. Both parameters have a considerable influence on computation time. The length of the assembly line seems to have a small effect on execution time. It was suggested that for a problem with m models the line balance could be done by first solving m single-model problems, and for models with less stations than the models having the largest number of stations, the work elements should be rearranged to insure that all products have the same number of stations. The rearranging of the work elements must be done respecting the precedence constraints. It was concluded that this approach would require more computational effort than the branch and bound algorithm and therefore the branch and bound algorithm appeared to be the best method (among those studied) for the line balancing problem.

Deutsch (1971) also develops a branch and bound algorithm that solves the line balancing problem for mixed-model assembly lines by minimizing the number of stations and, given the obtained number of stations minimizes the cycle time. This type of methodology that initially minimizes the number of stations and then attempts to

[1]Note that by that time computer execution time was an aspect of great concern; today, due to great advances in computer technology, this seems less relevant.

minimize the cycle time given that number of stations is very common and is known as a multi-phase technique. Sneider (1980) presents a five stage procedure to solve optimally the single-model line balancing problem. The particular aspect of this work is that the method will solve the line balancing problem by minimizing both the number of stations and the cycle time (the majority of line balancing methods solve the problem by minimizing just one of the two variables). The first stage bounds the problem based on the input data. The second phase uses a heuristic that generates a good feasible solution and further bounds the problem. If the solution obtained in the second stage is not optimal, a Mixed Integer Program exploits the problem structure and the heuristic solution. The optimal solution may be found in the second, third, fourth or in the fifth stage if it was not previously found. This method has the advantage of a heuristic coupled with the optimal quality of an algorithm. Although this work was done for single-model problems the extension to mixed-model may be done by using task-group balancing.

7.1.2. Comments.

The optimization algorithm method presented by Villa (1970) has the apparent advantage of not needing to know the total production requirement. Balances are achieved based on a desired cycle time (which is the time that a unit is available to a workstation) rather than on total times. Therefore the balance solution obtained will not be dependent on production requirements but only on element times. The requirement that the number of workstations should be the same for each model is logical, because if this was not the case, there would be stations that would be completely idle for some models launched into the line, and the objective of a smooth flow would not be accomplished. The fact that some stations are idle would contribute to a decrease in line efficiency.

Task-group balancing is directly dependent on the quantities to be produced of each model (remember that $TT_k = \sum_{j=1}^{m} \sum_{k=1}^{K_j} Q_j T_{e_{jk}}$). Therefore, the assignment of tasks to stations obtained for a given production level is not necessarily the same as obtained for a different production output (i.e. for different Q's), even if the number of models and the nature of the tasks is the same. When using an aggregated task method, if the production level changes it may be necessary to rebalance the line, whereas with the Villa's branch and bound algorithm this theoretically does not happen. On the other hand, Villa's algorithm also depends on the production requirement by the fact that it is based on the cycle time and the cycle time is dependent on production requirements.

7.1.3. Unbalancing of Production Lines.

Hillier and Boling (1979) found that when operation times are variable, the mean rate of production of an unpaced production line with more than two stations can be maximized if the stations workloads are appropriately unbalanced. The optimal assignment of workloads to stations followed the "bowl phenomenon" which allocates the greatest workloads to the first and last stations in the line and less amounts of work to the stations in the middle. Because workloads are not evenly distributed over all stations the result is an unbalanced line. The use of bowl allocation of workloads resulted in an increase in the mean production rate over the perfectly balanced line.

In a simulation study to evaluate the efficiency of unbalancing production lines with normally distributed operation times it was verified that there is an improvement in the efficiency of the line if finite buffer storage is used combined with an appropriate unbalance of the line, even when a small variability of task times is present. By appropriate unbalance is meant a bowl allocation of workloads where "the optimal bowl allocation is symmetric and is relatively flat in the middle and very steep towards the beginning and the end of the line" (So 1989). The results of the simulation study showed that although the improvement was generally very small, in most of the cases tested, it was statistically significant even for a level of significance of 95%.

In general, when the unbalance was not done in an appropriate way (did not follow the optimal bowl allocation) a decrease in efficiency resulted. The study also suggested that methodologies to attain perfectly balanced lines are likely to give near optimal performance in cases when the variability in processing times is low.

7.1.4. Comments.

Heuristic methods of line balancing seem to be an acceptable methodology for line balancing. The complexity of "real life" problems is likely to preclude the use of algorithms that yield perfectly optimal solutions. These algorithms are always sophisticated procedures and the complexity of some problems (e.g. great number of restrictions) will render them inefficient. They may provide optimal solutions for deterministic cases, but the same may not happen for the stochastic situation common in manual assembly lines. Because they are less sophisticated and can solve complex problems, heuristic methods seem to be the best approach for solving the line balancing problem.

7.2. Sequencing.

The sequencing problem for mixed-model assembly lines was first presented by Kilbridge and Wester (1963). For an assembly line with a moving conveyor and non-overlapping stations, they suggested two approaches for the solution of the sequencing problem, based on two model launching systems. The criteria for finding the optimal sequence was to minimize idle time and avoid congestion. It was assumed that the workload of each model was evenly divided among the stations; therefore the amount of time that a given model would spend at one station would be equal to the amount spent in each of the other stations. Consequently, each product will spend in each station an amount of time equal to the cycle time for that product. However, this ideal situation (evenness in station workloads for each model) is not very likely to happen in the reality. The two launching systems were the Variable Launching Rate (VLR) and the Fixed Launching Rate (FLR).

For the conditions studied the results were that with VLR, units can be fed into the line in any order and yet not cause station idleness or congestion. Therefore, for this launching system the sequence of products can be randomly generated[2]. The disadvantage of this system is that difficulties of integrating the assembly line with other

[2]This conclusion is only valid for the particular case studied where each model was assumed to have the same service time over all the stations, although the service time could differ for the models.

production lines may arise. Also, it is usual in industry to schedule the delivery of components to the lines at fixed periods of time. The use of a VLR may prevent the line to cope with these scheduled deliveries. The work carriers (e.g., hooks) in the moving conveyor are usually equidistantly spaced and a variable launching rate would introduce the difficulty of where to place the carriers. The fixed launching interval has the advantage of providing a uniform rate of production and is easily adaptable to standard industrial techniques. This type of launching system is very appropriate when several production lines are synchronized to feed the assembly line. However the sequence of products has to be carefully determined. In general, if overlapping stations is permitted the line efficiency will increase (provided that there is not interference between operators).

One method for sequencing is presented by Thomopoulos (Prenting and Thomopoulos 1974). Stations with open boundaries are considered and penalty costs for various types of inefficiencies are included. Work performed outside predetermined station limits is considered an inefficiency. The optimal sequence will be chosen as the one that minimizes the cost of the inefficiencies for the assembly line. The launching system is the fixed launching interval and a moving conveyor with constant speed is assumed. The method does not provide optimal solutions, but studies showed that the results are close to optimum. For a given line configuration (station types, station lengths, station passage times, conveyor speed) and a given penalty cost associated with each inefficiency (given in cents per time unit) the resultant total inefficiency cost is calculated for launching each model (each model will result in a certain inefficiency time at each station on the line). The model that yields the least inefficiency cost is launched. The next model to be launched is the one that results in the least inefficiency cost given the previous launches. The methodology is successively applied until all Q models have been launched.

Difficulties of this method are that it may be difficult to define the penalty costs for the inefficiencies and also that it may difficult to define the station boundaries. According to Dar-El (1978) "the negative aspects of this approach are that it does not provide help for determining station lengths for a new design situation and that depending on the station length any sequence can be made to look good". Another aspect that may be important is that for complex problems (where the number of models and stations may be large) the computer effort may be considerable because the inefficiency cost of launching each model remaining to be produced has to be evaluated for every station on the line!

Dar-El and Cother (1975) argue that it is difficult to define the limits of operator movements and this makes it difficult to calculate the inefficiencies of the line. The only known previous method for the general mixed-model sequencing problem (Thomopoulos heuristic[3]) determines the sequence that minimizes the penalty cost of inefficiencies. However the inefficiencies are measured and defined by the extent of operator movements outside his workstation, and the extent of these movements is extremely

[3]Kilbridge and Wester method was for the mentioned particular situation.

difficult to determine. The arbitrary selection of station lengths affects directly the generated product sequence. The authors argue that the choice of a line length (or equivalently the choice of station lengths), because it affects the inefficiencies, can make any sequence become efficient or inefficient. This is considered a weakness of the Thomopoulos method.

Instead of determining the sequence that minimizes the inefficiencies, Dar-El and Cother presented a method to generate the sequence that minimizes the overall assembly line length for no operator idle time or other inefficiencies. Therefore, the method will provide the station lengths for the required sequence and not the inverse. Two line configurations were assumed: open stations and closed stations and the line is assumed to be perfectly balanced. The products to be assigned to the sequence are referred as the "pool". The lower bound for each station is calculated, and the station lengths are initially taken as the lower bound. A heuristic method of selection, which ranks each model in a descending order of priority, will assign a model from the pool. Successive products from the pool are assigned until the pool is empty. Once a product is selected for the sequence it will have to satisfy an "acceptance heuristic" which means that the model chosen cannot violate the stations limits. If the limits are violated then the model is returned to the pool and the next model (in the priority order) is tested. If at some stage no model satisfies the "acceptance heuristic", the station length limits are increased by equal amounts, all products are returned to the pool and the sequencing procedure starts again. In this study it was assumed that the operator moves instantaneously from the position where he leaves the unit on which he was working to the position where he reaches the next unit. This method has some disadvantages, namely the fact that when a product satisfying the "acceptance heuristic" cannot be found, all stations are incremented by an equal amount and the sequencing procedure has to start again.

The algorithm can be used for determining new sequences if small changes in the production requirements or in the model design occur once the assembly line is in operation. The authors state that "idle time will be minimized but the production requirement may not be achieved although the models would be sequenced in proportion to the demand".

The influence on the overall assembly line of five factors was studied: (1) number of models, (2) number of stations, (3) model cycle time deviation factor, (4) model production requirement deviation factor, (5) and operator-time deviation factor, where the deviation factors are respectively a measure of the variation between model cycle times, measure of the production requirement variation and a measure of the operation times variability. The conclusions were that the overall assembly line length is predominantly influenced by number of models, model cycle time variation and operator-time deviation factor. It was shown that for a given model-mix the assembly line length decreases when better line balances are achieved. A sequence resulting in an overall line length 10% bigger than the lower bound was considered to be good.

In order to achieve maximum line efficiency, the use of open stations and overlapping is also suggested.

Dar-El and Cucuy (1977) proposed a sequencing procedure for perfectly balanced lines. The method results in the optimal solution (minimum overall assembly line length for no operator idle time), but would only be applicable for perfect balanced lines, which are unlikely to occur. The authors indicate that imbalance will result when technological constraints occur (e.g. when zoning constraints occur - this will result in different workloads at the stations for each model, even though the station workloads per shift may be the same). Because of this, the method is only applicable to perfectly balanced lines.

Dar-El (1978) defines two objectives that can be found in mixed-model sequencing problems: (1) to find the sequence that minimizes the overall assembly line length and (2) to find the sequence that minimizes throughput time, the latter being the typical flow-shop sequencing problem. Several variations of assembly lines were studied. According to the type of assembly line a different objective is suggested for the sequencing problem. Lines where the product to be assembled can be moved independently of the movement of the conveyor, i.e. Products Movable lines have the greatest flexibility in coping with changes in production because the layout is not dependent on the sequence. This is not the case with Product Fixed lines where each time that production changes it is necessary to find a new solution for the station lengths. The author indicates that for the case of items fixed to the conveyor and many stations in the system, minimizing throughput time and minimizing facility length are practically the same. When the number of stations is small (4 or 5 stations) a tradeoff between the objectives of minimizing facility length and minimizing throughput time needs to be considered. For products removable from the conveyor and for stationary assembly lines the size of the facility becomes irrelevant and the concern should be to find the sequence that minimizes throughput time (the classical flow-shop problem). In these cases if the launching system used is FLR the only station that will be fed with a constant rate of subassemblies is the first one and, the author states that VLR becomes the only meaningful launching discipline.

Dar-El concludes that the greatest flexibility is achieved in 'Product Movable' lines, because of its flexibility to cope with changes in production, since the layout will be independent of the sequence; therefore, when changes occur the facility length is not affected. When the product is fixed to the conveyor then stations should have open boundaries and the preferred launching is VLR. For closed stations (and products fixed to the conveyor), the author states that "There is not evidence which launching system is to be used. If all models are approximately well balanced over the n stations, then there are advantages in using the VLR discipline, since any sequence would minimize both the overall line length as well as the throughput time".

Bard et al. 1992, formulated analytically the sequencing problem for situations with several different parameters. To solve these problems a relaxed linear integer programming procedure was used. The authors suggest that for the case of lines with closed stations a tradeoff exists between the length of the line and idle time. A late start schedule will allow the operators to work continuously because "a late start indirectly assures that a sufficient amount of work-in-process inventory is available on the conveyor to avoid starvation". With a late start schedule the stations are designed in such a way that an operator never has to wait for arrival of the next unit at the station boundary. However, the result of using a late start is that an increase in the facility length is very likely. An early start schedule is likely to produce smaller facility lengths but will induce a certain amount of idle time.

For open stations the use of a late start or an early start does not seem to have the same implications as for the case of closed stations. With open stations, the operator does not need to wait at the upstream boundary for the next job to arrive, and therefore an early start will allow start of work on the first unit in the sequence immediately as this unit enters the station.

The study performed by these authors also suggests that when possible the Minimum Part Set (MPS) should be used. The choice of a particular line configuration (closed or open stations, products fixed or removable, overlapping stations, etc.) depends on the conditions of each particular situation and, therefore, the sequencing objective can vary. For example, if the size of the facility is fixed and the production runs are small the objective for the best sequence should be to minimize throughput time; if the work element times are characterized by high variability, then the best sequence should be the one that maximizes idle time, or equivalently, minimizes facility length. Here it can be seen that the authors suggest the use of slack time to reduce task time variability, similar to Hoffman (1990).

The conclusions of Bard et al. study are that the use of an early start schedule rather than a late start schedule could in fact yield significant reductions in line length. The use of open stations (rather than closed stations) results in a decrease in the throughput time.

Open stations also reduce the overall assembly line length, and therefore should be used when possible. The problems tested revealed that to find the sequence that minimizes line length or to find the sequence that minimizes throughput will not result in lines substantially different. These two measures were always within 5% of each other when the remaining parameters were held constant. The use of Variable Launching Rate yields the best results with respect to overall line length, throughput time and idle time. It was observed that use of VLR will decrease idle time significantly, whereas the decrease in line length and throughput time may not be significant.

Other papers concerning the mixed-model sequencing problem include Macaskill (1973), previously mentioned, who simulated four different situations for a mixed-model assembly line of the moving belt type with station overlap and certain products fixed. These situations were: (1) complete sequencing of a shift with concurrent work allowed, (2) complete sequencing of a shift but no concurrent work, (3) sequencing of a limited number of products in a special situation, with concurrent work allowed, and (4) the same as (3) but no concurrent work. For balance purposes it was assumed that residual work between shifts is approximately the same; however, for the simulation exact details of location and extent of completion of each residual were required.

The sequencing method used that was basically the same as Thomopoulos, but included "steps to avoid build-up of a residue of high penalty models that reduced performance in the Thomopoulos method". It was concluded that task-group balancing, although it may result in uneven distribution of workloads, is an acceptable method for general use in mixed-model balancing. It was verified that if concurrent work is allowed the damaging effects of task variability can be greatly reduced, and apparently it facilitates the generation of good sequences. It is inferred that concurrent work has substantial advantages in assembly line operation.

Okamura and Yamashina (1979) developed a sequencing heuristic that minimizes the risk of stopping the conveyor in a complex mixed-model assembly line under situations of variable task times. The line was composed of closed stations. When the work cannot be completed within the station limits, the conveyor has to be stopped due to the high cost of incomplete items. The sequence that maximizes the distance from the furthest operator downstream displacement to the station downstream boundary for all stations is the sequence that minimizes the risk of stopping the conveyor. The method generates randomly a sequence that will be iteratively improved into an optimal or near-optimal sequence by inserting a product into another position and interchanging product pairs in the sequence.

According to Kao (1981), the disadvantage of this method is that, as with the Thomopoulos method, any sequence can be made to appear efficient or inefficient depending on the choice of the station lengths.

Dar-El and Navidi (1981) applied the Dar-El and Cother sequencing method to a problem composed of 8 different models, 50 work elements, and a required production per shift of 17 units. The Dar-El and Cother method was extended to include cases where the balance may not be perfect, and this resulted in an additional station lower bound. A scheme with multi-function workers permits reduction of the number of workers in the line. It is suggested that it may be beneficial to introduce buffer stocks at the stations with greater task time variability. Buffer stocks will be helpful to lessen the effect of task time variability.

Wang and Wilson (1986) compared three assembly line designs with respect to total station idle time, incomplete units and production rate. The three designs were: (1) a moving belt with products fixed, (2) a moving conveyor with products movable, and (3) an accumulation conveyor with products removable. A sequencing heuristic is proposed and a simulation study evaluates the performance of the different line designs. Station times were assumed variable. It was concluded that the method proposed is effective and that an accumulation conveyor with products movable improves throughput, reduces worker idle time and utility work compared to a moving belt with products fixed. The moving belt with products fixed revealed the greatest idle time and the greatest number of incomplete jobs. Although the simulation was not done for an extensive number of sequencing methods and problems, the conclusions appear to be valid.

Bolat (1988), developed a method that gives the sequence which minimizes total setup and utility work cost. A branch and bound algorithm becomes inefficient for more than 20 jobs and therefore heuristics were used.

Kao (1981) modifies the Dar-El and Cother method and suggests that the splitting of unavoidable idleness between units has potential for lowering the station lower bound and therefore a potential to decrease line length. A new fixed launching rate that would deal with the possibility of multiple shifts is proposed. The launching rate for mixed-model lines (proposed by Dar-El and Cother 1975, Thomopoulos 1967) would result in incomplete jobs when the sequence determined was repeated for the next shift. The proposed method was shown to perform better than the Dar-El and Cother method.

Some methods deal with the sequencing of mixed-model lines in just-in-time multi-level production systems. The assembly line is the highest level of the production system, and the demand at the assembly line will trigger the demand on the lower levels. In order to control production in the lower levels (i.e. to have smooth production of components parts and sub-assemblies) an appropriate sequence at the assembly line is required.

These methods (Bancroft 1987, Monden 1983, Miltenburg 1989, Miltenburg and Sinnamon 1989) determine the sequence that results in constant consumption of the component parts used in the assembly line. The idea behind this is that keeping a constant usage of component parts at the assembly line will result in smooth production of these parts at the lower levels; there is a potential to reduce work-in-process inventories, which is one of the goals of a JIT production system.

Sumichrast et al. (1992) statistically compared five sequencing procedures for mixed-model assembly lines in a just-in-time multi-level production system. The evaluation of these procedures was based on four measures of inefficiency: (1) assembly work deficiency (which they defined as work not completed, a different definition than the one given in this research), (2) worker idleness, (3) worker home time, and (4) mean square deviation from linear usage of components. Workers were allowed to pass the upstream and downstream station limits to a certain extent, and concurrent work was allowed. An overloaded worker could be helped by a worker that was under-utilized, if the latter was not too distant.

The main objective of three of the sequencing methods compared is to determine a sequence that results in uniform consumption of component parts. These Methods are Toyota's Goal Chasing methods I and II and the Miltenburg (1989) algorithm 3 using heuristic 2. The other two sequencing methods were the traditional batch sequencing procedure and the Time Spread method developed by the authors. The Time Spread method smooths the workload in each station on the assembly line. The assembly line is assumed to be balanced so that each station will be able to process the average amount of work without the need of extra workers beyond the one assigned to the station. If several units with excessive work content are consecutively launched, the operator will not be able to finish the work within his station limits and the unit will leave the station incomplete; or if this is to be avoided, a utility worker will be needed to help the worker with such a unit.

The study showed that in general the batch sequencing method performed poorly, while the other methods gave good performance. Among these, the Time Spread and the model developed by Miltenburg, appear to be the most efficient (i.e. for the efficiency parameters considered, they produced the best results). Overall, the Time Spread method showed best results with respect to idle time and work not completed. T-tests with a significance level of 0.05 showed that the Time Spread was statistically better in respect to operator home time. The authors concluded that the Time Spread method is slightly preferable when the structure of the product is not considered and when assembly efficiency (such as idle time and work not completed) is the main objective. The Miltenburg method is preferable when uniform parts consumption is to be achieved. Therefore, for a multi-level production system the Miltenburg algorithm seems preferable, particularly when the structure of the product is complex, whereas in a assembly line where the levels and product structure are not directly of concern, the Time Spread seems more appropriate.

It is not unusual to find in the industry assembly lines where the sequencing of units into the line is done without applying any mathematical algorithm. Empirical considerations may influence the selection of models to be launched. The procedure often used is to space the models at fixed intervals. As an example, if the total daily production is of one hundred units, and there are 10 units of model A, 20 units of B, etc., then every tenth unit will be a model B, every fifth unit will be a B, and so on.

7.2.1. Comments.

Goals such as leveling the station workloads, reducing setups, maintaining a constant usage of components, minimizing penalty costs, and minimizing inefficiencies have been considered in this literature review; however, other objectives for the sequencing (such as maximizing the product variety, etc.) may need to be considered. To minimize the setup cost in a mixed-model line does not make too much sense. A mixed-model line is typically a production system where line changeovers are not important, or at least not a major point of concern. Therefore, minimizing changeover costs is not likely to be an option in such production lines. If changeovers are in fact important, then the line will no longer be a mixed-model line; it will become a batch-model line. In these lines changeovers are aspects that need to be considered and sequencing methods can be obtained that aim to minimize changeover costs.

7.3. The Minimum Part Set (MPS).

When it is possible to partition the total production requirements into identical smaller requirements over several schedules, the result will be that each schedule is much more manageable because the problems are reduced to a practical size which facilitates the calculations, and the efficiency of the assembly line will increase (Prenting and Thomopoulos 1974, Bard et al. 1992). Thomopoulos (Prenting and Thomopoulos 1974) showed that if partitions of the total part set are used, the inefficiency costs will decrease. The different partition sizes studied ranged from 100 to 1. It was observed that the inefficiency costs decreased until a partition size of 10. Smaller partition sizes resulted in an increase in the inefficiency costs. McCormick et al. (1989 - referred in Bard et al. 1992), showed that the use of partitioned schedules rather than the total production requirement enables line operation to be achieved more quickly. Bard et al. (1992), state that: "previous research (Prenting and Thomopoulos 1970, Dar-El and Cotter 1975) have suggested that when heuristic methods are used the results obtained with the use of partitioned schedules are far better than the ones achieved with use of the full part set".

In the automobile industry, for example, it is not unusual to have a number of different models that can be as high as 3000 or 4000, and required daily production of 1000 units (Monden 1983). In such cases it would be desirable to partition the daily production in a more manageable number of units. If a cyclic pattern could be found, then instead of having to schedule a significant number of units the calculations would be confined to a more manageable number. Once the best solution for the MPS has been found, it will be cyclicly applied to the following set of units until the total production requirement is achieved.

7.4. Summary.

Previous research suggests that for the type of systems considered in this thesis (manual mixed-model assembly flow lines, continuous conveyor, products fixed to the conveyor and no buffer storage), there are advantages in using: (1) an early start, because it will better handle the variability in task time (early start allows for a certain amount of "slack time"), (2) concurrent work and multi-function workers because it allows a decrease in idle time and throughput time, (3) open stations because they may reduce the facility size and throughput time, and (4) the fixed launching rate because it allows a better integration of other lines into the assembly line. In some occasions it may be beneficial to introduce buffer stocks - in a continuous flow-line this may be done by increasing the distance between two consecutive stations. If possible, a line where Products Movable are allowed should be used, because the layout will not be dependent on the sequence.

When applicable, the Minimum Part Set should be used. Savings in calculations effort are likely to be achieved and previous research indicates that heuristic sequencing methods yield best results with partitions than with the total set part.

It seems to be a consensus that because mixed-model problems are likely to result in a large number of variables and constraints, the solution derived from an optimization algorithm is only possible at the expense of a large number of computations and considerable memory requirements. Heuristics methods (for the line balancing and sequencing problems) are fast and can solve large problems and seem to be the best option over all. However with the development in high speed computers, algorithms that were formerly put aside because they were slow and required heavy computer effort are more feasible. On the other hand, several factors, such as variability in task times, inappropriate sequence of models, size and complexity of real problems may lessen the quality of the solution obtained through sophisticated optimization methods. In the case of heuristics, it may not be possible to tell how far from the optimality is the resultant solution.

It is therefore questionable if the price paid for a sophisticated procedure that theoretically yields an optimal solution (and in practice may prove to be other than the optimal solution) is really a good option. The use of heuristic methods has proven to perform very well (Macaskill 1972 and 1973, Prenting and Thomopoulos 1970) and is less costly and less complicated than using sophisticated methods.

8. Comparative Analysis of Sequencing Methods and Line Balance Solutions

This chapter is dedicated to a comparison between the method proposed in section 6.1 and several other methods for the sequencing problem. A comparison between several solutions for line balancing and the use of different launching rates is also discussed.

8.1. Comparative Analysis for Different Line Balancing Solutions.

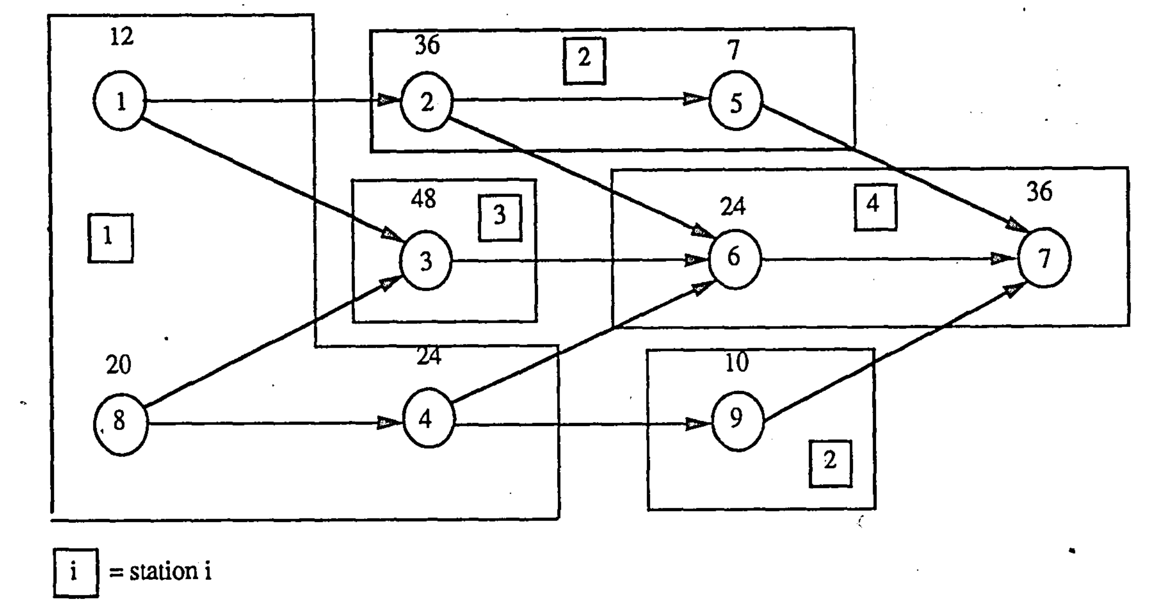
To compare different solutions for the line balancing problem the example presented in section 5.3.1.1 is used. The three balance solutions are given by: (1) the Largest Set Rule method, (2) the R.P.W. method, and (3) a bowl allocation of workloads to stations.

The bowl allocation was used to investigate if the conclusions drawn by Hillier and Bolling (1979) and So (1989) for unpaced lines could be extended to the case of paced lines (which is the type of line considered in this thesis). Recall (section 7.1.2) that the mentioned authors verified that when operation times are variable, an appropriate unbalance of the line[1] and the use of finite buffer stocks could result in an improvement of the mean rate of production over a perfectly balanced line. A manual assembly line

[1]Where appropriate unbalance means to use a bowl allocation of workloads to the stations. This means that the stations on the extremes of the line will receive a heavier workload than the stations in the middle; the optimal bowl allocation would be, according to So (1989), "symmetric and relatively flat in the middle steep towards the end of the line".

is, as already mentioned, characterized by variability in the task times. Therefore, in order to investigate if the output of an unbalanced manual mixed-model assembly line could be improved over the output of a balanced line, a bowl allocation of workloads to stations was used (see Figure 8.1). Note that for the example given it was not possible to achieve a perfect balance.

Figure 8.1: Assignment of Elements to Stations for the Bowl Allocation.



8.1.1. Model Parameters.

Different launching rates were used in order to evaluate how the line would be affected by them. These launching rates included: (1) variable launching rate, (2) FLR proposed by Kao (1980), where $\gamma = \max\left(\frac{TT_{si}}{Q}\right)$, $i = 1,\ldots,n$, (3) FLR calculated by the method presented in section 6.3, i.e. $\gamma = \frac{TT_{s1}}{Q}$, (4) FLR calculated by $\gamma = \frac{TT_{wc}}{Q \times n}$, (5) FLR calculated by the Kilbridge and Wester (1963) method, i.e. $\gamma_{K\&W} = \frac{\sum_{j=1}^{m} Q_j T_{cj}}{Q}$, where T_{cj} is the cycle time for each model (the maximum station service time for each model), and (6) FLR given by $\gamma = \frac{TT_{s1} - T_{1Q}}{Q-1}$[2].

The evaluation parameters were overall idle time, throughput time (defined as the time at which the operator at the last workstation finishes work on the last unit), line length, and number of incomplete items at the last station.

In order to maximize the efficiency of the line, open stations were assumed; however, concurrent work was not permitted (it was assumed that worker i+1 could only start work in unit l if the operator i had finished work on this unit). If concurrent work had been allowed, the result would most probably have been a decrease in throughput time. Because the upstream walking speed is usually much greater than the conveyor speed, it was assumed that the repositioning time can be neglected and therefore the operator's upstream walking distance is $L_w = V_c\gamma$. V_c was taken as equal to 1 and therefore $L_w = \gamma$. The first station was assumed closed to the left.

The line balance solutions are given in Tables 8.1, 8.2 and 8.3.

The application of the proposed sequencing method to three different solutions resulted in the same sequence {1 2 1 2 1 2 1 1 2 1 2 1}. The results of the study are

---

[2]If a single shift is to be scheduled, i.e. after the launch of the Q^{th} unit, there will be nothing else to be launched, then Eq. 6.12 would result in this launching rate (T_{1Q} is the time required at station 1 to assemble the unit launched at position l=Q).

presented in appendix A.1.

Overall, it was observed that the 5th launching rate produced the worst results of all, except for the line length that was the smallest. These results may be easily understood if we remember that the Kilbridge and Wester method is for the particular case mentioned in section 6.3, and for situations that deviate from that case it may not performed very well (as it was observed here).

It was observed that the idle time increased with an increase in the launching rate. An increase in the launching interval also resulted in smaller line lengths. This was already expected, because it is intuitive that a greater launching interval will increase idle time - the operators will be able more often to finish their work before the arrival of the next unit and will need to wait for the next unit to arrive. A smaller line length results because the operators in these conditions will more often start to work on the arriving units in a position that is closest to the station upstream limit; this way the displacement in the flow direction will be smaller. However the effect of greater launching intervals is to increase throughput time. The greatest launching intervals were given by the Kilbridge and Wester launching rate, which achieved the poorest results of all.

Table 8.1: Line Balancing Solution for the Largest Set Rule

Station i	Station time (T_{si})	$T_{s_{ij}}$ 1	$T_{s_{ij}}$ 2
1	55	5	4
2	54	2	8
3	48	4	4
4	60	5	5

S.I = 14.3; d = 9.6%

Table 8.2: Line Balancing Solution for the R.P.W.

Station i	Station time (T_{si})	$T_{s_{ij}}$ 1	2
1	56	3	7
2	58	4	6
3	60	5	5
4	43	4	3

S.I. = 17.6; d = 9.6%

Table 8.3: Line Balancing Solution for the Bowl Allocation

Station i	Station time (T_{si})	$T_{s_{ij}}$ 1	2
1	56	3	7
2	53	4	5
3	48	4	4
4	60	5	5

S.I. = 14.6; d = 9.6%

Among the fixed launching rates the best results were obtained with $\gamma = \frac{TT_{wc}}{Q \times n}$ and $\gamma = \frac{TT_{s1}}{Q}$. Those two achieved the lowest throughput times, less idle time and smaller overall line lengths. $\gamma = \frac{TT_{wc}}{Q \times n}$ performed slightly better than $\gamma = \frac{TT_{s1}}{Q}$ in relation to idle time and throughput time, however $\gamma = \frac{TT_{s1}}{Q}$ resulted in smaller line lengths (exactly the opposite of what happened in another problem tested).

However, differences in throughput and idle time resulting from the use of $\gamma = \frac{TT_{s1}}{Q}$ instead of $\gamma = \frac{TT_{wc}}{Q \times n}$ were insignificant, whereas the differences in line length were not so insignificant. Therefore, and because the results are almost the same the use of $\gamma = \frac{TT_{s1}}{Q}$ for the launching rate is suggested.

For the line configuration considered, the Variable Launching Rate performed better in all the evaluation parameters, i.e. achieved smaller throughput time, less idle time, smaller line length and smaller number of incomplete units at the last station. It seems that in a line with open stations, if possible, the VLR should be the preferred launching system.

Among the three line balance solutions, the bowl allocation showed the best performance, although the difference was not very significant.

A scheme with closed stations (appendix A.2) was evaluated and again the bowl allocation performed the best among the three balance solutions. In this case the launching interval calculated by $\gamma = \frac{TT_{s1}}{Q}$ yielded the best results for throughput time, the best results for idle time being obtained by the use of $\gamma = \frac{TT_{wc}}{Q \times n}$. Dar-El (1978) stated that for a line composed of closed stations, if the models were well balanced over the stations, a variable launching interval would be more appropriate; if not, it was not evident which of the launching systems would be best. For the three balancing solutions, where the balance delay could be considered acceptable (d=9.6%), the VLR did not achieve the best results; it was preferable to use a fixed launching interval. It was also verified that the effect of closed stations is to increase the line length, this confirming what Dar-El (1978) had found.

The example used here seems to prove that an appropriate unbalance will in fact increase the mean throughput rate (because it will decrease the throughput time) and, therefore the conclusions drawn by Hillier and Boling (1979) and So (1989) may be extended to a manual assembly line working under paced conditions. Although the difference relative to the other two methods was not considerable, the bowl allocation showed an increase in the throughput time. If the results of this study can be generalized, it is possible to conclude that the appropriate unbalance of an assembly line will result in better performance (smaller throughput time, smaller number of incomplete units and smaller line length) over the balanced line. Note that the unbalanced solution did not always result in less idle time (it did for the RPW solution but not for the Largest Set Rule Solution). This is easily comprehensible; the line balance has the purpose of minimizing idle time, and therefore it is natural to expect that a nonbalanced line should result in an amount of idle time greater than the balanced solution.

Using the workload allocation given by the Largest Set Rule method, and considering the optimal Fixed Launching Rate system, the proposed sequencing method was compared with the Kilbridge and Wester (1963) and Time Spread sequencing methods. The proposed method showed better results than the other two, the worst results being given by the K&W method. This was probably due to the fact that the K&W method uses a different launching interval, and the solution provided by this method is based on that interval. In order to compare the three methods against the same reference, a new comparison was done. This time the launching interval used was calculated by the Kilbridge and Wester method. Again the proposed method performed better but now the K&W performed better than the TS - this can probably by explained by the fact that the TS results in a batch sequencing for this example.

8.2. Comparative Analysis of Sequencing Methods and Interpretation of the Results.

To evaluate the proposed sequencing method, an example presented by Bard et al. (1992), for which the optimal solution is known, was used. An early start schedule was assumed. The proposed method, the optimal solution, the TS method and the K&W method were compared. Again it should be mentioned that the K&W method is based on the K&W launching interval and therefore, the basis for comparison may not be the same. Trying to compensate for this aspect, the K&W method was also used with the same launching interval as the other three methods. Line configurations with open and closed stations were studied.

The optimal sequence that minimizes throughput time resulted in the same as the optimal sequence that minimizes line length for the case of open stations. For closed stations the optimal sequence that minimized throughput time was different than the optimal sequence that minimized line length. The objective of the proposed method and of the TS method is not directly to minimize throughput time and/or line length and is not affected by the station type. The Kilbridge and Wester method is also not directly dependent on station lengths however, it assumes that stations are open to the right (because work congestion is allowed). The proposed method and the Time Spread will provide a solution that is independent of the station type. Again concurrent work was not allowed. The data used for this comparison is given in Table 8.4.

Table 8.4: Data for the Test Problem ($Q_1 = 5$, $Q_2 = 3$, $Q_3 = 2$)

Station i	Station time (T_{si})	$T_{s_{ij}}$ 1	2	3
1	58	4	8	7
2	65	6	9	4
3	70	8	6	6
4	51	4	7	5

Total work content = 244

In order to compare the results obtained against the same conditions, the launching interval was the same as that used to obtain the optimal solution given by Bard et al., i.e. $\gamma = \dfrac{TT_{wc}}{Q \times n} = 6.1 \simeq 6.$

8.2.1 Open Stations

The results obtained for a line composed of open stations are shown in Tables 8.5 through 8.7.

Table 8.5: Sequence Solution for the Different Methods

Optimal Solution	Proposed Method	Time Spread	K&W
1	1	3	2
1	2	1	1
1	3	3	1
2	1	1	1
1	2	1	1
3	1	1	1
2	1	1	2
3	3	2	3
2	2	2	2
1	1	2	3

Table 8.6: Results Obtained with the Different Methods

	Idle Time	Throughput Time	Line Length
Optimal Solution	33	87	34
Proposed Method	32	89	36
Time Spread	37	90	36
K&W $\gamma = 6$	27	93	40
$\gamma = 8.8^{(1)}$	64	99	30

(1) calculated by $\frac{\sum_{j=1}^{m} Q_j T_{c_j}}{Q}$.

As can be seen, the proposed method is the closest to the optimal value for throughput time and achieves a line length close to the optimal value.

Using the optimal sequence (for $\gamma = 6$) and the proposed method, a new launching interval was calculated (given by $\gamma = \frac{TT_{s1}}{Q}$). The results were:

Table 8.7: Results Obtained for $\gamma = \frac{TT_{s1}}{Q}$

	Idle Time	Throughput Time	Line Length
Optimal Solution	30.6	86.4	35
Proposed Method	30.2	88.8	37.4

The throughput and idle time decreased with the use of this launching interval.

8.2.2. Closed Stations.

The optimal solution for closed stations depended on the objective, as seen in Table 8.8:

Table 8.8: Optimal Solution for the Different Objectives.

Minimize Throughput Time	Minimize Line Length
2	2
3	1
1	1
1	3
2	1
3	2
1	1
1	3
2	1
1	2

Table 8.9: Results Obtained with the Different Methods

	Idle Time	Throughput Time	Line Length
Optimal Sol.(2)	24 (24)	95 (96)	43 (42)
Proposed Method	24	95	43
Time Spread	27	106	52
K&W $\gamma = 6$	108	102	32
$\gamma = 8.8$	24	101	48

(2) the values X are for the sequence that minimizes throughput time; the values (X) are for the solution that minimizes line length.

8.3. Summary.

The conclusions that can be drawn from these comparisons are that the proposed method performed better than the TS and the K&W methods. Independently of the objective considered, the proposed method yielded results within 2% of the optimum throughput time and within 5% for the optimum line length. Those values seem perfectly acceptable for a method that does not seek the optimum. Remember that Dar-El (1978) and Bard et al. (1992) refer that it is practically the same to minimize line length or throughput time (this is particularly true when the number of stations is large). The proposed method, although not aiming directly to minimize these two parameters, is likely to provide solutions that are acceptably close to the optimum. The optimal solution may be impossible to reach by branch and bound algorithms or other type of methods seeking the optimum for large problems (which is the common situation in industry). The proposed method will also not give the optimum solution, but it is much simpler to apply and computation effort is smaller when compared with methods seeking the true optimal solution. The results demonstrate that the quality of the solution obtained by the proposed method is acceptable.

It was also seen that the launching interval should be the one given by $\gamma = \frac{TT_{s1}}{Q}$.

As expected, open stations provided better results than closed stations (i.e. smaller throughput time, less idle time, smaller line length, and also fewer incomplete items).

If concurrent work had been allowed, the result would be an increase in line performance. The example tested showed that for closed stations, the best performance is achieved with a fixed launching interval. In a line with open stations the best performance is achieved with a variable launching interval. Finally, the bowl allocation showed an improvement in the throughput time over the balanced line.

9. Conclusions and Suggestions for Further Research

9.1. Conclusions.

The following conclusions are drawn from this study of mixed-model assembly lines:

1.) Heuristics are still the best option for solving the line balancing and sequencing problems of mixed-model assembly lines. The size and complexity of the problems found in industry precludes the efficient use of algorithms that seek the optimal solution. These algorithms are usually sophisticated and the restrictions imposed on the problems may restrict their use.

2.) Heuristics available to solve the line balance of single-model lines can be extended to deal with the mixed-model assembly line. The use of task-group balancing will result in efficient solutions. Unevenness in assignments (models having different service times across the workstations) that may result from the task-group balancing (allocate work elements using total time per scheduled period) will have a harmless effect, provided that the sequence is carefully determined. Inappropriate sequences will significantly degrade the performance of the line.

3.) Heuristic methods also seem better than optimum-seeking algorithms to solve the sequencing problem. In general the solution obtained by such methods will be acceptable. Sequencing methods, in general, have a particular objective that depends on the line configuration, constraints imposed on the problem, etc. When possible, the most flexible line configuration should be chosen, i.e. open stations, concurrent work allowed, cross-training of operators, etc. As a general rule, the more flexible the assembly line, the better will be the assembly efficiency (less idle time, shorter throughput time, shorter line length, etc.).

4.) Open stations provide the best results for moving conveyor lines. If possible a variable launching interval should be used. For closed stations a fixed launching interval is preferred. In open stations greater launching intervals seem to decrease the line length but will increase idle time and throughput time.

5.) The evaluation study showed that the proposed method for solving the sequencing problem seems to be efficient and that the solution will be acceptably close to the optimal solution. The proposed method selects models to be launched at the position l in the sequence ($l=1, 2, \ldots, Q$) according to the following procedure: for each model remaining to be launched the difference is calculated between the average time to perform l units ($l \cdot \overline{T}_{si}$) and the actual time to perform l units ($X_{i,l}$) at station i ($i=1, 2, \ldots, n$). This difference is calculated for all workstations in the line, and the differences are added. The model selected to be launched at position l is the one that results in the least sum of the mentioned differences. In the examples tested the throughput time was within 2% of the optimum throughput time and line length was within 5% of the optimum line length. These values seem perfectly acceptable. This method can be easily implemented in a computer and should solve complex problems without major difficulties.

6.) The examples tested indicated that a bowl allocation of workloads to stations (thus creating an unbalanced line) will result in a slight improvement in the throughput rate over balanced mixed-model assembly lines. Therefore, an appropriate unbalance of the paced line will improve throughput over the balanced line. The same result was verified for an unbalanced and unpaced line.

7.) The use of the Minimum Part Set (MPS), which means a partition of the total production requirement into several smaller schedules seems to facilitate calculations, decrease inefficiencies and achieve faster line operation. When heuristic methods are used, the results obtained with the use of partitions of the total part set are better than the ones achieved with use of the total part set.

9.2. Suggestions for Further Research.

This study focused on paced manual assembly lines and in the comparison evaluation many simplifications were assumed. One of the simplifications was the use of deterministic task times. However, deterministic times are almost impossible to achieve in systems where human work is involved. Stochastic task times should be used in order to verify the validity of the conclusions obtained for deterministic times.

The proposed method for sequencing units onto the line needs to be extensively tested. A simulation study, involving a considerable number of test problems, should be done in order to effectively validate the method (or perhaps discredit it). Because this method only depends on average station times, for cases where the stations lengths are fixed, and station passage times are very different from the average station times, the performance of the method may suffer. This aspect should be investigated.

The comparison study revealed that for a line composed of closed stations, the best launching system is to use a fixed launching interval. Dar-El (1978) stated that when models are approximately well balanced over the workstations, the best launching system is the VLR. The case studied in this thesis showed that for an acceptable balance delay the best solution was to use the FLR. Thus we have a contradiction, and a more thorough study of this aspect should be done. Particularly several line balances with different values for balance delay should be investigated.

The study showed that a bowl allocation of workloads resulted in a decrease in throughput time (for the problem tested). However, this decrease was not considerable, and therefore, it is questionable whether the effort to find the optimal bowl allocation should be taken, or if it is just better to balance the line. A simulation of unbalanced lines, using a significant number of test problems, should be carried out to confirm that this allocation will result in an improvement in throughput, and whether this improvement justifies the bowl allocation. The effect of stochastic task times on the unbalance of the line should also be analyzed.

When the Minimum Part Set is being used, an effective method of dealing with the incomplete items should be devised. When the first set of products is finished, a residue of units may have been left. These will need to be completed in the next run or shift. What steps should be taken to minimize the possibility of residuals? A possible objective for a sequencing method could be to minimize the residuals.

Another aspect that may be interesting to explore is for what degree of dissimilarity (between the models assigned to the same assembly line) the line becomes inefficient; in short, how the line efficiency depends on the similarity index.

This research did not consider the assembly line as being part of a multi-level production system. Unbalancing the assembly line will certainly affect the other production levels. For a multi-level, just-in-time production system, unbalancing the assembly line may prove to be harmful for the precedent levels. This aspect may be of some interest.

Due to advances in computer technology, methods that seek the optimal solution (in line balancing as well as in sequencing) are more likely to be used today than they were at the time that the majority of the studies in mixed-model assembly lines were performed. A cost/effort comparison between heuristic methods and optimization methods may be of interest.

Appendix A.1

Results for the application of different launching intervals to example presented in chapter 5. The sequencing method used was the one proposed in chapter 6, and the obtained sequence was {1 2 1 2 1 2 1 1 2 1 2 1}. Line composed of open stations. The last column (Incomp) refers to the number of units that will be incomplete in the last station (because the shift finished). The number X is the time at which the operator at the last station would finish the first incomplete unit.

Largest Set Rule Solution

	Idle Time (min)	Throughput (min)	Line (m)	Incomp
VLR	19.0	76.0	21.0	4 (61)
$\gamma = \frac{\max TT_{si}}{Q} = 5$	30.0	76.0	21.0	4 (61)
$\gamma = \frac{TT_{s1}}{Q} = 4.58$	20.5	76.0	25.6	4 (61)
$\gamma = \frac{TT_{wc}}{Q \times n} = 4.52$	19.4	76.0	26.3	4 (61)
$\gamma_{K\&W} = 6.25$	81.3	88.5	21.0	5 (62.3)
$\gamma = \frac{TT_{s1} - T_{11}}{Q-1} = 4.55$	19.8	76.0	26.0	4 (61)

RPW Solution

	Idle Time (min)	Throughput (min)	Line (m)	Incomp
VLR	27.0	75.0	21.0	3 (65)
$\gamma = \dfrac{\max TT_{si}}{Q} = 5$	38.0	77.0	22.0	4 (61)
$\gamma = \dfrac{TT_{s1}}{Q} = 4.67$	33.7	76.7	25.3	4 (60.7)
$\gamma = \dfrac{TT_{wc}}{Q \times n} = 4.52$	33.1	76.5	26.8	4 (60.5)
$\gamma_{K\&W} = 6.25$	88.0	89.5	21.0	5 (63.25)
$\gamma = \dfrac{TT_{s1} - T_{11}}{Q-1} = 4.82$	35.0	76.8	23.8	4 (60.8)

Bowl Allocation

	Idle Time (min)	Throughput (min)	Line (m)	Incomp
VLR	20.0	74.0	21.0	3 (64)
$\gamma = \dfrac{\max TT_{si}}{Q} = 5$	34.0	76.0	21.0	4 (61)
$\gamma = \dfrac{TT_{s1}}{Q} = 4.67$	26.7	75.7	24.3	4 (60.7)
$\gamma = \dfrac{TT_{wc}}{Q \times n} = 4.52$	26.1	75.5	25.8	4 (60.5)
$\gamma_{K\&W} = 6.25$	84.0	88.5	21.0	5 (62.25)
$\gamma = \dfrac{TT_{s1} - T_{11}}{Q-1} = 4.82$	29.5	75.8	21.0	4 (60.8)

Appendix A.2

Results for the application of different launching intervals to example presented in chapter 5. The sequencing method used was the one proposed in chapter 6, and the obtained sequence was {1 2 1 2 1 2 1 1 2 1 2 1}. Line composed of closed stations. The last column (Incomp) refers to the number of units that will be incomplete in the last station (because the shift finished). The number X is the time at which the operator at the last station would finish the first incomplete unit.

Largest Set Rule Solution

	Idle Time (min)	Throughput (min)	Line (m)	Incomp
VLR	9.0	79.0	29.0	4 (64)
$\gamma = \frac{TT_{s1}}{Q} = 4.58$	9.6	79.1	28.7	4 (64.1)
$\gamma = \frac{TT_{wc}}{Q \times n} = 4.52$	8.4	79.4	29.7	4 (64.4)
$\gamma_{K\&W} = 6.25$	75.8	90.8	22.0	5 (65.7)

RPW Solution

	Idle Time (min)	Throughput (min)	Line (m)	Incomp
VLR	14.0	81.0	29.0	5 (62)
$\gamma = \frac{TT_{s1}}{Q} = 4.67$	14.7	80.3	29.0	5 (61.7)
$\gamma = \frac{TT_{wc}}{Q \times n} = 4.52$	12.8	82.7	32.9	6 (60)
$\gamma_{K\&W} = 6.25$	74.8	90.8	22.0	5 (65.8)

Bowl Allocation

	Idle Time (min)	Throughput (min)	Line (m)	Incomp
VLR	13.0	80.0	29.0	4 (65)
$\gamma = \frac{TT_{s1}}{Q} = 4.67$	11.7	77.3	26.0	4 (62.3)
$\gamma = \frac{TT_{wc}}{Q \times n} = 4.52$	8.5	78.3	28.6	4 (63.3)
$\gamma_{K\&W} = 6.25$	74.8	89.8	21.0	5 (64.8)

List of References

ARCUS, A.L., 1966, "Comsoal, A Computer Method of Sequencing Operations for Assembly Lines", Int. J. Prod. Res., Vol. 4, No. 4, (cited in Macaskill 1973)

BANCROFT, W.E., 1987, A Scheduling Sequencing Algorithm for Mixed-Model Assembly Lines, Unpublished Ph.D. Dissertation, Illinois Institute of Technology

BARD, J.F., DAR-EL, E.Z. and SHTUB, A., 1992, "An Analytic Framework for Sequencing Mixed Model Assembly Lines", Int. J. Prod. Res., Vol. 30, No. 1, pp. 35-48

BOLAT, A., 1988, Generalized Mixed Model Assembly Line Sequencing Problem, Unpublished Ph.D. dissertation, The University of Michigan

BOOTROYD, G., POLI, C. and MURCH, L.E., 1982, Automatic Assembly, Marcel Dekker & Inc., New York

BUFFA, E.S., 1980, Modern Production/Operations Management, 6th Ed., John Wiley & Sons, New York

BUXEY, G.M., SLACK, N.D. and WILD, R., 1973, "Production Flow Line Systems Design -A Review", AIIE Transactions, Vol. 5, No, 1, pp. 37-47

BUZACOTT, J.A., 1990, "Abandoning the Moving Assembly Line: Models of Human Operators and Job Sequencing, Int. J. Prod. Res., Vol. 28, No. 5, pp. 821-839

DAR-EL, E.M., 1973, "MALB - A Heuristic Technique for Balancing Large Single-Model Assembly Lines", AIIE Transactions, Vol. 5, No. 4, pp. 343-356

DAR-EL, E.M., 1975, "Solving Large Single-Model Assembly Line Balancing Problems - A Comparative Study", AIIE Transactions, Vol. 7, No. 3, pp. 302-310

DAR-EL, E.M. and COTHER, F., 1975, "Assembly Line Sequencing for Model Mix", Int. J. Prod. Res., Vol. 13, No. 5, pp. 463-477

DAR-EL, E.M. and CUCUY, S., 1977, "Optimal Mixed-Model Sequencing for Balanced Assembly Lines", OMEGA, Vol. 5, No.3, pp. 333-342

DAR-EL, E.M., 1978, "Mixed-Model Assembly Line Sequencing Problems", OMEGA, Vol. 6, No.4, pp. 313-323

DAR-EL, E.M. and NAVIDI, A., 1981, "A Mixed-Model Sequencing Application", Int. J. Prod. Res., Vol. 19, No. 1, pp. 69-84

DAR-EL, E., 1986, Productivity Improvement: Employee Involvement and Gainsharing Plans, Elsevier, New York

DEUTSCH, D.F., 1971, A Branch and Bound Technique for Mixed-Model Assembly Line Balancing, Unpublished Ph.D. Dissertation, Arizona State University

DOWNEY, B.S. and LEONARD, M.S., 1992, "Assembly Line with Flexible Work-Force", Int. J. Prod. Res., Vol. 30, No. 3, pp. 469-483

GROOVER, M.P., 1987, Automation, Production Systems, and Computer-Integrated Manufacturing, Prentice-Hall, Englewood Cliffs, New Jersey

HELD, M., KARP, R.M. and SHARESIAN, R., 1963, "Assembly-Line Balancing-Dynamic Programming with Precedence Constraints", Operations Research, Vol. 11, No. 3, (cited in Mastor 1970)

HEGELSON, F. and BERNIE, D.P., 1961, "Assembly Line Using the Ranked Positional Weight Technique", J. Ind. Eng., Vol. 12, No. 6, pp. 244-254

HICKS, C.R. and YOUNG, H.H., 1962, "A Comparison of Several Methods for Determining The Number of Readings in a Time Study" J. of Ind. Eng., Vol. 13, No. 2, (cited in Moodie and Young 1965)

HILLIER, F. and BOLING, R., 1979, "On the Optimal Allocation of Work in Symmetrically Unbalanced Production Line Systems with Variable Operation Times", Management Science, Vol. 25, No. 8, pp. 721-728

HOFFMAN, T., 1990, "Assembly Line Balancing: A Set of Challenging Problems", Int. J. Prod. Res., Vol. 28, No. 10, pp. 1807-1815

KAO, C.C., 1981, An Investigation of the Mixed-Model Assembly Line Sequencing Problem, Unpublished Ph.D. dissertation, Texas Tech University

KILBRIDGE, M. and WESTER, L., 1961, "A Heuristic Method of Assembly Line Balancing", J. of Ind. Eng., Vol. XII, No.4, pp. 292-298

KILBRIDGE, M. and WESTER, L., 1963, The Assembly Line Mixed-Model Sequencing Problem, Proc. of the Third International Conference on Operations Research, Dunod Editeur, Paris

MACASKILL, J.L.C., 1972, "Production-Line Balances For Mixed-Model Lines", Management Science, Vol. 19, No. 4, pp.423-434

MACASKILL, J.L.C., 1973, "Computer Simulation for Mixed-Model Production Lines", Management Science, Vol. 20, No. 3, pp.341-348

MANSOOR, E.M. and BEN-TUVIA, S., 1966, "Optimizing Balanced Assembly Lines", J. of Ind. Eng., Vol. XVII, No. 3, pp. 126-132

MASTOR, A.A., 1970, "An Experimental Investigation and Comparative Evaluation of Production Line Balancing Techniques", Management Science, Vol. 16, No. 11, pp. 728-746

MILTENBURG, G.J., 1987, "A Theoretical Basis for Scheduling Mixed-Model Lines for Just-In-Time Production Systems", Management Science, Vol. 35, No. 2, pp. 192-207

MILTENBURG, G.J. and SINNAMON, G., 1989, "Scheduling Mixed-Model Multi-Level Just-In-Time Production Systems", Int. J. Prod. Res., Vol. 27, No. 9, pp. 1487-1509

MONDEN, Y., 1983, Toyota Production System, Industrial Engineering and Management Press, Norcross, Georgia

OKAMURA, K. and YAMASHINA, H., 1979, "A Heuristic Algorithm for the Assembly Line Model-Mix Sequencing Problem to Minimize the Risk of Stopping the Conveyor", Int. J. Prod. Res., Vol. 17, No. 3, pp. 233-247

PRENTING, T. O. and THOMOPOULOS, N. T., 1974, Humanism and Technology in Assembly Systems, Hayden Book Company, Inc, Rochelle Park, New Jersey

SNEIDER, R.M., 1980, A Methodology for Optimal Assembly Line Balancing, Unpublished Ph.D. Dissertation, Purdue University

SO, K.C., 1989, "Efficiency of Unbalancing Production Lines", Int. J. Prod. Res., Vol. 27, No. 4, pp. 717-729

SUMICHRAST, R.T., RUSSEL, R.R., TAYLOR, B.W., 1992, "A Comparative Analysis of Sequencing for Mixed-Model Assembly Lines in a Just-In-Time Production System", Int. J. Prod. Res., Vol. 30, No. 1, pp. 199-214

TONGUE, F. 1965 "Assembly Line Balancing Using Probabilistic Combinations of Heuristics", Management Science, Vol 11, No. 7, pp. 727-235

VILLA, C., 1970, Multi Product Assembly Line Balancing, Unpublished Ph.D.

Dissertation, University of Florida

WANG, F. and WILSON, R., 1989, "Comparative Analyses of Fixed and Removable Item Mixed Model Assembly Lines", IIE Transactions, pp.313-317

WILD, R., 1972, Mass Production Management, John Wiley, Chichester, UK.

Vita

Carlos João L. de Lima Fernandes was born May 8, 1965, in Oporto, Portugal. In August 1990, he received the "Licenciatura" degree in Mechanical Engineering from the University of Oporto. After a summer internship at the Polytechnic University of Lublin, Poland, he enrolled in January 1991 the Manufacturing Systems Engineering program at Lehigh University.

END OF

TITLE