ORDER-DRIVEN FLEXIBILITY MANAGEMENT IN MAKE-TO-ORDER COMPANIES WITH FLEXIBLE SHOPS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BEKİR İLKER SÜER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

SEPTEMBER 2009

Approval of the thesis:

ORDER-DRIVEN FLEXIBILITY MANAGEMENT IN MAKE-TO-ORDER COMPANIES WITH FLEXIBLE SHOPS

submitted by **BEKİR İLKER SÜER** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan ÖZGEN
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Nur Evin ÖZDEMİREL
Head of Department, **Industrial Engineering** _______________

Prof. Dr. Sinan KAYALIGİL
Supervisor, **Industrial Engineering Dept., METU** _______________

Examining Committee Members:

Prof. Dr. Ömer KIRCA
Industrial Engineering Dept., METU _______________

Prof. Dr. Sinan KAYALIGİL
Industrial Engineering Dept., METU _______________

Prof. Dr. Meral AZİZOĞLU
Industrial Engineering Dept., METU _______________

Assoc. Prof. Dr. Haldun SÜRAL
Industrial Engineering Dept., METU _______________

Prof. Dr. İhsan SABUNCUOĞLU
Industrial Engineering Dept., Bilkent University _______________

Date: **11.09.2009**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name	: Bekir İlker SÜER
Signature	:

ABSTRACT

ORDER-DRIVEN FLEXIBILITY MANAGEMENT IN MAKE-TO-ORDER COMPANIES WITH FLEXIBLE SHOPS

SÜER, Bekir İlker
M.S., Department of Industrial Engineering
Supervisor: Prof. Dr. Sinan KAYALIGİL

September 2009, 125 pages

In this study, an operational (short term) flexibility management approach is proposed for make-to-order companies with flexible shops. Order Review and Release (ORR) techniques and typical Flexible Manufacturing System (FMS) decisions are combined in this method. The proposed method prepares a shop environment by allocating process and routing flexibility types at different levels to the shop in each production cycle. Variety, volume, and criticality of the part types in the pool and the anticipated orders constitute the main inputs for flexibility allocation. A flexibility management policy is introduced and determination of the proper policy is realized with the integrated utilization of mathematical programming and simulation modeling. An experimental study is performed to investigate the effects of proposed method on a hypothetical flexible shop. Results show that with an appropriate policy, periodical and online flexibility management can be an effective tool to cope with uncertainty in demand if combined with ORR techniques.

Keywords: Make to Order, Flexible Manufacturing System, Flexibility Allocation, Order Review and Release, Mathematical Programming

ÖZ

ESNEK ATÖLYELERE SAHİP SİPARİŞE ÜRETİM YAPAN FİRMALARDA SİPARİŞE DAYALI ESNEKLİK TAHSİSİ

SÜER, Bekir İlker
Yüksek Lisans, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Sinan KAYALIGİL

Eylül 2009, 125 sayfa

Bu çalışmada, esnek atölyelere sahip siparişe üretim yapan firmalar için bir operasyonel (kısa vadeli) esneklik yönetimi yaklaşımı önerilmektedir. Bu yöntemde, sipariş inceleme ve sürme teknikleri ile tipik esnek imalat sistemi kararları birleştirilmiştir. Önerilen yöntem, her üretim döneminde atölyeye farklı seviyelerde işlem ve rotalama esneklikleri tahsis edilmesiyle bir atölye ortamı hazırlamaktadır. Sipariş havuzunda bulunan ve gelmesi öngörülen parça tiplerinin çeşitliliği, hacmi ve kritikliği, esneklik tahsisi için ana girdileri oluşturmaktadır. Bir esneklik yönetimi politikası tanıtılmış ve uygun politikanın belirlenmesi matematiksel programlama ve benzetim modellemesinin ortak kullanımı yoluyla gerçekleştirilmiştir. Deneysel bir çalışma yapılarak, önerilen yöntemin kuramsal bir esnek atölye üzerindeki etkileri incelenmiştir. Sonuçlar, uygun bir politika izlenen dönemsel ve çevrimiçi esneklik yönetiminin, sipariş inceleme ve sürme teknikleri ile birleştirildiğinde talepteki belirsizlik ile başa çıkmak konusunda etkin bir araç olabileceğine işaret etmektedir.

Anahtar Kelimeler: Siparişe Üretim, Esnek İmalat Sistemi, Esneklik Tahsisi, Sipariş İnceleme ve Sürme, Matematiksel Programlama

To My Mother

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Prof. Dr. Sinan Kayalıgil for his continuous encouragement, guidance, support and insight throughout this thesis study. I consider myself privileged to have a mentor like Prof. Dr. Sinan Kayalıgil with the strong work ethic, never-ending knowledge, unyielding patience and tolerance.

I would like to thank to my colleague Özer Tekinşen for his support. I would also like to thank to Ziya Çiftçi and Hüseyin Dengizek for their understanding and tolerance.

I wish to thank the public institution, The Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting me with the National Scholarship Programme throughout this thesis study.

My special thanks go to Tuğba for her encouragement, support and many other things. And finally, I am deeply in debt and grateful to my mother Beyhan Özkan for her endless love and moral support she gave me throughout my life. I am so lucky to have a strong and invincible mother like her.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Make-to-order refers to a manufacturing policy where the products are produced only after they are ordered. This policy is known to have several advantages such as supplying the product to the customer with the exact requirements and reduced finished goods inventory. However, make-to-order policy has drawbacks such as long manufacturing lead times and unavoidable high set-up costs. Moreover, this policy is known to need care in addressing the responsiveness to the demand fluctuations in the market.

Order review and release (ORR) techniques have been developed in order to cope with uncertainty in demand and to have a reasonable level of utilization of the shop while decreasing congestion on the shop floor. These techniques create a buffer zone between job arrivals and job release to the shop floor for better control. Orders can be accepted or rejected based on the shop conditions. Accepted orders are collected in a pre-shop pool and prioritized according to some rule before they are released for production. This activity is performed in a controlled manner by determining when an order is to be sent to the shop floor and which order is to be sent.

Although, ORR techniques can control the flow of orders and reduce congestion on the shop floor, the "releasability" of the parts is another factor to consider. This is related to the flexibility of the shop. The more flexible the shop, the more it can produce different part types and the better it can respond to orders with varying volumes. Flexible manufacturing systems (FMS) are highly suitable for this purpose.

FMSs consist of programmable and reconfigurable machines that are linked by an automated material handling system. This integrated system can be configured to serve a wide variety of part types in different volumes of orders.

Recent studies show that the flexibility can be handled in an operational sense. Rather than offline solutions to order flow in shop, it is suggested to provide the necessary flexibility and let the real-time dispatching govern how orders will flow.

The main purpose of this thesis is to combine ORR with a flexible shop environment by controlling the flexibility at discrete instants and in a cyclical manner. A Flexibility Management (FM) approach integrated with the order release is proposed by allocating two different competing flexibility types dynamically. Our view is to affect "releasability" of the existing pool orders through configuring an appropriate flexibility mix. This (flexibility control rather than capacity control) can be considered an added dimension to output control in make-to-order manufacturing. Our approach consists of a mathematical model and a simulation model for allocating flexibility to the system and testing the effectiveness of the applied FM policy. After the determination of the appropriate policy for a shop with this method, the policy can be considered for application in real shop environments with proper parameterization in every shop case.

In this study, process and routing flexibilities are selected to respond to uncertainties mentioned in the literature. As these flexibility types can be adjusted dynamically by changing only operation-machine assignments in the shop, they are suitable for a routine operational management control concept. Different types of measures are defined in order to quantify and manage flexibility. The required flexibility types and levels are tied to the variety of the part types, their current volumes and urgency levels in the pre-release pool. Moreover, estimated job arrivals are also considered. Thus, flexibility is allocated to the parts according to their priorities.

An experimental study is performed to investigate the impacts of flexibility management approach. Different measures are utilized to monitor the performance of the FM policies. The outcomes of the study are analyzed and some of the characteristics of the proposed approach are revealed.

This thesis study is organized as follows: Chapter 2 covers the survey of relevant literature. In Chapter 3, flexibility measures are defined, a general view of the approach is given and the elements that are used for flexibility allocation are mentioned. In Chapter 4, the mathematical model and the simulation model are presented, and the interactions between them are explained. Chapter 5 is dedicated to the experimental study where the settings of the experiment and performance measures are mentioned, and the results of the experiment are discussed. Finally, in Chapter 6, concluding remarks on the thesis study and some suggestions for future research directions are presented.

CHAPTER 2

LITERATURE SURVEY

In this section, a detailed literature survey is presented to reflect recent developments in ORR strategies, make-to-order shops, FMSs and flexibility concept.

This study was initiated with an investigation of ORR concepts in the literature. Although, there are several approaches for order release, this study focused on workload control (WLC) in the shop.

Wisner [1] reviewed the relevant literature on the order release problem. Related research was categorized into three topics: Descriptive/empirical studies, analytical studies, and simulation based studies.

Descriptive research section of Wisner [1] included general discussion papers, case studies and survey research. It was signified that there is a controversy about the benefits of controlled release in the descriptive research area. Although, some of the reviewed articles in this section mentioned that controlled release results in lowering inventory costs, WIP costs and tardiness, other articles mentioned the risk of increasing flow times by delaying the release of the jobs. Necessity of further research was emphasized in this section.

Analytical research section of Wisner [1] included optimization techniques in different articles. It was concluded that the techniques can be applied to machine shop models in order to minimize costs resulting from job delays. However, the author also highlighted the static shop characteristics used in the application of these techniques. It was mentioned that these techniques can become less effective in the face of dynamism in order arrivals and shop conditions.

Simulation based research section of Wisner [1] included various studies. The studies mainly compared immediate release to controlled release methods. The benefits resulting from either the controlled or immediate release methods were shown. However, necessity of modeling realistic shop conditions was suggested, as is the case for analytical research section.

Bergamaschi et al. [2] reviewed and classified existing ORR work published in major journals from 1970 through 1997. The oldest ORR work in publications was Wight [3]. Wight [3] (as cited in Bergamaschi et al. [2]) examined manufacturing problems in make-to-order companies. It was found that effective planning and input/output control of the plant can be used to deal with the problems.

Bergamaschi et al. [2] mentioned that majority of the research on the subject of ORR from Wight [3] to date was dedicated to workload balancing among machines and the timing convention. It was pointed out that sufficient research has not been made on "capacity planning". Most of the papers reviewed in this study had considered the capacity of the system as given and beyond management control. Moreover, only 3 out of 18 papers had considered an "extended" schedule visibility, which corresponds to allowing a reduction in shop performance in the current period for an advantage in the future. In conclusion, the authors mentioned that ORR effectiveness can be improved through "active capacity planning" and "extended schedule visibility", and they suggested these topics for future research.

Land and Gaalman [4] discussed the "Workload Control" (WLC) concept, which was started in the early 1990s. The main tool of the WLC concept was the controlled job release (based on various functions of the existing workload in the shop) from the order pool to the shop. The jobs were allowed to be released if doing so would not cause exceeding predefined norms computed specifically for the workstations. According to the authors, every workload norm was based on a series of assumptions such as stationary shop floor and stationary order pool, and hence was open to be questioned.

Land [5] explored the effects of workload threshold levels (norms) and control parameters on the performance of WLC concept. According to the results of the study, performance of the shop increased when the workloads were calculated on station basis and if the planned station throughput times were determined correctly. The workload balance between stations improved when the orders in the pool were evaluated frequently and release was performed in real-time. Results of this study also showed that the selection of an appropriate norm level, which is neither too high nor too low, improves system performance by reducing throughput times.

Cigolini and Portioli [6] investigated the effects of workload limiting policies on shop performances by conducting a simulation study. Three workload control methods were considered: Setting an upper bound for workload at every machine; setting upper and lower bounds for workload at every machine; minimization of overloads / underloads at each machine. Results showed that "upper bound only" method is the best while the "upper and lower bound" method is the worst in terms of overall performance measures. Although, the "workload balancing" method yielded similar results with the "upper bound only" approach, it was shown that "workload balancing" method is less affected by changes in the experimental factors.

Sabuncuoglu and Karapinar [7] proposed an ORR method that utilizes job due date and shop load information. The method was called DLR (Due date and Load-based Release). The main aim of the proposed algorithm was to finish the jobs on time while keeping the shop load at reasonable levels. They conducted simulation studies and compared the performance of DLR with other ORR methods. It was shown that using both the load and due date information simultaneously improves system performance. It had also been shown that DLR is the most robust method among the compared methods, at varying levels of system load and processing times. They pointed out that releasing jobs to the system in a controlled manner is a better policy than immediate release of jobs to the system. Moreover, it was also highlighted that shop performance improves when a due-date-oriented dispatching rule is integrated with ORR methods.

Kingsman and Hendry [8] examined the effects of capacity adjustment on WLC concept. Thus, these authors explored output control besides the classical input control by conducting simulation experiments. The output control was realized by either reallocation of workers, allowing overtime or both allowing reallocation and overtime at the same time. This control can also be seen as a form of flexibility application. It was shown that by reallocating the workers in a flexible manner according to the workload in the system, manufacturing lead times are lowered and throughput of work is increased.

As can be seen from the relevant literature that order release techniques have major effects on shop performance. Controlled release, varying capacity and online job flow control have been advocated. On the other hand, as the visibility for forthcoming orders is limited, preparation by setting the proper shop parameters in response to uncertain events constitutes an essential issue. This matter is related to flexibility.

Flexibility in manufacturing has been made an issue since mid 80s. Many studies have been published about the categorization, definition and quantification of flexibility concepts. Sethi and Sethi [9] reviewed the relevant literature on flexibility in manufacturing and defined 11 different types of flexibility. They also explained the means to obtain each of those flexibility types and suggested measures to quantify each of them according to the approaches of several authors. The types of flexibility, as defined in Sethi and Sethi [9], are:

- Machine Flexibility: “Machine flexibility (of a machine) refers to the various types of operations that the machine can perform without requiring a prohibitive effort in switching from one operation to another.”
- Material Handling Flexibility: “Flexibility of a material handling system is its ability to move different part types efficiently for proper positioning and processing through the manufacturing facility it serves.”
- Operation Flexibility: “Operation flexibility of a part refers to its ability to be produced in different ways.”
- Process Flexibility: “Process flexibility of a manufacturing system relates to the set of part types that the system can produce without major setups.”
- Product Flexibility: “Product flexibility is the ease with which new parts can be added or substituted for existing parts.”
- Routing Flexibility: “Routing flexibility of a manufacturing system is its ability to produce a part by alternate routes through the system.”
- Volume Flexibility: “Volume flexibility of a manufacturing system is its ability to be operated profitably at different overall output levels.”
- Expansion Flexibility: “Expansion flexibility of a manufacturing system is the ease with which its capacity and capability can be increased when needed.”
- Program Flexibility: “Program flexibility is the ability of the system to run virtually untended for a long enough period.”

- Production Flexibility: "Production flexibility is the universe of part types that the manufacturing system can produce without adding major capital equipment."
- Market Flexibility: "Market flexibility is the ease with which the manufacturing system can adapt to a changing market environment."

After the descriptions, the authors established links between different flexibility types. They categorized flexibility types into three groups, which are component (basic) flexibilities, system flexibilities and aggregate flexibilities. The linkages are illustrated in Figure 2.1. In the figure, the relations between different types of flexibility are shown. System flexibilities depend on basic flexibilities, and aggregate flexibilities are affected from various types of system flexibilities. This indicated that manufacturing strategy of a firm can only be realized if related flexibility levels are acquired. The organizational structure of a company and microprocessor technology were considered as the factors that underlie all flexibility types. The authors also pointed out the existence of tradeoffs between various flexibility types and other factors such as productivity, quality and degree of automation.

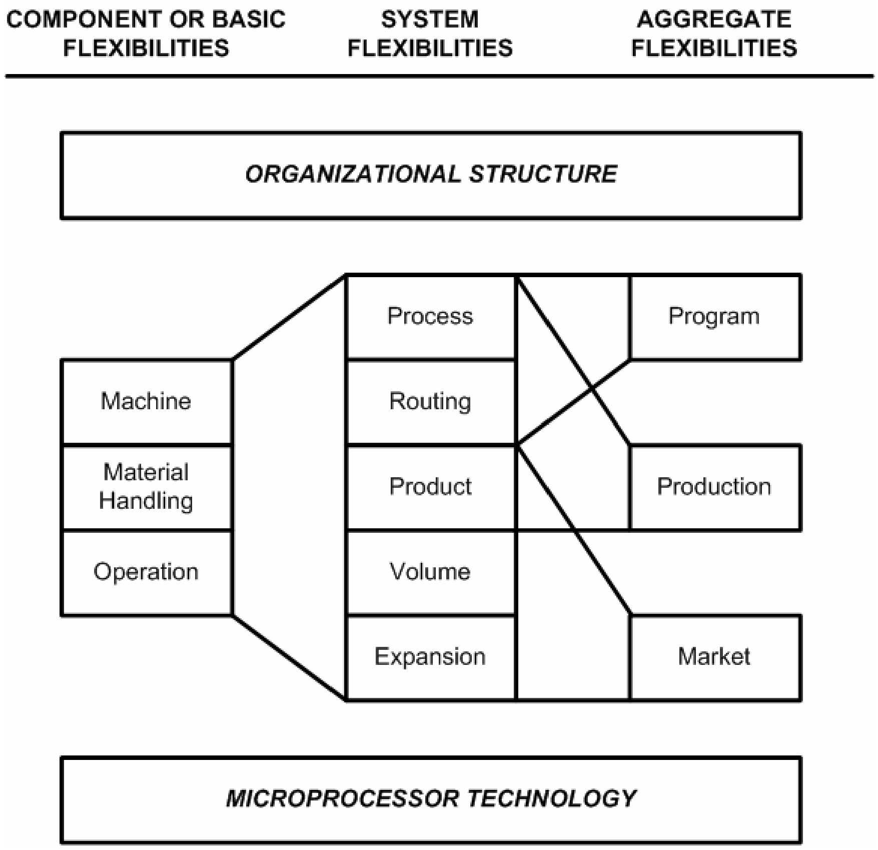


Figure 2.1 Linkages between Flexibility Types, (reproduced from Sethi and Sethi [9])

Stecke [10] handled the tool management and loading problems of an FMS. In this paper, issues like aggregate planning, machine grouping, machine loading, fixture management, scheduling and inventory management were covered. For the machine loading problem, a mathematical model was developed to allocate the operations to machines while balancing the workload between machines and providing redundancy for each operation. The routing flexibility (redundancy) was to be met by a constraint set. The flexibility level was determined by a preset parameter, which adjusted number of machines that an operation is assigned, and this same level was imposed to all operation types. It can be argued that, in this study, the routing flexibility concept is taken implicitly as given. The concept is not considered as a decision criterion that is to be optimized.

Lingayat et al. [11] developed an order release mechanism (ORM) for a flexible flow system (FFS). It was concluded from this study that routing flexibility is a key system characteristic besides the bottleneck machines. According to the authors, making routing decisions at the time of release using real-time information instead of at the pre-release planning stage using static information, minimizes the possibility of any machine being idle. Their release mechanism included the effect of flexibility and yielded a better performance than no release control.

Newman and Maffei [12] examined the effects of routing flexibility, order release mechanisms based on aggregate shop load, and sequencing rules. The results of this simulation based study showed that increasing routing flexibility improves job shop performance by decreasing flow times of orders. It was shown that flexibility outperforms other operational management tools in all performance measures. The reported results signified explicitly that as flexibility is increased, the marginal rate of performance improvement decreases. However, the authors concluded the subject by notifying that while it is a very powerful tool to cope with uncertainty, flexibility can be expensive and it may also hide organizational problems.

Chan et al. [13] analyzed the effects of increasing flexibility in systems with different physical and operational characteristics (processing time, transportation time, machine setting time, control strategies, tooling cost, performance of scheduling rule, etc.). The focus of this paper was to show that an increase in flexibility is not always advantageous. It can also get disadvantageous under different circumstances. A simulation study was conducted for this purpose. Routing flexibility was chosen as the flexibility type. It was shown that, maximizing flexibility level does not always improve system performance, it may also deteriorate performance. The main cause of this was attributed to increase in total processing times on alternative machines which is a combination of processing time, loading/unloading time, tool changing time, transportation time, machine setting time, etc.

Calvo et al. [14] mentioned the need of integrating flexibility into the system for successful decision making at operational (short-term) level. In their study, a utility function was defined in terms of cost, quality and time. Then, flexibility of production was defined as the temporal ratio of incremental utility of a pair of states. Using these concepts, flexibility was taken as a tool for decision-making in a mathematical model. In this model, the utility function was the objective to be maximized. They conducted a case study based on real-life data and showed that each flexibility type become more effective under different market conditions. It was concluded that a measure of flexibility should reflect the operational and strategic characteristics of manufacturing systems and it should be related with uncertainty caused by the quickly changing market environment.

Nomden and Zee [15] studied the effects of routing flexibility on virtual cellular manufacturing (VCM). In their flexibility level measure, only the "extra" routes (or machines) were counted. The results showed that high levels of flexibility do not improve performance of the system measured in terms of mean flow times, even deteriorates the performance. A low level was found to be sufficient since most of the benefits would already have been realized at that level.

Daniels et al. [16] examined a flow shop environment with partially skilled workers. They explored the operational benefits that can be achieved with the amount of resource flexibility present within the system and allocation of flexibility among the workforce members. They had introduced a "skill matrix", which stores workers' cross-training skills. According to the computational experiments in this paper, they concluded that consistent distribution of partial flexibility among the workers over the stations results in excellent system performance. This balanced situation was identified as a "chain" by the authors.

It was also concluded that allocating flexibility arbitrarily could result in poorer operational performance than that achieved by an inflexible system.

Wahab and Stoyan [17] developed measures for machine flexibility and routing flexibility. They mentioned the dynamic behavior of manufacturing systems by categorizing the sources of changes as either internal or external. It was shown that the suggested measures respond to the changes in the system more effectively than other measures in literature. This was attributed to the fact that the suggested measures capture several attributes of manufacturing systems that were not considered in other measures, such as the efficiency of a machine to process an operation, the efficiency of alternative routes, the availability of alternative routes, etc.

Chang [18] extended the model of measuring “single machine flexibility” into “machine-group flexibility”. It was emphasized that three attributes should be taken into account in order to correctly measure the “machine-group flexibility”. These attributes were “Versatility”, “Efficiency” and “Redundancy”. As a result, flexibility was treated as a function of these three attributes. It was also mentioned in this paper that the efficiency and versatility of the system are simultaneously improved if a system has the ability to learn to dampen the effects of disturbances that result from introducing new tasks into the system.

So far, we have reported on some relevant articles on flexibility in relation to performance outcome impacts. It can be seen that it is often suggested not to increase flexibility to its full limits to better the performance and flexibility is taken not simply as a static measure of strategical nature for potential to respond to arbitrary changes. Work on make-to-order and flexibility connection is rather new. From this point on, we review pertinent work.

Henrich et al. [19] emphasized that the semi-interchangeability of machines in the context of WLC had not been researched. Semi-interchangeability was defined as “the ability of the machines to perform similar operations”. In this paper, a simulation study was conducted to see the effects of different alternatives for capacity groups (one norm / separate norms), routing decisions (at order release / at dispatching), degree of interchangeability (from non-interchangeability to full interchangeability at different levels), and workload norms (at different tightness levels). The results showed that in a controlled release environment, interchangeability (routing flexibility) is a major factor that decreases total throughput time. However, it was also shown that increasing interchangeability results in decreasing marginal performance improvements. The empirical results also pointed to the importance of making routing decisions at the dispatching moments (not in advance, in the form of loading parts) and using separate norms for each individual machine.

Corsten et al. [20] examined the order release models in terms of flexibility aspects and developed the structure of a decision model for flexibility-driven order release. The authors suggested a flexibility oriented approach emphasizing the uncertainties in job-shop production. In their model, this uncertain environment was handled by “maintaining the openness of the decision field”, which was also suggested in Corsten and Gössinger [21]. That is decisions were made at the latest possible instant not to restrict (i.e. close) further action alternatives and also they are made in a way that guarantees largest scope of action for forthcoming decision moments. The model took flexibility as a decision criterion and specifically maximized flexibility as the objective.

Beach et al. [22] reviewed manufacturing flexibility concept and constructed a consolidation framework to make operational flexibility decisions. In this framework, the system flexibility level was determined according to strategic decisions of a company. This level constituted an input for the operational management control by setting boundaries in terms of automation, numerical control technologies, modular design techniques, etc. After this point, a cycle was initiated in order to monitor the performance of the system and take corrective actions by determining the appropriate type (process, routing, etc.) and level of operational flexibility.

The survey underscores performance of a shop can be improved if order release concepts are integrated with dynamically controlled flexibility in a flexible shop environment. Instead of planning the job flow in detail (either online or offline), allocating the necessary flexibility to the shop and controlling job flow as necessary seems an effective and more adaptable way to cope with uncertainty in make-to-order companies. We take the approach advocated in Beach et al. [22] to be performed in a cyclical manner. The approach is aimed at controlling operational flexibility on a routine basis. Thus, the flexibility allocation is considered as a subject of production planning like other issues such as capacity allocation or labor allocation.

CHAPTER 3

FLEXIBILITY MANAGEMENT APPROACH

In this chapter, flexibility management approach of this thesis study is introduced. The chapter is organized as follows: First, the flexibility measures used in the concept are defined. After that, a general view of the approach is presented. Finally, the elements that constitute the basis in flexibility allocation are mentioned.

3.1 FLEXIBILITY MEASURES

Flexible Manufacturing Systems are known to be effective at the medium levels of production volume and part variety. They fill the gap between high volume-low variety transfer lines and low volume-high variety job shops by effective use of numerical control in NC and CNC machines [23]. Figure 3.1 illustrates the application characteristics of FMSs.

As mentioned in Chapter 2, make-to-order manufacturing strategies contain inherent uncertainty. As the FMSs are not dedicated either to only high volume or to only high part variety situations, they can be re-configured to handle different demand characteristics of the market if a make-to-order policy is applied. In the approach of this thesis, two flexibility types have been included to respond, in short term, to the demand fluctuations in the market. These are process and routing flexibilities. Process flexibility is related to the part type variety, routing flexibility is related to the production volume. [24] mentions that these flexibility types place antithetical requirements on the system as the tool magazines are limited. One has to choose between increasing the number of distinct part types that can be produced simultaneously and increasing the number of alternative machines available for producing the same part type.

In order to quantify and manage flexibility effectively, process and routing flexibilities should be measured. The measures for each of the flexibility types are defined in the following sub-sections.

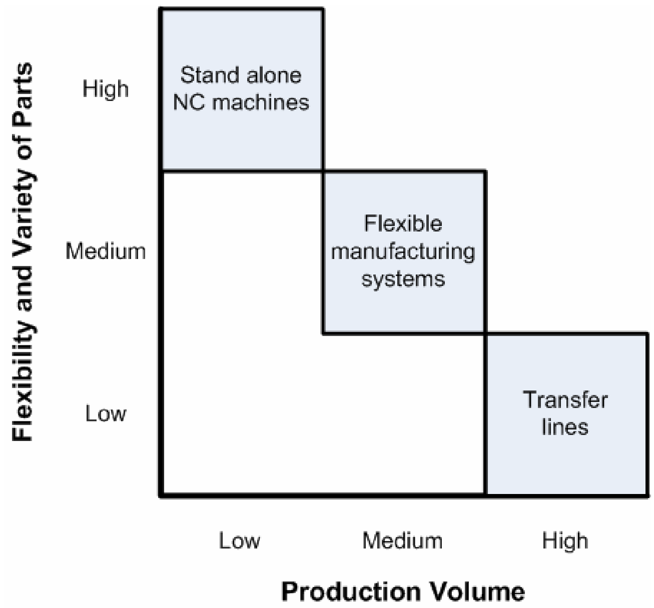


Figure 3.1 Application Characteristics of FMSs, (reproduced from [23])

3.1.1 Process Flexibility Measure

Sethi and Sethi [9] defined Process Flexibility as the set of part types that the system can produce without major setups. In Jaikumar [25] and Ettlie [26] (as cited in Sethi and Sethi [9]) firms were interviewed on their counts of part types produced to measure process flexibility. The Process Flexibility measure of this thesis is in line with the mentioned studies above. Process Flexibility of a shop is defined as,

$$PrFlex^{shop} = \frac{Part^{realizable}}{Part^{all}}$$

where $Part^{realizable}$ is the number of part types that can be realized in the shop with the current setup; $Part^{all}$ is the universe of part types that the shop can produce with the proper setup.

3.1.2 Routing Flexibility Measure

Sethi and Sethi [9] defined Routing Flexibility as the ability of a manufacturing system to produce a part by alternate routes through the system. Hence, the Routing Flexibility measure of this thesis for a part type *p* is defined as,

$$RtFlex_p = \frac{Route_p^{open}}{Route_p^{\max}}$$

where $Route_p^{open}$ is the number of available routes for part type *p* with the current setup; $Route_p^{\max}$ is the maximum number of possible routes that can be opened for part type *p*, theoretically (i.e. ignoring tool availability, machine availability, etc.).

The number of routes is calculated by multiplying the number of operation-machine assignments at each operation sequence of part type *p*:

$$\#\ of\ routes = \prod_{s \in S(p)} T_{sp}$$

where *S(p)* is the ordered set of operations sequence number for part type *p*; index *s* is the operation sequence number; T_{sp} is total number of operation-machine (Oper-M/C) assignments available for operation at sequence position *s* of part type *p*.

The maximum value T_{sp} can take is equal to the number of machines in the system. If *M* is defined as the set of all machines, and $COUNT_p$ as the total number of operations for part type *p*, then $Route_p^{\max}$ can be given as,

$$Route_p^{\max} = |M|^{COUNT_p}$$

where $|M|$ is the cardinality of set *M*.

Having defined the number of route calculations in factorial forms, $RtFlex_p$ will be called as the Factorial Form of Routing Flexibility Measure (FFRFM) throughout the following chapters of this thesis. Hence, *FFRFM* for part type *p* can be given as,

$$FFRFM_p = \frac{\prod_{s \in S(p)} T_{sp}}{|M|^{COUNT_p}}$$

The observed characteristics of this measure are,

- If a part cannot be realized in the system

  $\Rightarrow \prod_{s \in S(p)} T_{sp} = 0$

- If a part can be realized in the system with no extra routes

  $\Rightarrow \prod_{s \in S(p)} T_{sp} = 1$

- If a part can be realized in the system with extra routes

  $\Rightarrow \prod_{s \in S(p)} T_{sp} \geq 2$

- For different cases with the same number of tools available for a part, *FFRFM* gives the maximum value if the tools are distributed to operations as uniformly as possible.

The last characteristic is due to the fact that if sum of *n* positive real numbers is equal to constant *C*, then multiplications of *n* numbers is maximized if each number is equal to *C/n*. This shows that uniformity is always preferred in the routing flexibility measure proposed in this thesis.

3.2 GENERAL VIEW OF THE APPROACH

It has been mentioned in Chapter 2 that performance of a shop can be improved if order release concepts are combined with a flexible shop environment ([8], [11], [12], [19]). However, the outcomes of the literature survey also show that increasing one flexibility type only may not be appropriate for every situation of the order pool and the shop due to the inherent tradeoffs ([13], [14], [15], [16]). A consistent distribution of flexibility was suggested considering the operational and physical characteristics of the shop, and the market environment ([14], [16], [18]).

In our approach, allocation of flexibility to the shop is controlled by a flexibility management (FM) policy. This policy manages the type of operations to be assigned to the machines, and whether duplications of the same tool type will be allocated to several machines by considering the jobs in the pool and estimated job arrivals. Hence, it is proposed that by proper operation-machine assignments, process and routing flexibilities are controlled, and short-term market fluctuations are compensated. The approach is similar to the approach suggested in Beach et al. [22], where short-term fluctuations are handled with reactive use of flexibility in a cyclical manner. The area bounded by the dashed ellipse in Figure 3.2 shows the operational management control approach mentioned in Beach et al. [22]. "Operational Management Control" box reflects the controlling influence of management on the level of the flexibility requirement; “Required Operational Flexibility” box reflects the management action and control focus; “Methods for Delivering Flexibility” box represents the facilitators of the flexibility, labor, information technology, process technology, etc.; “Performance Measurement” box reflects monitoring and maintaining effective operational flexibility deployment policies of an organization.

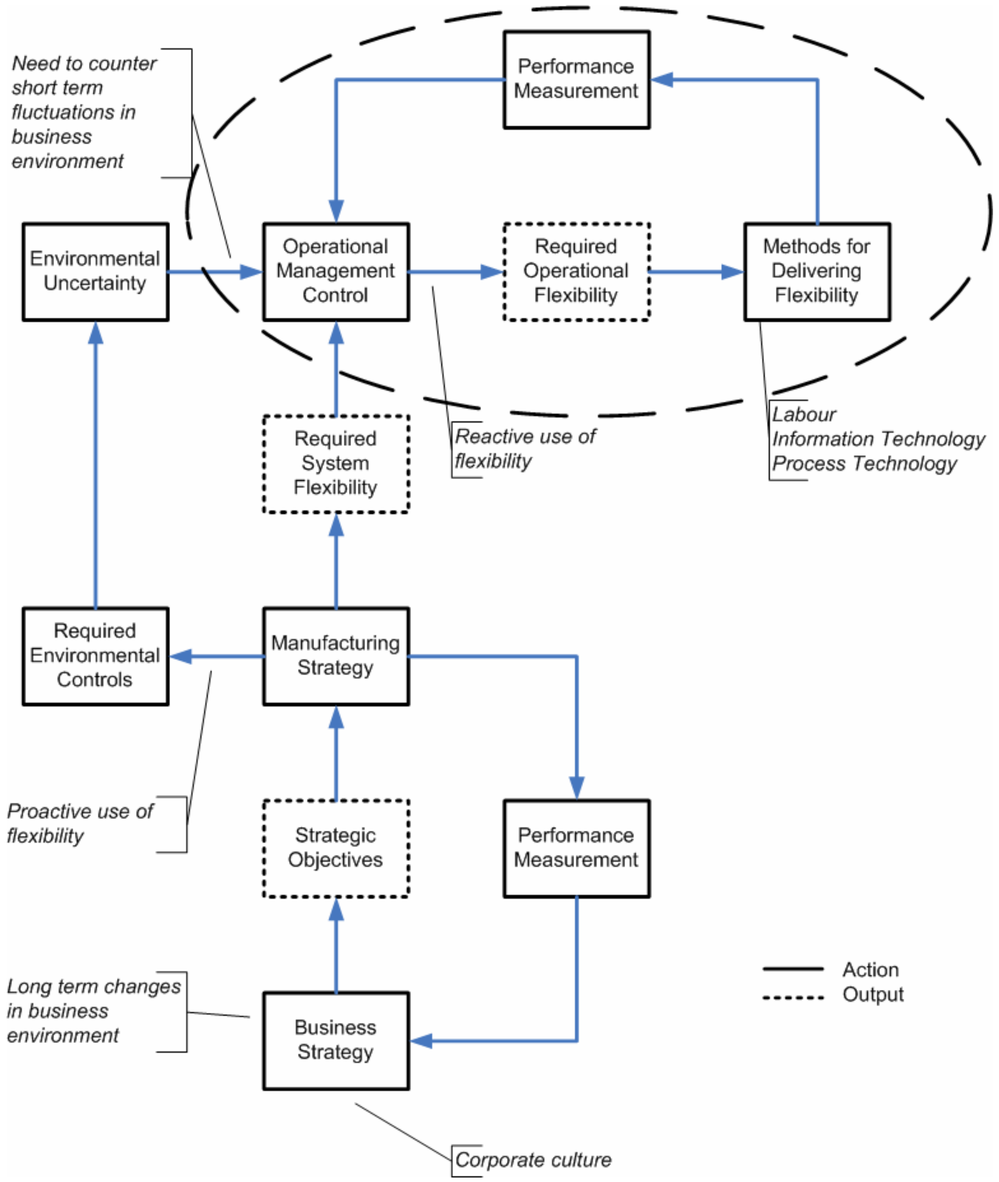


Figure 3.2 Consolidation framework, (reproduced from Beach et al. [22])

In order to find the proper flexibility management policy, two main tools are suggested to work together. The tools are a mathematical model for allocating the two types of flexibility to the shop and a simulation model to monitor the performance of the flexibility management policy.

Figure 3.3 illustrates the general view of the approach. The mathematical model-simulation model cycle is repeated until sufficient data is collected for the performance of the shop. According to the results, FM Policy is updated and the same cycle is repeated until the desired shop performance is achieved.

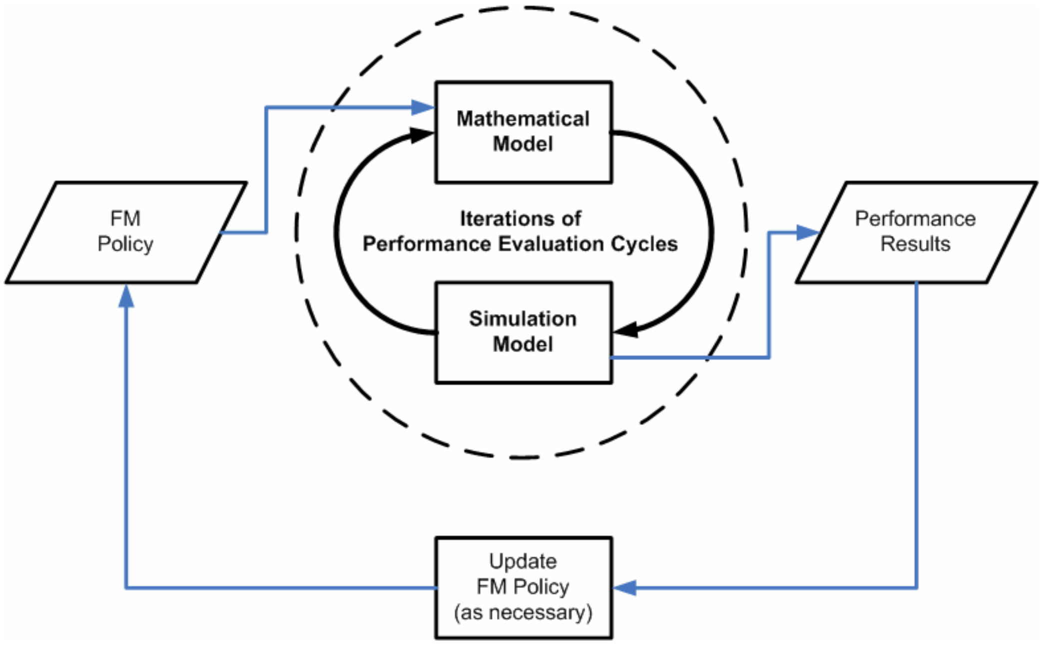


Figure 3.3 General View of the Flexibility Management Approach

There is an input/output relationship between the mathematical model and the simulation model. The mathematical model uses the data of jobs in the pool and anticipated orders, and finds a solution to operation-machine assignments (which reveals the allocation of flexibilities) by maximizing a combined function of process and routing flexibilities. Then, the operation-machine assignments are transferred to the simulation model. Although the outputs of the mathematical model are directly transferred to the simulation model, the outputs of the simulation model are pre-processed before they are used in the mathematical model. The elements that are used by the mathematical model for flexibility allocation are mentioned in the next section.

3.3 THE ELEMENTS FOR FLEXIBILITY ALLOCATION

We consider allocation of flexibility as the resultant degrees of process and routing flexibilities in response to some “flexibility demand”. Demand for flexibility is conceived as the relative need for having either one of the capabilities of,

- Processing as many parts as possible simultaneously, or
- Offering as many alternative processing paths as possible to a given set of parts.

In this study, flexibility allocation is performed at two levels:

- High level flexibility allocation is done between process and routing flexibilities. The flexibility demand is related to the variety of part types and their volumes in the system. This level is managed by the elements that are named as “Flexibility Inter-weights”. These weights determine how the mathematical model behaves while finding solutions for different mixtures of process and routing flexibilities in the system.
- Low level flexibility allocation is done among the part types. Flexibility demand of each part type is related to the due dates and volumes of the orders for the related part type. This level is managed by the elements that are named as “Part Intra-weights”. These weights prioritize the part types and determine the share of each part type that they take from the overall flexibility demand in the system.

These weights are the inputs for the mathematical model as illustrated in Figure 3.4. In the coming sub-sections, we mention how these weights are calculated.

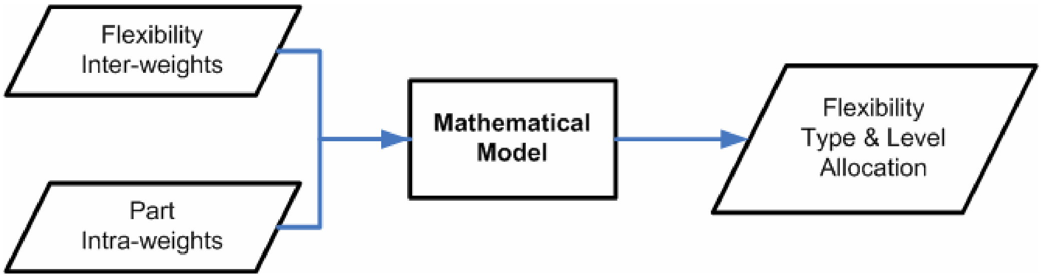


Figure 3.4 Input/Output Flow of Mathematical Model

3.3.1 Flexibility Inter-weights

As the part type variety and volume are the inputs for the calculation of inter-weights, an entropy measure similar to the one in [27] is used in order to indicate the variance in the system:

$$Entropy = -\sum_{p} (WL\%)_p * \log(WL\%)_p$$

where $(WL\%)_p$ is the ratio of workload (in work hours) of part type p to the total workload (in work hours) in the system (i.e. $WL_p / \sum_{p} WL_p$).

This entropy measure takes its maximum value when all $(WL\%)_p$ s are equal [27]. It can be shown that the maximum value for this measure is $\log P$, where P is the total number of part types.

In our approach, not only the jobs in the pool but also the expected arrivals that can happen during the next production interval are considered. The estimation is simply made using the probabilistic data of order arrivals as follows:

$$WL_p^{est} = \frac{PL}{\overline{IT}} * \overline{BS} * PA_p * UPT_p$$

where WL_p^{est} is the estimated workload for part type p (in work hours); PL is the production period length (in work hours); $\overline{IT}$ is the mean interarrival time of a generic order (in work hours); $\overline{BS}$ is the average batch size; PA_p is the probability of arrival of part type p; UPT_p is the unit processing time of part type p (in work hours).

Hence, we calculate overall workload of part type p (WL_p) as,

$$WL_p = \sum_n WL_{np}^{pool} + WL_p^{est}$$

where the index n covers the jobs in the pool for part type p; WL_{np}^{pool} is the workload of job n of part type p in the order pool (in work hours).

Next, we define the inter-weight of process flexibility (*WPROCINTER*) and inter-weight of routing flexibility (*WROUTINTER*) as follows:

$0 \leq WPROCINTER \leq 1$
$0 \leq WROUTINTER \leq 1$
$WPROCINTER + WROUTINTER = 1$

A similar approach was used in [18] where the attributes that are used to measure machine group flexibility were combined in a weighted manner and the weights summed up to 1. “Versatility” attribute was defined as the number of products that the system is able to produce; “Redundancy” attribute was defined as the number of machines that are capable of processing the same type of operations. The versatility and redundancy attributes in [18] are directly related to the process and routing flexibilities in our approach, respectively.

In order to establish the relationship between the entropy measure and inter-weights, we introduce a Flexibility Management Factor (FMF). Inter-weight of process flexibility is found by multiplying FMF and the entropy level. This concept is illustrated in Figure 3.5, where x axis represents the entropy level, y axis represents the flexibility inter-weight levels in this graph. The rays emanating from the origin represent different FMF Levels and θ is defined as the angle it makes with the abscissa. As can be seen from the graph, there is a zone where the inter-weight calculation is irresponsive to entropy level changes (i.e. inter-weight of process flexibility always equals to 1). The starting point of the irresponsive zone differs between each selected FMF Level. On the graph, the irresponsive zone starts earlier for the 3rd FMF Level than for the 2nd FMF Level. Taking this into consideration, the inter-weight of process flexibility is defined as,

$$WPROCINTER = \begin{cases} 1 & \text{if } FMF * Entropy\,Level > 1 \\ FMF * Entropy\,Level & otherwise \end{cases}$$

where FMF is equal to $\tan\theta$.

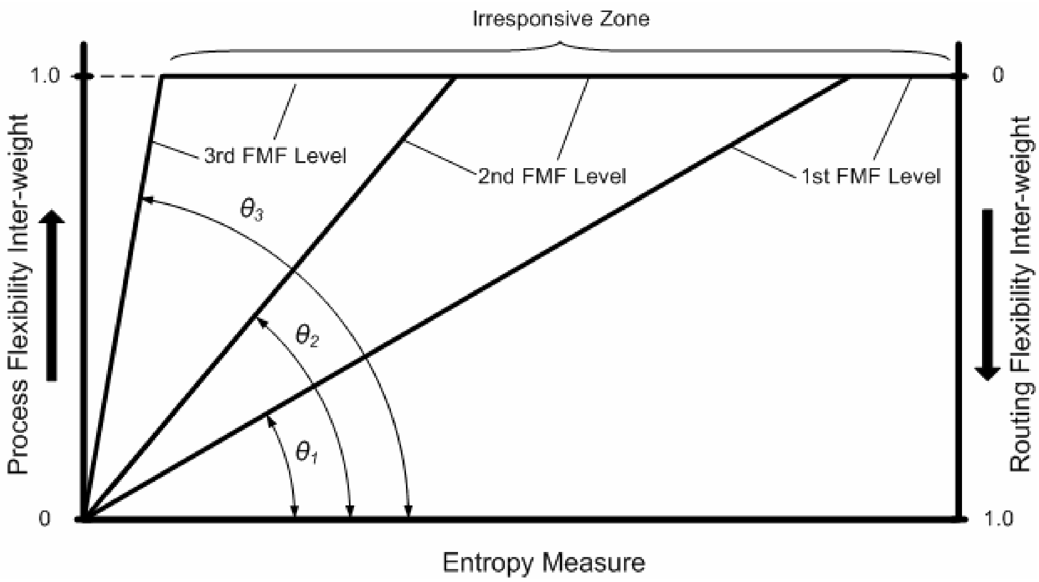


Figure 3.5 Graphical Representation of FMF Levels

It can be seen from the graph that as the entropy level increases the weight of process flexibility increases. However, the response characteristics of the lines are dependent on the FMF Level chosen. The higher the FMF Level the more sensitive the inter-weight of process flexibility becomes.

With the introduction of FMF, the FM Policy of the shop becomes dependent on only one factor. Hence, we propose that choosing the proper level of FMF for a shop will affect its performance. It is our purpose to reach one of the better performances as such, if not the best.

To illustrate the calculation of inter-weights, let us assume the shop produces 3 part types and there are 2 part types in the pool: Part type 1 and part type 2. There are 2 and 3 orders in the pool for part types 1 and 2, respectively. Let us assume we have the following parameters:

PL = 40 hours	$\overline{IT} = 4.5$ hours	$\overline{BS} = 10$ parts
$PA_1 = 0.15$	$PA_2 = 0.35$	$PA_3 = 0.50$

UPT_1 = 2 hours UPT_2 = 1.5 hours UPT_3 = 0.5 hours

WL_{11}^{pool} = 10 hours WL_{21}^{pool} = 16 hours

WL_{12}^{pool} = 15 hours WL_{22}^{pool} = 18 hours WL_{32}^{pool} = 19.5 hours

Estimated workloads are calculated as,

$WL_1^{est} = 26.66$ $WL_2^{est} = 46.67$ $WL_3^{est} = 22.22$

Hence, the overall workloads are calculated as,

$$WL_1 = 10 + 16 + 26.66 = 52.66$$

$$WL_2 = 15 + 18 + 19.5 + 46.67 = 99.17$$

$$WL_3 = 22.22$$

yielding a total load of 174.05 hours.

Converting these values to percentages,

$$(WL\%)_1 = 52.66/174.05 = 0.303$$

$$(WL\%)_2 = 0.570$$

$$(WL\%)_3 = 0.128$$

Using the workload percentage values entropy is calculated as,

$$Entropy = -0.303 * \log 0.303 - 0.570 * \log 0.570 - 0.128 * \log 0.128 = 0.411$$

If we take FMF Level 10, which corresponds to θ=10° and is rather routing sensitive, then the inter-weights are calculated as,

$WPROCINTER = 0.411 * \tan 10 = 0.073$, and thus

$WROUTINTER = 1 - 0.073 = 0.927$

Instead we can consider FMF Levels 40 and 80, and the calculations would then give, FMF Level 40: $WPROCINTER = 0.411 * \tan 40 = 0.345$

$WROUTINTER = 1 - 0.345 = 0.655$

FMF Level 80: $WPROCINTER = 1 \; (as \; 0.411 * \tan 80 = 2.33 > 1)$

$WROUTINTER = 1 - 1 = 0$

From these results, it can be seen that the FMF Level 10 shows a major need for routing flexibility while FMF Level 80 strictly demands process flexibility. Although, FMF Level 40 shows a need for both of the flexibilities, routing flexibility is demanded more than process flexibility. These cases are shown in Figure 3.6, where 0.477 (=log 3) is the maximum level of entropy for this example.

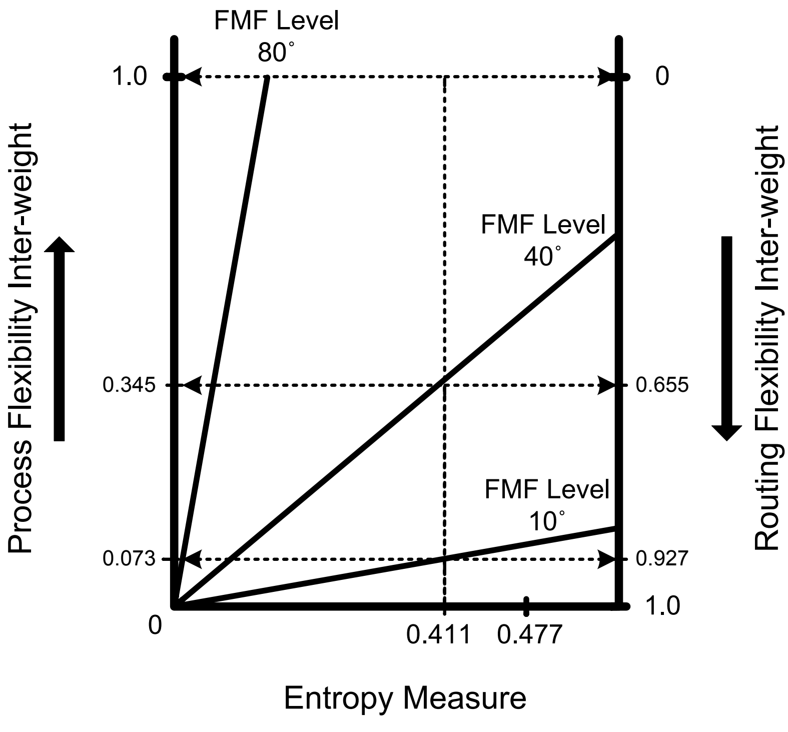


Figure 3.6 Sample Cases for Inter-weight Calculations

3.3.2 Part Intra-weights

In our approach, there are two criteria to sort the part types in terms of their contribution to flexibility allocation. These are due dates and volumes of the jobs. This is not done to use in releasing or dispatching orders. This is just to distinguish impact of orders on realized level of operational flexibility. In order to reflect the effect of both factors simultaneously, let us introduce three classes that are used to categorize urgency of the jobs:

- Overdue Orders, whose due dates are already passed.
- Urgent Orders, whose due dates are within total processing hours needed for the job from current time (*TNOW*).
- Normal Orders, whose due dates are later than total processing hours needed for the job from current time (*TNOW*).

With the classification given above, we define the Critical Ratio (*CR*) for each order in the pre-release pool as:

$$CR_{np}^{pool} = \begin{cases} \dfrac{UPT_p * BS_{np}}{DD_{np} - TNOW} & \text{if } DD_{np} \text{ - TNOW} \geq \delta \\ CR^{\max} & \text{otherwise} \end{cases}$$

where UPT_p is the unit processing time of part type p (in work hours); BS_{np} is the batch size of order n of part type p; DD_{np} is the due date of order n of part type p.

CR^{max} is a constant value assigned for overdue orders, which is greater than any value among the orders that are urgent. In order to realize this, a threshold is defined. The threshold δ ($\delta > 0$) is used in specification of CR^{max} as:

$$CR^{\max} > \frac{UPT_p * BS_{np}}{\delta} \text{ for any } n, p$$

As is the case for inter-weights, the expected arrivals that can happen during the next production interval are also taken into account. If "Total Work Content (TWK)" rule, as mentioned in [28], is used for due date assignment, then the estimated critical ratio for part type p (CR_p^{est}) is calculated as follows:

$$CR_p^{est} = \frac{UPT_p * \overline{BS}}{AT_p^{est} + F * UPT_p * \overline{BS} - TNOW}$$

where AT_p^{est} is the estimated arrival time for the next order of part type p; F is the flow allowance parameter of TWK rule ($F \geq 1$).

AT_p^{est} is estimated by the following equation,

$$AT_p^{est} = TNOW + \frac{\overline{IT}}{PA_p}$$

Substituting into the CR_p^{est} formula yields,

$$CR_p^{est} = \frac{UPT_p * \overline{BS}}{\frac{\overline{IT}}{PA_p} + F * UPT_p * \overline{BS}}$$

Since CR is calculated for each order of a part type p, we aggregate this measure by summing the CRs of all orders for every part type p:

$$CRTOT_p = \sum_n CR_{np}^{pool} + CR_p^{est}$$

There are similar applications in literature where the *CR*s are summed to find an aggregate measure, such as [29], [30], [31].

After calculating total critical ratios for each part type, intra-weights of parts are calculated by the ratio,

$$WINTRA_p = \frac{CRTOT_p}{\sum_p CRTOT_p} * 100$$

Thus, $\sum_p WINTRA_p = 100$.

To illustrate the calculation of intra-weights, let us use the same example from sub-section 3.3.1 by introducing the following additional parameters:

$TNOW = 20{:}00$ $\quad F = 6$

$DD_{11}^{pool} = 35{:}00$ $\quad DD_{21}^{pool} = 30{:}00$

$DD_{12}^{pool} = 40{:}00$ $\quad DD_{22}^{pool} = 55{:}00$ $\quad DD_{32}^{pool} = 24{:}00$

Critical ratios of the jobs in the pool are calculated as,

$CR_{11}^{pool} = 0.67$ $\quad CR_{21}^{pool} = 1.60$

$CR_{12}^{pool} = 0.75$ $\quad CR_{22}^{pool} = 0.51$ $\quad CR_{32}^{pool} = 4.88$

Estimated critical ratios are calculated as,

$CR_1^{est} = 0.13$ $\quad CR_2^{est} = 0.15$ $\quad CR_3^{est} = 0.13$

Hence, overall critical ratios are,

$CRTOT_1 = 2.4 \qquad CRTOT_2 = 6.29 \qquad CRTOT_3 = 0.13$

Finally, the part intra-weights are calculated as,

$$WINTRA_1 = \frac{2.4}{8.82} * 100 = 27.2\%$$

$$WINTRA_2 = \frac{6.29}{8.82} * 100 = 71.3\%$$

$$WINTRA_3 = \frac{0.13}{8.82} * 100 = 1.5\%$$

As can be seen from the results, the intra-weights prioritized each part type by considering the due dates and volumes of orders in the pool and estimated arrivals.

CHAPTER 4

MODELING

In this chapter, the flexibility management concept is detailed by developing the two main tools of the approach that are the mathematical model and the simulation model. Then, the interaction between the two models is established in order to complete the optimization cycle that is mentioned in Chapter 3.

4.1 MATHEMATICAL MODEL

In this section, MIP formulation for the order-driven flexibility allocation in a flexible shop is introduced. The model is developed to respond the needs of operational flexibility management by relating flexibility to the current processing and order delivery needs in the pool besides anticipated arrivals.

The flexibility allocation problem is similar in character to a set covering problem. In the classical set covering model, each constraint is forced to be included into the covered set at least once [32]. However, in the "extended" approach of this model, the process and routing flexibilities are in conflict as we have discussed in Chapter 3. Moreover, there is a competition among the parts to seize as much flexibility as possible to introduce and duplicate their required tools. As a result, it is not possible to fulfill all requirements at once. This model can be treated as a "multi-dimensional" and "extended" version of the classical set covering formulation.

This section is organized as follows: The assumptions of the model are stated in 4.1.1; the notation used in the MIP model is defined in 4.1.2; a linear approximation to non-linear *FFRFM* is proposed in 4.1.3 in order to use the routing flexibility measure in our MIP model; the objective function of the model is presented in 4.1.4; the constraints of the model are presented in 4.1.5.

4.1.1 Assumptions of the Model

Reflecting major and practical requirements is a necessity in order not to move away from the main goal of this study. For this purpose as some assumptions have been made, care has been given so that the assumptions do not mislead the results and oversimplify the handling of the problem. Without loss of generality, the following simplifying assumptions are made for the MIP model:

1) A finite set of standardized part types is produced in the shop. The set can be easily updated to include new part types or to exclude the existing ones.

2) Each operation needs one tool; so "operation" and "tool" are used interchangeably. If this assumption is found to be critical, an additional variable and a constraint set can be added to the model.

3) There are no alternative operation types. Each operation type is unique. This assumption can be relaxed by adding a constraint set.

4) Each tool needs only one slot in the tool magazine. This assumption can be easily relaxed by modifying the related constraint set.

5) Each tool type can only be duplicated if the copies are assigned to different machines. Since tool life has not been considered a part of this study, there is no reason to allocate another copy of any tool on the same

machine. Minor modifications to the model are needed in order to relax tool life assumption.

6) Tool capacities of all machines are identical. This assumption can be relaxed by giving different values to the related parameter that constrains the tool capacity for each machine.

7) No “Operation Flexibility” has been provided; that is processing needs of the parts are fixed and there are no alternate ways to produce each part. This assumption can be relaxed by introducing additional constraints.

8) Full “Machine Flexibility” has been provided; that is all machines can perform any operation on any part if the necessary tools are provided. This assumption can be easily relaxed by modifying the related constraint set.

9) Material handling systems and tool availability have not been considered as a part of this study. Tool availability assumption can be relaxed by adding a constraint set. Minor modifications to the model are needed in order to relax material handling systems assumption.

10) Setup times are ignored in the system due to the fact that the setups are done periodically all at once. Modifications are needed to include setup times as a criterion that is to be minimized and/or as a constraint set.

4.1.2 Notation

The notation used for the MIP model, definition of the sets, scope of the indices, parameters and decision variables are described in this sub-section.

4.1.2.1 Sets

The sets have been defined as follows:

O:	set of all operations
M:	set of all machines
P:	set of all part types
$S(p)$:	ordered set of operations sequence number for part type p
$O(p)$:	set of operation types associated with part type p
$O(p,s)$:	set of operation type associated with part type p, with sequence number s
$P(i)$:	set of part types which need operation type i

The scopes of indices corresponding to each set have been specified as follows:

$i = 1, 2, \ldots, |O|$;
$j = 1, 2, \ldots, |M|$;
$p = 1, 2, \ldots, |P|$;
$s = 0, 1, 2, \ldots, |S(p)|$;
where $|A|$ is the cardinality of the set A; sequence number "0" is a dummy operation sequence number introduced for circular calculation purposes.

4.1.2.2 Variables

The variables of the MIP model have been defined as follows:

X_{ij}:	Binary decision variable that assigns operations to machines. 1, if operation i is assigned to machine j; 0, otherwise

B_i: Binary variable that shows if operation i can be performed in the system.

1, if at least one tool is available in the system for operation i;

0, if no tool is available in the system for operation i

C_p: Binary variable that shows the producibility of part type p in the system.

1, if at least one complete route is open for part type p;

0, if no complete route exist for part type p

Y_{ijp}: Non-negative variable that shows if operation i of part type p that can be produced in the system is assigned to a machine.

1, if part p is producible in the system and operation i is assigned to machine j;

0, otherwise

T_{sp}: Non-negative variable that shows total number of tool-machine assignments available for operation at sequence position s of part type p.

E^{+}_{sp}: Non-negative variable that shows positive difference between number of machines available for operation at sequence position s and operation at sequence position s-1 for part type p (for $s \geq 1$).

E^{-}_{sp}: Non-negative variable that shows negative difference between number of machines available for operation at sequence position s and operation at sequence position s-1 for part type p (for $s \geq 1$).

D_p: Non-negative variable for the total number of extra tool-machine allocations available for part type p with the uniformity correction for all operations.

$CTOT$: Non-negative variable that shows the weighted total of parts that can be released to the system.

$DTOT$: Non-negative variable that shows the weighted sum of ratios of extra tool-machine assignments to the maximum extra tool-machine assignments.

WL_{ij}: Non-negative variable that shows possible share of workload of machine j for operation i.

4.1.2.3 ***Parameters***

The parameters of the MIP model have been defined as follows:

$WPROCINTER$: Inter-weight of process flexibility.
$0 \leq WPROCINTER \leq 1$

$WROUTINTER$: Inter-weight of routing flexibility.
$0 \leq WROUTINTER \leq 1$;
$WPROCINTER + WROUTINTER = 1$

$WPINTRA_p$: Intra-weight of part type p for process flexibility.
$0 < WPINTRA_p < 100$;
$\sum_{p \in P} WPINTRA_p = 100$

$WRINTRA_p$: Intra-weight of part type p for routing flexibility.

$0 < WRINTRA_p < 100$;

$$\sum_{p \in P} WRINTRA_p = 100$$

$COUNT_p$: Total number of operations for part type p.

PF_p: Penalty factor for uniformity correction of routing flexibility for part type p.

$DMAX_p$: Maximum number of alternative machines that can be allocated to operations of part type p in total.

$$DMAX_p = COUNT_p * (|M| - 1)$$

$MAXTOOL$: Tool capacity for all machines.

WL_{ip}: Total of existing pool workload and anticipated arrivals of operation i for part type p.

LF: Load factor for minimum workload on each machine.

CAP: Capacity of each machine within a production period (in work hours).

$BIGM$: A very big number.

4.1.3 Linear Approximation to FFRFM

As mentioned in Chapter 3 for the routing flexibility measure *FFRFM*, the number of alternative routes for a part is increased when tools are duplicated on several machines. However, if the number of complete routes is to be maximized, then the tools of the part have to be assigned as uniformly to machines as possible (i.e. each operation of a part shall approximately have the same number of tools allocated in the system).

The *FFRFM* was defined in Chapter 3 as,

$$FFRFM_p = \frac{\prod_{s \in S(p)} T_{sp}}{|M|^{COUNT_p}}$$

This measure has the nonlinear product term in the decision variables. This is replaced by another ratio that is linear instead in the same decision variables in order to use it in the MIP model. The approximate routing flexibility measure is defined as,

$$LA(FFRFM)_p = \frac{P(\sum_{s \in S(p)} T_{sp} - COUNT_p)}{COUNT_p * (|M| - 1)}$$

where the function *P(a)* is such that *P(a)* ≤ *a*. This is to distinguish between balanced and unbalanced operation assignments to machines (i.e. uniformity). $\sum_{s \in S(p)} T_{sp} - COUNT_p$ counts the number of "alternative" machines (or routes) for part *p*, as is the case in [15]. The denominator is the maximum number of alternative machines that can be allocated to operations of part type *p* in total. In order to capture the behavior of *FFRFM* we have defined *P(a)* as,

$$P(\sum_{s \in S(p)} T_{sp} - COUNT_p) = \sum_{s \in S(p)} T_{sp} - COUNT_p - PF_p * \sum_{s \in S(p)} E_{sp}^{-}$$

The last term is introduced to penalize the sum $\sum_{s \in S(p)} T_{sp} - COUNT_p$ by PF_p. This is a process plan (i.e. number of operations over given number of machines) dependent factor. As the total imbalance measured in $\sum_{s \in S(p)} E_{sp}^-$ increases, the effective total will be reduced by larger deductions. Hence, the more the operation assignments deviate from balance, the larger the $\sum_{s \in S(p)} E_{sp}^-$ expression and the smaller the effective sum, *P(a)*, becomes.

From now on, the numerator part of the *LA(FFRFM)* will be denoted by D_p, and the denominator part by $DMAX_p$ as defined in 4.1.2. Hence,

$$LA(FFRFM)_p = \frac{\sum_{s \in S(p)} T_{sp} - COUNT_p - PF_p * \sum_{s \in S(p)} E_{sp}^-}{COUNT_p * (|M| - 1)} = \frac{D_p}{DMAX_p}$$

An example to illustrate the *LA(FFRFM)* calculation is given below:

Consider a part with 3 operations and a shop with 3 machines. Suppose the part has the operation sequence (2-5-4). The required tools for operation type 2, type 5 and type 4 have been loaded on 3/3, 2/3, 2/3 machines, respectively. Figure 4.1 illustrates this situation:

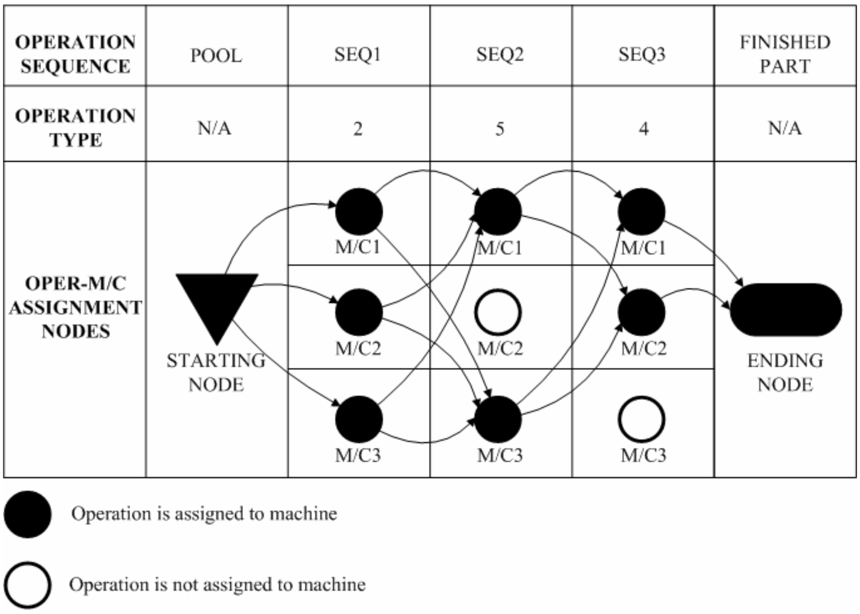


Figure 4.1 Example Case

T_{sp} values will be as follows:

$T_{1p} = 3$ $\quad$ $T_{2p} = 2$ $\quad$ $T_{3p} = T_{0p} = 2$

T_{sp} values equal to the number of filled circles for each operation in Figure 4.1. Using these values E^{+}_{sp} and E^{-}_{sp} values are calculated as follows:

$T_{0p} - T_{1p} = -1 \Rightarrow E^{+}_{1p} = 0, E^{-}_{1p} = 1$

$T_{1p} - T_{2p} = +1 \Rightarrow E^{+}_{2p} = 1, E^{-}_{2p} = 0$

$T_{2p} - T_{3p} = 0 \Rightarrow E^{+}_{3p} = 0, E^{-}_{3p} = 0$

The part has a total of 3 operations. So $COUNT_p = 3$. Say PF is taken as 0.5. Then, D_p is calculated as follows:

$D_p = (3+2+2) - (3) - (0.5*(1+0+0)) = 3.5$

$DMAX_p$ for this part is calculated as follows:

$DMAX_p = COUNT_p*(|M|-1) = 3*2 = 6$

The value is the summation over all operations of this part for maximum number of alternative machines that can be dedicated to each operation. The calculation corresponds to the area surrounded by dashed lines in Figure 4.2. This area is selected arbitrarily for illustration purposes in order to show the concept of excluding one machine for each operation (for the producibility of the part) from calculations. Only the extra machines for the operations are counted.

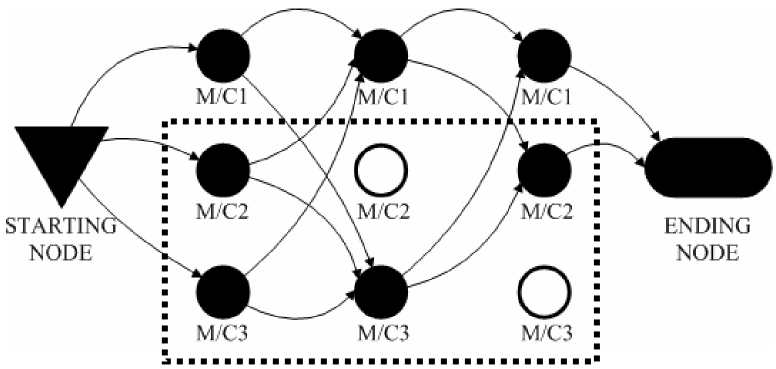


Figure 4.2 Additional Tool-Machine Assignments

Finally, $D_p/DMAX_p$ gives 3.5/6=0.583, which can be interpreted as a better situation than 3 additional machines which are associated uniformly with the three operations (i.e. one for each) where $D_p/DMAX_p$ would give 3/6=0.5, and a worse situation than 5 additional machines where $D_p/DMAX_p$ would give 4.5/6=0.75. The latter two cases are illustrated in Figure 4.3.

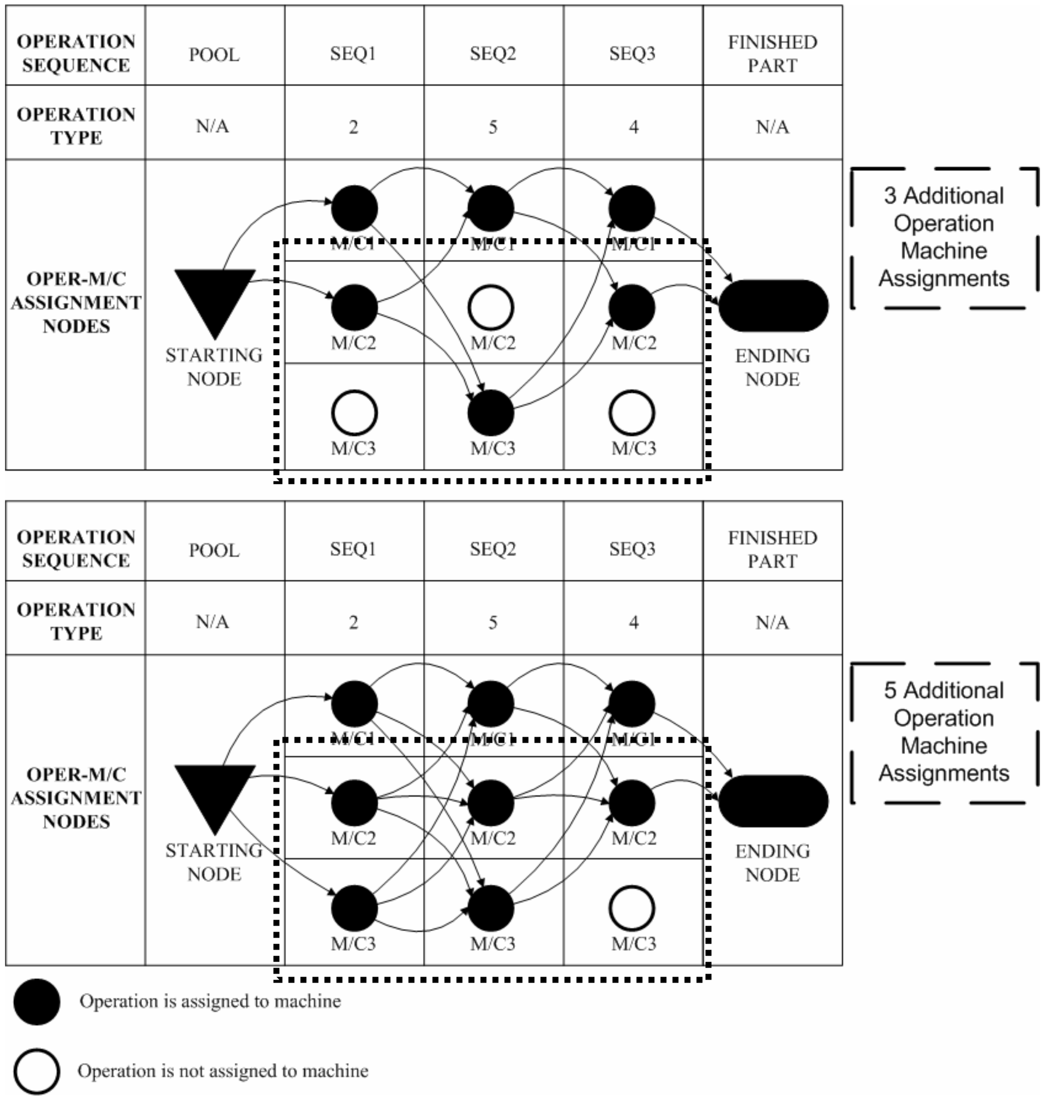


Figure 4.3 Example Case with 3 and 5 Additional Operation-Machine Assignments

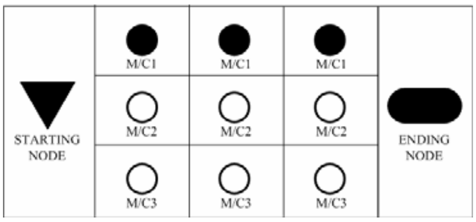
Let us take sample cases to illustrate the behavior of both *FFRFM* and *LA(FFRFM)*. The cases are shown in Figure 4.4. There are 3 machines in the shop and 3 operations for the part. The illustration concept is the same with Figure 4.1. In Case 1, there is only one complete route from start to finish. Thus, D_p equals to 0. In Case 2, there is one extra tool in the system assigned to the first operation. With the addition of an extra tool, there are 2 routes in the system. By arbitrarily selecting *PF* parameter as 0.5, D_p equals to,

$D_p = 4 – 3 – 0.5*(0 + 0 + 1) = 0.5$

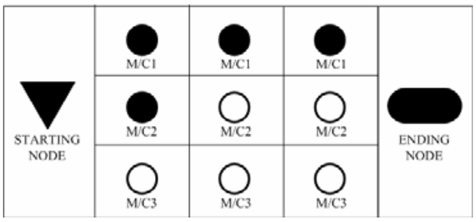
This is due to the fact that one extra tool opens an additional route but also creates a deviation from the uniformity of tool distribution among operations.

In Case 3, there are two extra tools in the system, and they are both assigned to the first operation. The number of routes in this case is 3, and D_p value is 1 (D_p=5–3–0.5*(0+0+2)). In Case 4, again there are two extra tools in the system. However, in this case, one tool is assigned to the first operation and the other is assigned to the second operation. The number of routes is 4, and D_p is 1.5 (D_p=5-3-0.5*(0+0+1)). This shows that the linear approximation captures the behavior of *FFRFM* in terms of ranking different tool-machine assignment cases.

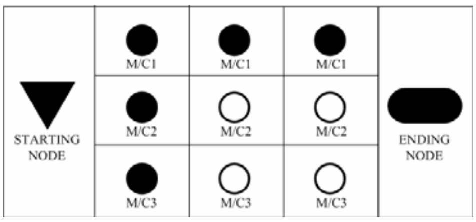
Case 5 and Case 6 show the situations in which the tools are uniformly distributed to different operations. In Case 5, the number of routes is 8 and D_p equals to 3. In Case 6, the number of routes is 27 and D_p equals to 6, where full routing flexibility is provided to the related part type. As can be seen from the D_p values, in both of the balanced allocation cases, penalty factor ceases due to the uniform distribution of tools.



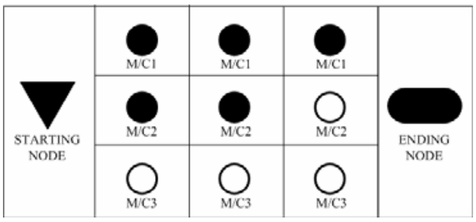
<u>Case 1</u>
Total Routes = 1 → *FFRFM* = 1/27
D_p = 0 → *LA(FFRFM)* = 0



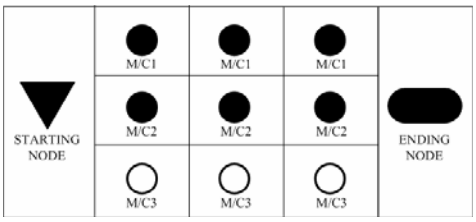
<u>Case 2</u>
Total Routes = 2 → *FFRFM* = 2/27
D_p = 0.5 → *LA(FFRFM)* = 0.5/6



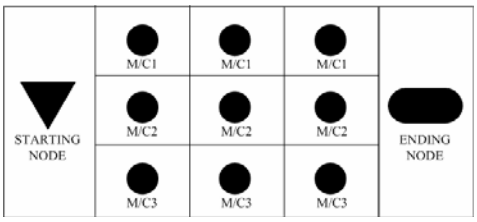
<u>Case 3</u>
Total Routes = 3 → *FFRFM* = 3/27
D_p = 1 → *LA(FFRFM)* = 1/6



<u>Case 4</u>
Total Routes = 4 → *FFRFM* = 4/27
D_p = 1.5 → *LA(FFRFM)* = 1.5/6



<u>Case 5</u>
Total Routes = 8 → *FFRFM* = 8/27
D_p = 3 → *LA(FFRFM)* = 3/6



<u>Case 6</u>
Total Routes = 27 → *FFRFM* = 1
D_p = 6 → *LA(FFRFM)* = 1

Figure 4.4 Sample Cases to Illustrate LA(FFRFM) Behavior

The non-linear form and the linearly approximated form of routing flexibility measure in these cases are given in Table 4.1:

Table 4.1 Flexibility Measures in Sample Cases

Cases	1	2	3	4	5	6
$D_p/DMAX_p$ (***Linear Apprx.***)	0	0.5/6	1/6	1.5/6	3/6	1
Routes/Max.Routes (***FFRFM***)	1/27	2/27	3/27	4/27	8/27	1

As can be seen from the example, *PF* is used to penalize deviations from uniform distribution of tools during routing flexibility allocation. *PF* can take values between 0 and 1. If it takes a value of “0”, then the model ignores uniformity. If it takes a value of “1”, then uniformity is strictly considered. However, after the addition of the first extra tool to a “pure” uniform state (as in Case 2), D_p does not change, which is an undesired situation (i.e. the case is not favorable in terms of routing flexibility, since there would be no difference between Cases 1 and 2). For the same reason, *PF* cannot take values greater than 1, since it can incorrectly indicate a decrease in routing flexibility when an extra tool is added to the system. Hence, the interval for *PF* is $0 < PF < 1$.

Instead of giving an arbitrary value to *PF*, it can be simply determined graphically by minimizing the differences in slope and ordinate measure between $D_p/DMAX_p$ and *FFRFM*. This minimization can extend over the range from no routing flexibility to full routing flexibility for the ideal cases. Ideal cases are the ones that have their D_p value the highest among some other cases that have same number of extra tools. For example, if Case 3 and Case 4 are compared, Case 4 is an ideal case for the same number (i.e. 5 tool allocations) of extra tools. Since, the number of combinations increases rapidly as the number of operations for a part and number of machines in the system increases, and minimizing the differences of ideal cases also minimizes the differences of non-ideal cases, including only ideal cases can give a fairly good approximation while simplifying the calculation process.

The reason for minimizing the slope differences is that there are fixed points on linear approximation graph that are not affected by the Penalty Factor. These are the points where the tools are uniformly distributed to operations. Hence, minimizing the ordinate measure differences only can result in *PF* values greater than or equal to 1, which is not desired. We propose finding an interval for *PF* where the total difference in slope does not change significantly and then selecting the factor that minimizes sum of differences in ordinate measure. This can be a fairly good approximation for the linear measure to *FFRFM* while preserving the general behavior of the graph.

To illustrate the concept consider a shop with 4 machines and a part with 3 operations. For every ideal case (from no routing flexibility state to full routing flexibility state), $D_p/DMAX_p$ and *FFRFM* values are calculated. Using these values, the rate of increase at every step is calculated. Then, the differences between the two measures in terms of the rate of increase are minimized by trial and error. After finding the *PF* interval that minimizes the total of slope (rate of increase between successive steps) differences, the *PF* that minimizes the sum of ordinate measure differences is selected. Figure 4.5 shows the final status for 4 machines, 3 operations. The *PF* value for this case is found to be 0.43. On the graph, the dashed circles represent the fixed points that are not affected by the value of the penalty factor.

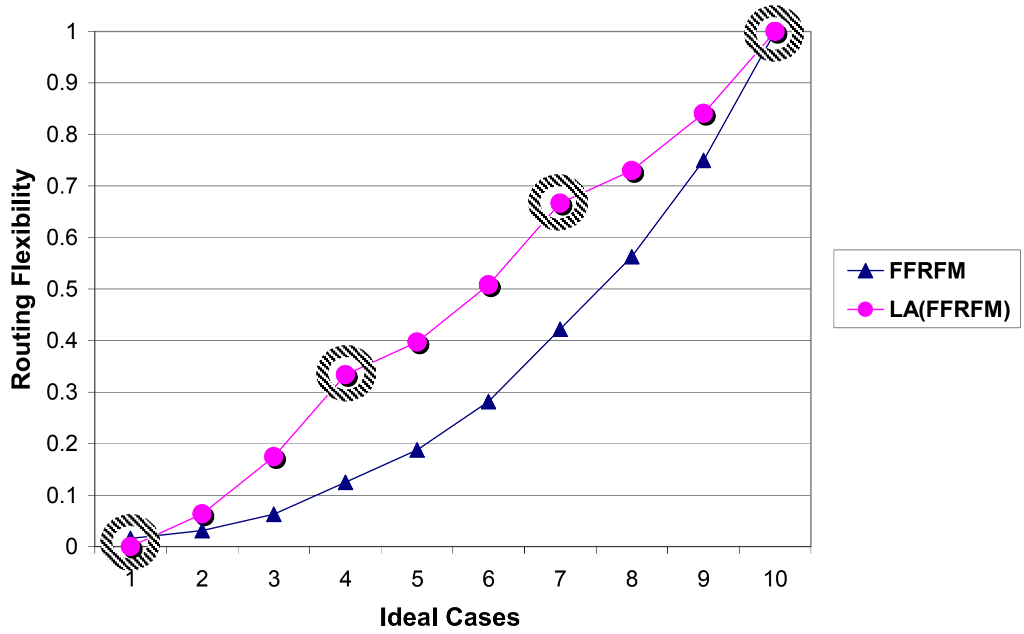


Figure 4.5 PF Corrected Graph for a Shop with 4 Machines and a Part with 3 Operations

The same calculations are carried out for two other shop and part configurations in order to examine the change in *PF*s. The second configuration consists of a shop with 4 machines and a part with 2 operations and the third configuration consists of a shop with 3 machines and a part with 3 operations. The *PF* values have been determined as 0.25 and 0.77 for the second configuration and the third configuration, respectively. This shows that *PF* is a function of two factors:

- Number of machines in the system
- Total number of operations for a part type

The resultant graphs are shown in Figure 4.6 and Figure 4.7 for the second configuration and the third configuration, respectively. The dashed circles represent the fixed points in each graph. Table 4.2 shows the sum of slope differences and sum of ordinate measure differences for the three configurations.

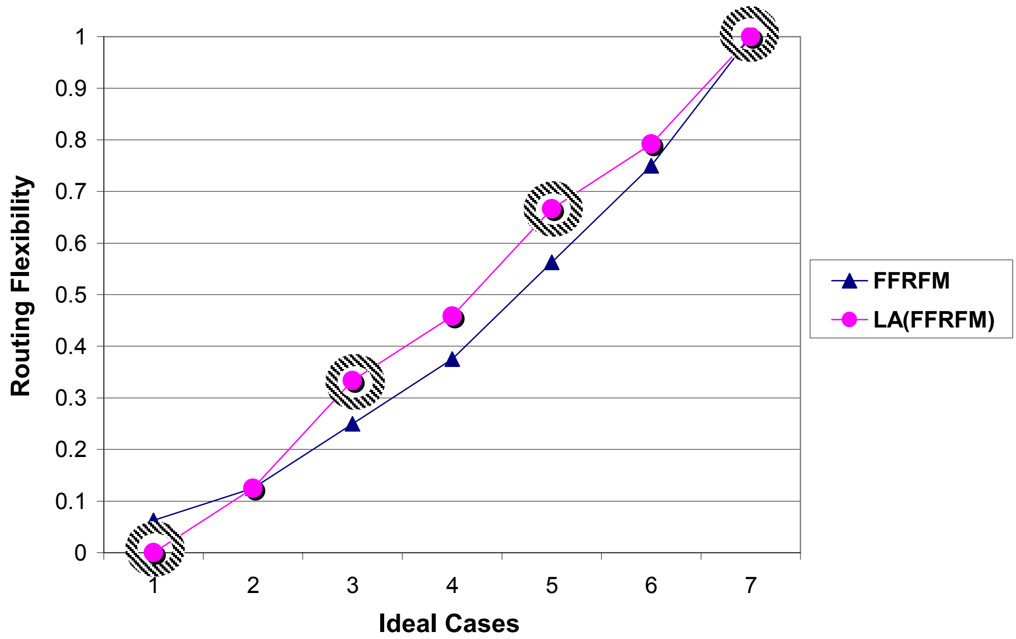


Figure 4.6 PF Corrected Graph for a Shop with 4 Machines and a Part with 2 Operations

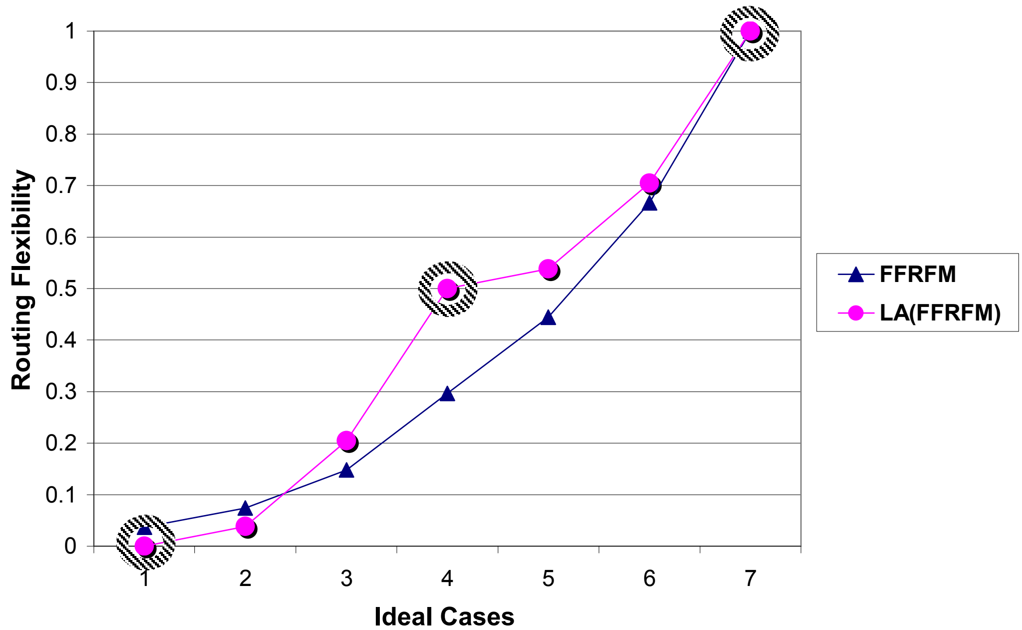


Figure 4.7 PF Corrected Graph for a Shop with 3 Machines and a Part with 3 Operations

Table 4.2 Differences in Total Slope and Total Ordinate Measure for the three configurations

Configurations	4 M/C 3 Oper.	4 M/C 2 Oper.	3 M/C 3 Oper.
Total Slope Difference	0.505	0.271	0.444
Total Ordinate Measure Difference	1.307	0.375	0.466

4.1.3.1 An Alternative Method to Linearly Approximate FFRFM

In this sub-section, another method is proposed to linearly approximate *FFRFM*, which eliminates the fixed-point occurrences on the graph; thus makes a better approximation.

In 4.1.3 we have defined D_p as,

$$D_p = \sum_{s \in S(p)} T_{sp} - COUNT_p - PF_p * \sum_{s \in S(p)} E_{sp}^{-}$$

This expression captures the behavior of *FFRFM*. However for balanced cases, that is when all assigned machines have identical number of the operations assigned, $\sum_{s \varepsilon S(p)} E_{sp}^{-}$ vanishes. This may inflate the approximate routing flexibility measure beyond what it deserves. Moreover, in defining D_p to pass through such "bumps", with a constant PF_p, the balanced cases may cause overvalued flexibility in the unbalanced allocations. In this alternative method, a fixed term strictly less than unity is appended to D_p in balanced cases (i.e. when

$\sum_{s\in S(p)} E^{-}_{sp} = 0$). However, this should exclude the 0 and maximum values of $\sum_{s\in S(p)} T_{sp} - COUNT_p$ for a part. This is due to the fact that if "0" case is affected by the fixed term, then the flexibility measure for this case will show a negative value and adjusting the "maximum" case is not necessary, even incorrect, since both the non-linear and linear measures have the same value of 1 in this case. Hence, we define,

$$D'_p = \sum_{s\in S(p)} T_{sp} - COUNT_p - PF_p * \sum_{s\in S(p)} E^{-}_{sp} - PF_p * (1-\varepsilon) * \delta_{BALANCED}$$

where

$$\delta_{BALANCED} = \begin{cases} 1 & \text{if } \sum_{s\in S(p)} E^{-}_{sp} = 0,\ \sum_{s\in S(p)} T_{sp} - COUNT_p \neq 0 \, or \, COUNT_p * (|M|-1) \\ 0 & otherwise \end{cases}$$

and $0 < \varepsilon < 1$.

This way even the least unbalanced cases (i.e. $\sum_{s\in S(p)} E^{-}_{sp} = 1$) is guaranteed to get a larger penalty in D'_p than the balanced cases. Hence, we define the new linear approximation to *FFRFM* as,

$$LA'(FFRFM) = \frac{D_p'}{DMAX_p}$$

We now explain how the PF_ps are found by a linear program to make the approximate measure *LA´(FFRFM)* fit the *FFRFM* as much as possible, subject to some regularity constraints (i.e., non-negativity, monotone increasing property). The linear program is given as follows:

Minimize $\quad DevMax$

Subject To

C1) $LA'(FFRFM)_k - FFRFM_k \geq -DevMax$ *for every k*

C2) $-LA'(FFRFM)_k + FFRFM_k \geq -DevMax$ *for every k*

C3) $LA'(FFRFM)_k > LA'(FFRFM)_l$ *for every appropriate k, l*

C4) $DevMax \geq 0$

The index k covers all the possible tool-machine assignment configurations on the interval $1 < \prod_{s \in S(p)} T_{sp} < |M|^{COUNT_p}$. Index l covers the tool-machine allocation configuration(s) which is (are) smaller than but nearest (in terms of $\prod_{s \in S(p)} T_{sp}$) to configuration k.

The objective function minimizes the maximum deviation between *FFRFM* and *LA′(FFRFM)* over all possible configurations. The interval for the configurations in this model is $1 < \prod_{s \in S(p)} T_{sp} < |M|^{COUNT_p}$. Constraints C1 and C2 covers all possible configurations for each number of additional machines. These constraints take differences of *FFRFM* and *LA′(FFRFM)* measure at every possible configuration. Constraint C3 creates a relation between *LA′(FFRFM)* values such that the rankings of the configurations in terms of *FFRFM* values preserved. For instance, if *FFRFM* values are given as 4, 6, 9 for configurations $FFRFM_1$, $FFRFM_2$, $FFRFM_3$, respectively, then the resultant value of *PF* preserves the relation $FFRFM_1 < FFRFM_2 < FFRFM_3$ by establishing the relation $LA'(FFRFM)_1 < LA'(FFRFM)_2 < LA'(FFRFM)_3$. Constraint C4 assures the parameter *DevMax* to be non-negative in order to have the absolute value of maximum deviation with constraints C1 and C2.

Let us consider the same shop environment with 4 machines and a part type with 3 operations from sub-section 4.1.3. Table 4.3 shows the related data for this condition.

Table 4.3 Configuration Data for a Shop with 4 Machines and a Part Type with 3 Operations

Extra Tool-Machine Assignments	Config. ID	Extra Machines for 1st Operation	Extra Machines for 2nd Operation	Extra Machines for 3rd Operation	Number of Combs.	FFRFM	$\sum E^{-}_{sp}$
0	*0*	0	0	0	1	1	0
1	*1*	1	0	0	3	2	1
2	*2a*	2	0	0	3	3	2
	2b	1	1	0	3	4	1
3	*3a*	3	0	0	3	4	3
	3b	2	1	0	6	6	2
	3c	1	1	1	1	8	0
4	*4a*	3	1	0	6	8	3
	4b	2	2	0	3	9	2
	4c	2	1	1	3	12	1
5	*5a*	3	2	0	6	12	3
	5b	3	1	1	3	16	2
	5c	2	2	1	3	18	1
6	*6a*	3	3	0	3	16	3
	6b	3	2	1	6	24	2
	6c	2	2	2	1	27	0
7	*7a*	3	3	1	3	32	2
	7b	3	2	2	3	36	1
8	*8*	3	3	2	3	48	1
9	*9*	3	3	3	1	64	0

On the table, 1st column shows the additional tool-machine assignments in the system. 2nd column shows the different configurations that can be applied for the same level of alternative tools. For instance, if there is a total of 7 additional tools in the system for that part, then there are 2 configurations:

- 3 extra machines for the first operation, 3 extra machines for the second operation, 1 extra machine for the third operation → {3/3/1}
- 3 extra machines for the first operation, 2 extra machines for the second operation, 2 extra machines for the third operation → {3/2/2}

The configurations are shown in the 3rd, 4th and 5th columns of the table. For both of the configurations, the tool-machine assignments can be realized in 3 different ways ({3/3/1}, {3/1/3}, {1/3/3} & {3/2/2}, {2/3/2}, {2/2/3}). The number of combinations is shown in the 6th column. On the table, the related *FFRFM* and ΣE^{-}_{sp} values are also presented.

Due to the characteristics of the LP model, the number of combinations is not relevant. For this reason, our model for this case contains 18 constraints for each configuration from *FFRFM* values 2 to 48. ε is taken as 0.1. LINDO software has been used to solve the model.

The solution gave the optimum *PF* value as 0.909, which is greater than the *PF* value found using the trial-and-error slope and ordinate measure approximation in 4.1.3. Figure 4.8 shows the graph with *FFRFM*, *LA(FFRFM)* approximation and *LA′(FFRFM)* approximation. As there are no fixed points on *LA′(FFRFM)* graph in this method, the linear measure makes a better approximation to non-linear measure while preserving the relations in the ordinate. The maximum deviation for the *LA′(FFRFM)* is found to be 0.173296 while the value for the *LA(FFRFM)* is 0.273333.

Looking at the LA’(FFRFM) one may be compelled to using a lower envelope instead of the piecewise function generated. However, this impression is misleading. LA’(FFRFM) can yield as many overestimates as underestimates for the true FFRFM. Hence whether to use a lower or and under envelope for smoothness of the approximation needs further work each time the approximation is made. In this regard, a strict choice for either direction (i.e., over- versus under- estimation) is hard to make. Extraneous rewarding of routing flexibility may block release beyond the need or cause a limited flexibility in reality; whereas, if routing flexibility is judged below what it actually is, excessive tool duplications may happen.

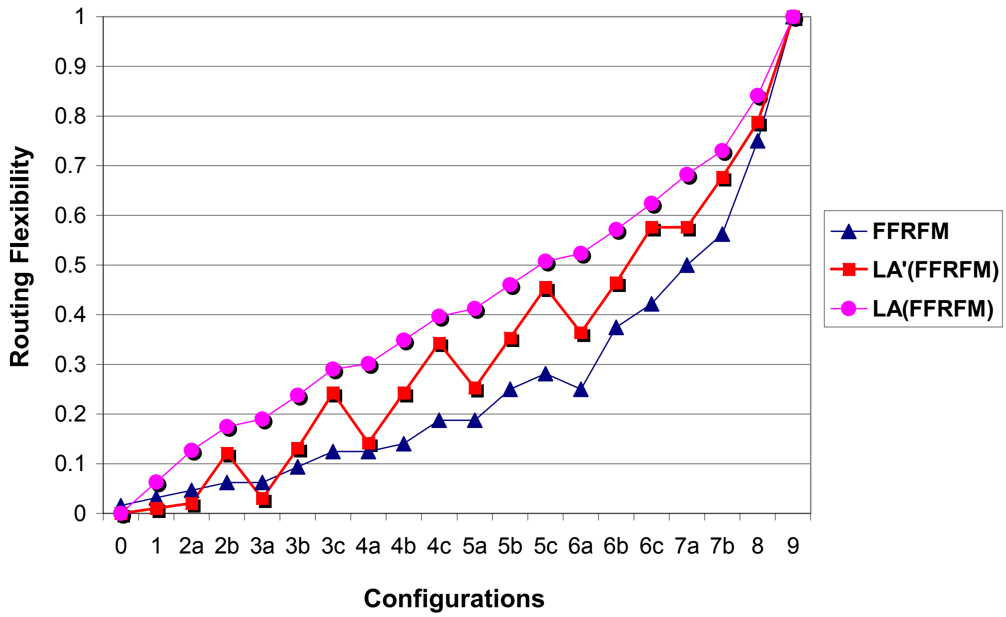


Figure 4.8 PF Corrected Graph with the Two Methods for a Shop with 4 Machines and a Part with 3 Operations

4.1.4 Objective Function of the Model

As mentioned in Chapter 3, make-to-order companies have dynamically changing orders. Control can be achieved by the choice of appropriate flexibility level dynamically in order to respond to these order fluctuations. This level needs to be affected from the current variety and density of parts in the pool and parts that are expected to arrive in the future. Moreover, urgency of orders and routing requirements of different parts are also factors to consider.

In this mathematical model, flexibility control is done at two levels. As described in Chapter 3, flexibility allocation at higher level occurs between process and routing flexibilities according to entropy measure. The lower level flexibility allocation, on the other hand, occurs between different part types. This latter allocation is performed according to the respective critical ratios. The objective function of the MIP model reflects this idea:

$$\text{Maximize} \quad WPROCINTER * CTOT + WROUTINTER * DTOT \qquad (1)$$

The objective function is based on improving a composite measure of flexibility in the shop. This is a weighted sum of process flexibility and routing flexibility, according to the parts that can be released to the shop. The *WPROCINTER* and *WROUTINTER* weights add up to 1. Maximum values of *CTOT* and *DTOT* are set equal by proper scaling and each can assume values up to 100. The first term improves process flexibility of the system by covering as many parts as possible to be produced following a setup (changeover). The second term improves routing flexibility of the system by covering as many alternative routes as possible for parts that can be released to the shop. This objective function is similar to the approach in [14] where a utility function was defined in terms of cost, quality and time and the function was maximized. A composite function is preferred rather than having one of the flexibility criteria in the objective, because the latter would need a lower bound appended to the constraints. As the model will run in a dynamic environment this would restrict model application.

The relative values of *WPROCINTER* and *WROUTINTER* weights depend on the current and anticipated future characteristics of the order pool. The more the workload of the pool is distributed among different part types with balanced shares, the higher will be the entropy value as discussed in Chapter 3. This, in turn, causes an increase in the weight of process flexibility as the machinery will be needed to permit more part types simultaneously. On the other hand, if the workload of the current pool favors a rather restricted portion of the part spectrum in the pool and this involves rather high work content (in machine-hours) relative to the workload of the future jobs, then this would result in a shift towards routing flexibility. Hence, we propose that a high entropy (i.e. variety of parts) promotes process flexibility and low entropy calls for higher routing flexibility.

4.1.5 Constraints of the Model

The constraints of the MIP model have been categorized and defined as follows:

- Process Flexibility Related Constraints
- Routing Flexibility Related Constraints
- Maximum Tool Capacity Constraints
- Workload Level Constraints
- Variable Restrictions

Process Flexibility Related Constraints:

$$\sum_{j \in M} X_{ij} \geq B_i \qquad \forall i \in O \tag{2}$$

$$\sum_{i \in O(p)} B_i \geq COUNT_p * C_p \qquad \forall p \in P \tag{3}$$

$$CTOT = \sum_{p \in P} WPINTRA_p * C_p \tag{4}$$

Routing Flexibility Related Constraints:

$$Y_{ijp} \leq X_{ij} \qquad \forall i \in O(p), j \in M, p \in P \tag{5}$$

$$Y_{ijp} \leq C_p \qquad \forall i \in O(p), j \in M, p \in P \tag{6}$$

$$T_{sp} = \sum_{i \in O(p,s)} \sum_{j \in M} Y_{ijp} \qquad \forall p \in P, s \in S(p) \qquad (7)$$

$$T_{(s-1)p} - T_{sp} = E_{sp}^{+} - E_{sp}^{-} \qquad \forall p \in P, s \in S(p), s \geq 1 \qquad (8)$$

$$\sum_{\substack{s \in S(p) \\ s \neq 0}} T_{sp} - COUNT_p * C_p - PF_p * \sum_{\substack{s \in S(p) \\ s \neq 0}} E_{sp}^{-} \geq D_p \qquad \forall p \in P \qquad (9)$$

$$DTOT = \sum_{p \in P} WRINTRA_p * D_p / DMAX_p \qquad (10)$$

Maximum Tool Capacity Constraints:

$$\sum_{i \in O} X_{ij} \leq MAXTOOL \qquad \forall j \in M \qquad (11)$$

Workload Level Constraints:

$$\sum_{i \in O} WL_{ij} \geq LF * CAP \qquad \forall j \in M \qquad (12)$$

$$WL_{ij} \leq BIGM * X_{ij} \qquad \forall i \in O, j \in M \qquad (13)$$

$$\sum_{j \in M} WL_{ij} \leq \sum_{p \in P(i)} WL_{ip} * C_p \qquad \forall i \in O \qquad (14)$$

Variable Restrictions:

$$X_{ij}, B_i, C_p = 0 \text{ or } 1 \qquad Y_{ijp}, T_{sp}, E_{sp}^{+}, E_{sp}^{-}, D_p, CTOT, DTOT, WL_{ij} \geq 0 \qquad (15)$$

Constraint (2) is generated for each operation type. B_i takes a value of "1" if the operation is assigned to at least one of the machines.

Constraint (3) is generated for each part type. C_p shows the producibility of part type p according to the LHS of the equation. C_p is a key consequent variable that is used in the objective function to improve process flexibility of the system.

Constraint (4) is the weighted summation of the C_p values yielding *CTOT* as one element of the objective function. According to the process flexibility intra-weights and producibility of the parts, *CTOT* can take values between 0 and 100, since C_p is a binary variable and process intra-weights of the parts sum up to 100 ($\sum_{p \in P} WPINTRA_p = 100$). Process flexibility intra-weights of the parts are calculated as defined in 3.3.2.

The first three constraints and the *CTOT* part of the objective function serve in favor of process flexibility. The logic tries to,

- Maximize number of producible part types in the system,
- Maximize tool type variety as much as part type variety demands,
- Minimize duplication of the same tool in the system. i.e. No alternative routes for any of the parts is promoted.

Improvement of process flexibility is realized by considering the process intra-weights of different part types and tool sharing among them. Intra-weights get higher the more urgent and larger the orders get. Moreover, the more a tool (i.e. operation) is shared by many part types the higher the intra-weights will add up to and the more that operation will be favored in process flexibility.

Constraints (5) and (6) are generated for each part type's operations and for each machine. Constraint (5) counts operation-machine assignments for a part if operation is available on any machine. Constraint (6) counts operation-machine assignments for a part if this part can be produced in the system. As a result, Y_{ijp} variable takes a value of "0" when at least one of the conditions above is not satisfied. Y_{ijp} is not a binary variable, but the objective function and the following constraints are sufficient to force it to be either 0 or 1 and no other value.

Constraint (7) is generated for each part type's operation sequences. This constraint set is used to count all available machines for an operation of a particular part type. If the part can be realized in the system, then T_{sp} gets at least equal to 1.

Constraint (8) is generated for each part type's operation sequences. This constraint calculates the differences between total number of machines available for successive operations of a part in a circular manner. This is done to measure deviations from uniformity between all successive pairings in a circular arrangement. The positive differences are held by E^{+}_{sp} variable and the negative differences are held by E^{-}_{sp} variable. Sequence number "0" is a dummy operation sequence number. It reserves a space for the last operation type of a part type for calculation purposes. For instance, if a part type has a total of 5 operations, then the operation type at sequence number "0" will be equal to the operation type at the 5th position.

Constraint (9) is generated for each part type. The first term is the summation of T_{sp} variables. This summation shows the total coverage of machine-operation assignment pairs. The second term is the multiplication of total number of operations for a part and its producibility with the given setup. This term subtracted from the ΣT_{sp} shows the number of "alternative" machines (or routes) for the part. The third term is the summation of negative differences between total number of machines available for successive operations of a part multiplied by the Penalty Factor. By subtracting this term from the other terms, "uniformity correction" is provided to approximate the factorial routing measure. As a result, D_p variable shows alternative machines available for all operations of a part with a uniformity correction. It is also possible to use summation of positive differences instead of summation of negative differences since they will be identical.

Constraint (10) is the weighted summation of the $D_p/DMAX_p$ values over all the parts yielding *DTOT*. This is another element of the objective function similar to *CTOT*. Since $D_p/DMAX_p$ can take values between 0 and 1 and routing intra-weights of the parts sum up to 100 ($\sum_{p \in P} WRINTRA_p = 100$), *DTOT* can also take values between 0 and 100. Hence, the common scaling of the relative weights for process and routing flexibilities is achieved. Routing intra-weights of the parts are calculated as defined in 3.3.2 with a slight modification. The purpose of this modification is to avoid giving an unfair advantage to the part types with lesser number of operations in the objective function. Thus, a "routing correction factor" (RCF) has been introduced and the $CRTOT_p$ values are multiplied with this factor before intra-weights are calculated. The *RCF* for part type *p* is calculated as,

$$RCF_p = \frac{COUNT_p}{\dfrac{\sum_{p \in P} COUNT_p}{|P|}}$$

The RCF_ps are calculated for each part type and multiplied with $CRTOT_p$ to find corrected total of critical ratios ($CRTOT'_p$),

$$CRTOT'_p = RCF_p * CRTOT_p$$

Then, corrected total critical ratios are used to calculate routing intra-weight of part type *p*,

$$WRINTRA_p = \frac{CRTOT'_p}{\sum_p CRTOT'_p} * 100$$

Constraints (5), (6), (7), (8), (9), (10) and the *DTOT* part of the objective function serve in favor of routing flexibility. This logic,

- Reduces the number of producible part types in the system,
- Maximizes the number of alternative routes by duplicating the same tool types for a rather small set of producible part types in the system,
- Minimizes tool type variety in favor of duplicating tools as much as possible.

Constraint (11) is generated for each machine. This constraint ensures that tool allocations do not violate magazine capacity of each machine.

Constraint (12) is generated for each machine. This constraint ensures that the operation machine assignments fill the machines with a minimum workload level for the next production period. This results in a relatively more balanced shop configuration at any flexibility level and allocation. The load factor in this constraint can be selected according to the planned utilization level of the shop.

Constraint (13) restricts WL_{ij} variable to take a value greater than zero only if the required tool i is assigned to machine j.

Constraint (14) together with constraints (12) and (13), ensures that if there are duplications of a tool in the system, then the whole workload associated with the tool will possibly be shared between all copies of the tool.

Constraint set (15) represents binary and non-negativity restrictions on the variables.

4.2 SIMULATION MODEL

In this section, a simulation model is developed to represent essential features of a typical flexible shop. As mentioned in Chapter 3, in our approach, simulation is the companion of the mathematical model due to the inherent uncertainty and dynamics of the flexible shop operation. This model constitutes the part of the study that is used to reveal the effects of different flexibility types and levels applied on a shop. The shop environment is discussed in this section. The interactions between the simulation model and the mathematical model are mentioned in section 4.3.

4.2.1 Assumptions of the Model

During the development of the simulation model, some assumptions are made without loss of generality. The aim is to reflect the details that may make major differences on the outputs, and neglect the ones that have no significant effect on the main goal of this study. The following simplifying assumptions have been made in addition to the assumptions of mathematical model mentioned in 4.1.1:

1) Machines are assumed to work without breakdowns and preventive maintenance.

2) Each operation has a fixed processing time and it depends only on the operation type. Even if two different part types share the same operation type, the unit processing times will have the same value for that particular operation type.

3) Batch splitting is allowed in the system.

4) If a part has consecutive operations that can be performed on the same machine, then the part is not allowed to leave that station before the particular operations are processed.

5) A fixed production interval with a periodic and once for all setup convention is used.

4.2.2 Model Structure

The simulation model mainly consists of 6 areas. Each area has its own logic. These areas are:

- Order Arrivals Area
- Pool Area
- Release Area
- Machines Area
- Dispose Area
- Setup Area

4.2.2.1 Order Arrivals Area

The sequence of "Order Arrivals Area" activities is as follows:

1. A generic part type arrives randomly to the system. Exponential distribution with a mean interarrival time is used for this purpose as it is done in many such modeling, such as [33].

2. Part type is specified randomly according to a discrete empirical distribution. According to this specification, number of operations,

operation types, operation sequences, and operation times of the part type are taken from the predefined tables and assigned to the order.

3. Batch size is assigned randomly according to a discrete uniform distribution as in [33].

4.2.2.2 __Pool Area__

The sequence of “Pool Area” activities is as follows:

1. Due date assignment is done using “Total Work Content (TWK)” rule as mentioned in [28]. [34] mentions that the rule is quite effective for setting due dates. The formula for setting due date for order *n* of part type *p* is as follows:

$$DD_{np} = AT_{np} + F * TPT_{np}$$

   where DD_{np} is the due date of order n of part type p; AT_{np} is the arrival time of order n of part type p; F is the flow allowance ($F \geq 1$); TPT_{np} is the total processing time of order n of part type p.

2. The parts are sent to the “Pre Shop Pool” and simply ordered according to the EDD rule. This rule is mentioned by [35] as a local selection rule for orders that are placed in the pool and also used by [36] for pool order prioritization. The orders are then kept in the pool until the next release opportunity.

Figure 4.9 illustrates the activities of Order Arrivals Area and Pool Area:

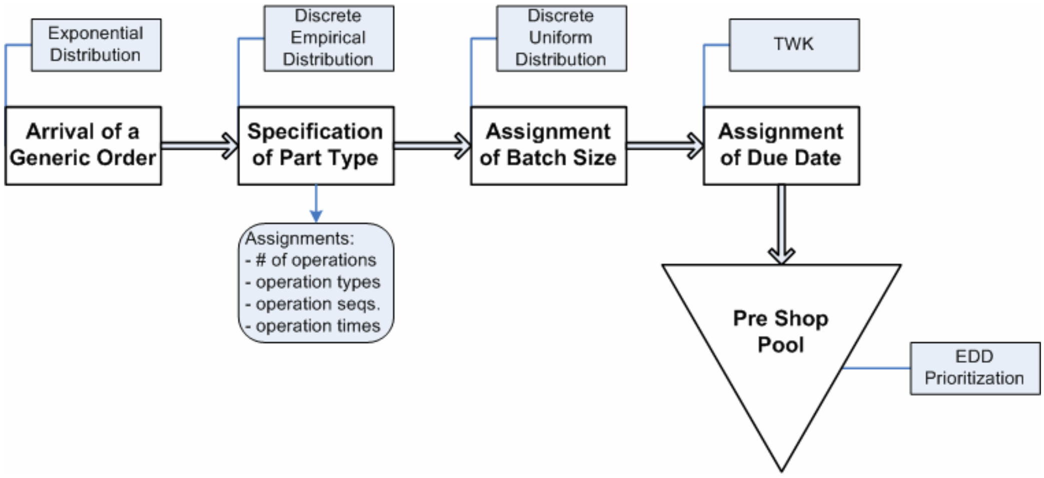


Figure 4.9 Order Arrivals Area and Pool Area Activities Sequence

4.2.2.3 <u>Release Area</u>

The sequence of "Release Area" activities is as follows:

1. Order release logic is initiated when;
   - "Aggregate Load" of any machine falls below a predefined norm or,
   - "Direct Load" of any machine becomes close to zero or,
   - A new order arrives to the pool or,
   - A setup change has just been completed.

   The aggregate load measure of a machine in this study is taken as in [37],

   $$\text{Aggregate Load} = \text{Direct Load} + \text{Upstream Load}$$

   [35] defines timing convention as the term that determines when a release can take place. There are two types of timing conventions in [35]: Continuous or Bucketed. Due to the characteristics that are defined in this step, our release rule can be considered to be the continuous timing convention.

2. The pool is searched to find an order to be released if periodic setup time is not reached. We apply order release as mentioned in [37]. An order is only selected for release if its release does not cause the workload norm of any machine to be exceeded, which is also suggested in [6].

3. If at step 2, no suitable order is found to be released, then another condition is searched: If direct load on any machine drops below a minimum level, which indicates that the specific machine is "starving", then an order with the first operation on that idling machine is released. This release method is mentioned in [38] with references to Hendry [39] and Tatsiopoulos [40]. As mentioned in [38] both authors suggest that when a machine becomes idle, an intermediate pull release can be triggered by the foreman of that particular station. However, as suggested in [41] a maximum allowance is also defined for this step: The "pull release" is permitted only, if release of the order does not cause a "second level" workload norm of any machine set higher than the first to be exceeded. As a result, idleness of the machines is avoided without losing control over the aggregate workload.

4. If an order is released to the system, then its workload is distributed to all related downstream machines for the aggregate load computations.

The routing decisions of the parts are performed dynamically with split batches. Every part selects the next machine in its operation sequence individually according to "Minimum Waiting Time in Queue" (MWTQ) dispatching rule. In this rule, processing times of parts waiting in the input buffer of each machine that can process the current operation of the part are summed. Then the machine with the lowest value is selected as suggested in [42]. As a result, it is difficult to know at release moment the exact path of a part from the start to the finish if there are tool duplications on several machines. In order to overcome this difficulty an extended aggregate calculation is needed. If there are more than one copies of a tool in the system, the aggregate loads are calculated as follows:

- At the entrance to the shop, the workload of order *n* of part type *p* on tool *i* is distributed "equally" to every machine that has tool *i* on its magazine.

$$\text{+Agg. Load to each M/C with Tool } i = \frac{(\text{Batch Size})_{np} * (\text{Unit Processing Time})_i}{(\text{\# of Machines with Tool } i)}$$

- When operation *i* on any machine is completed for each unit of part type *p* the aggregate loads are decremented equally on every machine that has tool *i*.

$$\text{-Agg. Load from each M/C with Tool } i = \frac{(\text{Unit Processing Time})_i}{(\text{\# of Machines with Tool } i)}$$

5. The batch is split at the release instant and the parts are routed to the first operation using the MWTQ rule. Individual parts may get distributed to several machines if there are tool duplications for any operation. Here instantaneous queue sizes of related machines are considered and queue sizes get incremented as every unit of a part is added to it.

Figure 4.10 illustrates the logic of this area:

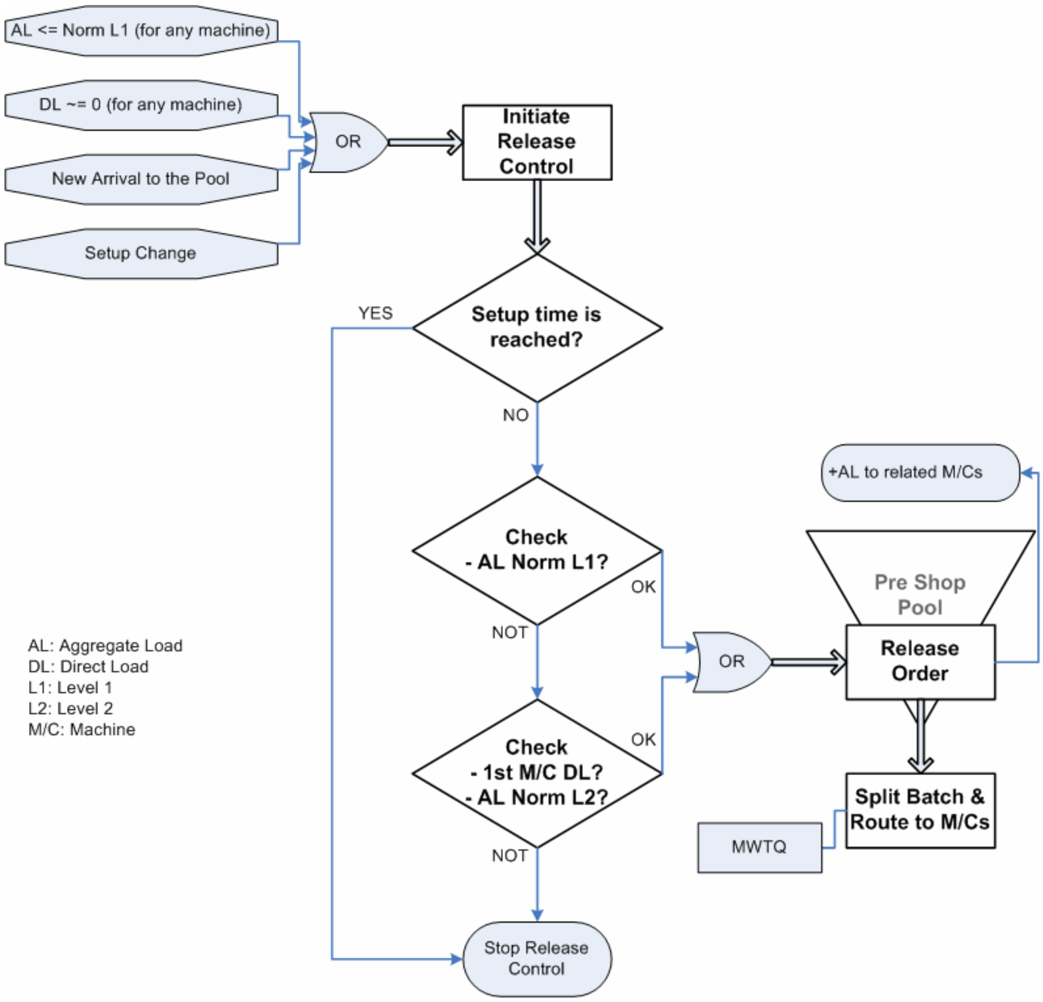


Figure 4.10 Release Area Activities Sequence

4.2.2.4 Machines Area

The sequence of activities in each "Machine Area" is as follows:

1. Parts arrive to the machine station and line up in the queue of the station according to EDD rule. This sequencing rule is selected to provide a consistency between pool prioritizing and job dispatching. Usage of a due-date oriented rule is also suggested in [7]. The direct load of the station is increased according to the consecutive operations of the part on the machine.

2. After the current operation of the part is finished, aggregate loads are decreased as defined in 4.2.2.3. Processing time for the current operation sequence of the part is subtracted from the direct load of the station.

3. A decision point is defined to check routing of the part:
   - If all operations of the part is finished, then the part is routed to "Dispose Area".
   - If the part's next operation can be realized on the same machine, then the part stays on that machine in order to have its next operation processed.
   - If it is not possible to process the next operation on the same machine, then the part is routed to another machine. The machine is selected from the set of machines that have the capability to process the next operation according to the MWTQ rule.

4. As soon as the machine completes a part, the next part with the earliest due date is pulled from the queue and its current operation is processed.

Figure 4.11 illustrates the logic of this area:

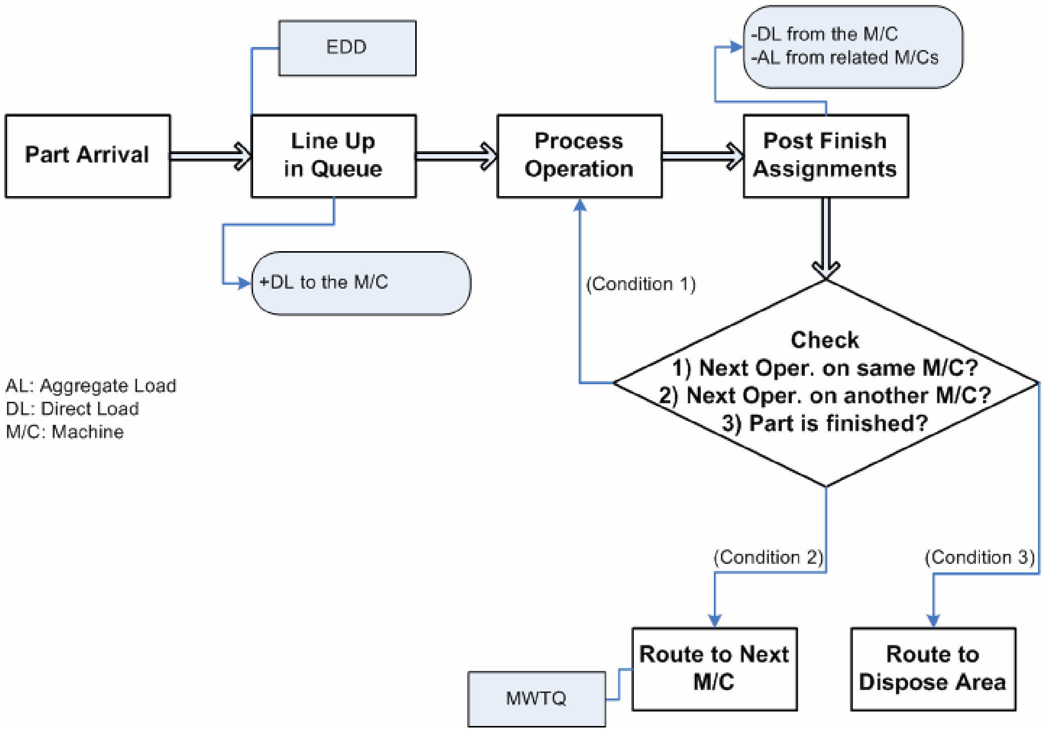


Figure 4.11 Machine Area Activities Sequence

4.2.2.5 Dispose Area

The sequence of "Dispose Area" activities is as follows:

1. Finished parts arrive to the dispose area.

2. The finished parts are delayed in the area until all the parts from the same batch are collected.

3. Statistical data are recorded and then the batch is disposed.

Figure 4.12 illustrates the logic of this area:

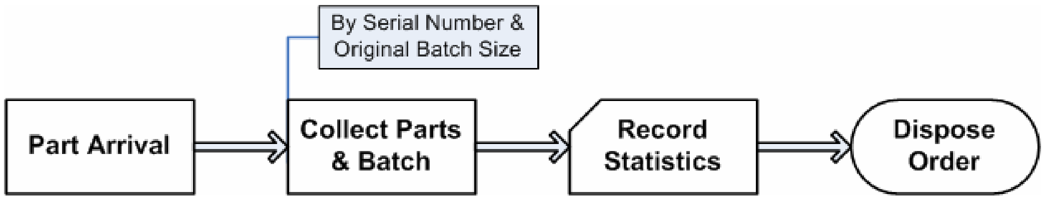


Figure 4.12 Dispose Area Activities Sequence

4.2.2.6 Setup Area

The activity sequence of the "Setup Area" is as follows:

1. Control is initiated at the beginning of every hour to check if both of the setup conditions are satisfied. These conditions are;
   - Periodic setup time is reached.
   - The shop is empty.

   A fixed production interval with a periodic and once for all setup convention is used in order to measure the effectiveness of the mathematical model.

   The logic asks to wait for the parts in the shop to be finished and disposed even if the periodic setup time is reached. This situation may be considered to correspond to making overtime whenever necessary.

   The logic has similarities with the "Extended Gating Cyclic Service Model" mentioned in [43] in the following aspects:
   - Production interval is fixed and setup changes occur on a periodic basis.
   - Existing or newly arrived orders can be sent to the shop during the entire production interval so long as they can get processed.
   - A setup can be performed (purely for the anticipated orders) at the beginning of a production interval even if there are no orders

waiting. This is subject to mathematical model solution yielding allocations of those operations.

However, there are also differences between the two approaches:

- The model in [43] has a "cyclic" service discipline. As mentioned by the author, production intervals are part type specific.
- Overtimes are not allowed in the "Extended Gating Service Model".

In order to reduce the need for overtime and cease the release of pool orders smoothly before the next setup instance, the following modifications are made to the workload control concept:

- The workload norm level 1 is decreased linearly as the setup time (i.e. end of a production interval) approaches. i.e. If t hours are left before the setup moment and t is less than the latest workload norm level 1, then the level 1 norm will be set equal to t.
- The workload norm level 2 is also decreased linearly. Since level 2 norm is higher than level 1, its decrementing starts earlier.

Figure 4.13 shows the behavior of norm levels in time:

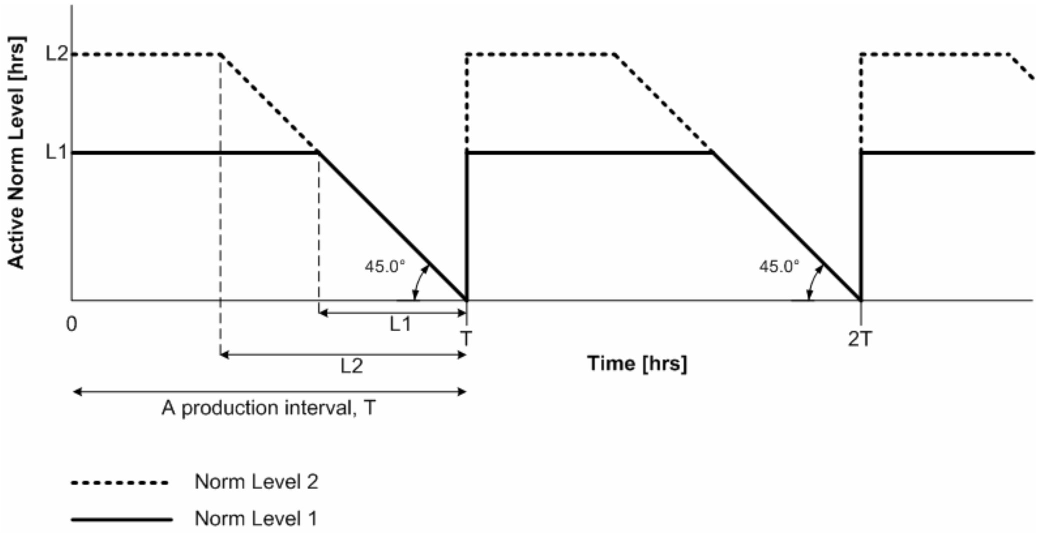


Figure 4.13 Change in Workload Norm Levels in Time

2. Setup is performed if the conditions at Step 1 are satisfied. Otherwise, production continues.

Figure 4.14 illustrates the logic of this area:

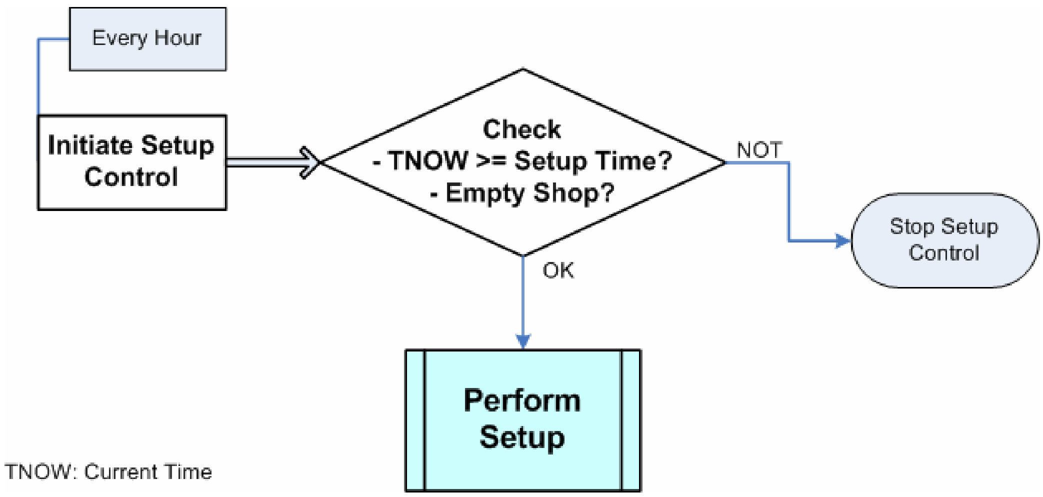


Figure 4.14 Setup Area Activities Sequence

4.3 MATHEMATICAL MODEL-SIMULATION MODEL INTERACTION

The mathematical model serves as a companion to the simulation model. According to the jobs in the pool at the setup moment and the anticipated jobs, a new shop configuration is generated by the mathematical model. Thus, there is an input/output relation between the two models. The interaction between the two models is as follows:

1. The following inputs for the mathematical model are calculated at the setup moment using the saved simulation data of part types in the pool and parts that are expected to arrive in the future:
    - Workload status of part types (WL_p)
    - Inter-weights of flexibilities (*WPROCINTER*, *WROUTINTER*)
    - Intra-weights of part types ($WPINTRA_p$, $WRINTRA_p$)

2. MIP model finds the optimal tool-machine assignments (hence, flexibility allocation).

3. The new shop configuration is transferred to the simulation model and a new production interval begins.

Figure 4.15 shows the interaction between the two models:

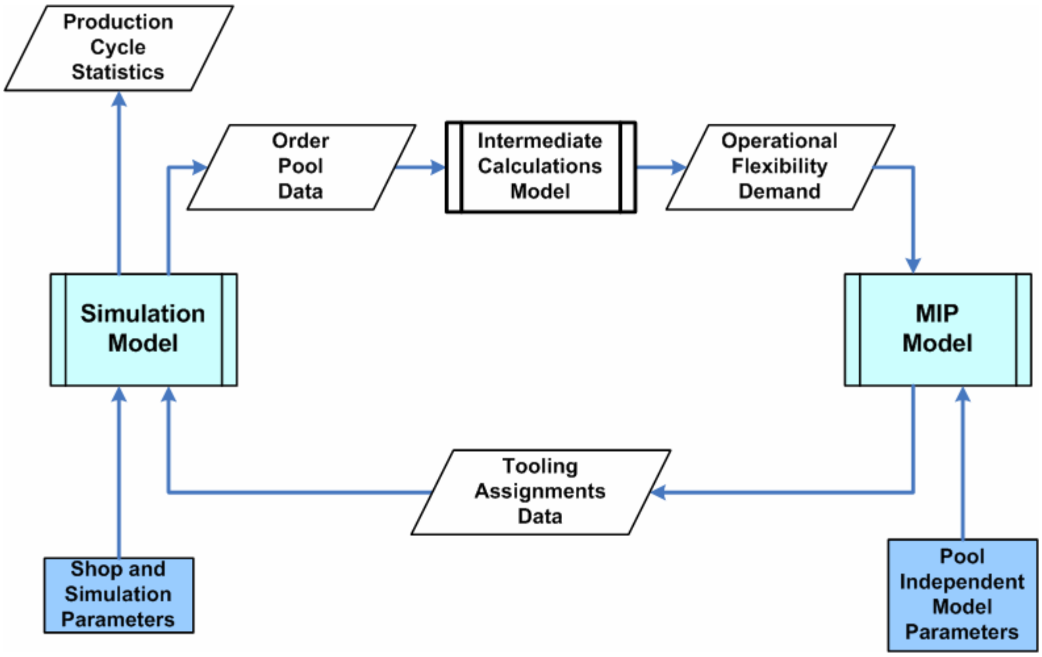


Figure 4.15 The Combined Operation and the Interaction between the MIP and the Simulation Models

CHAPTER 5

EXPERIMENTAL STUDY

In this chapter, an experimental study is set up to study the impacts of operational flexibility management approach on a typical flexible shop. Main aim of the study is to show how much and in response to what the shop performance is affected while FMF levels are changed.

For simulation modeling, ARENA software has been used; for mathematical modeling GAMS modeling language has been used. The intermediate calculations have been carried out using Microsoft Excel spreadsheet software. VBA (Visual Basic for Applications) codes have been used to automate the interactions between all the components. GAMS and VBA codes are given in the appendix.

The data that are used in the experiment have been compiled from [33], [44], [45].

5.1 SETTINGS OF THE EXPERIMENT

In this section, specifications and settings of the experiment are defined. In order to make the simulation model as realistic as possible, most of the parameters have been taken directly from related papers in literature. Others have been determined according to pilot experiments.

5.1.1 Number of Machines

In the study of [33], data from different sources were inspected in order to find widespread cases for the number of machines in FMSs, and four machines were included in the simulation model. In this study, number of machines in the system is taken as 4 in accordance with simulation study reported in [33].

5.1.2 Job Arrival

It is mentioned in [33] to use either exponential or Erlang distribution for interarrival times in FMSs. In this study, exponential distribution with a mean of 4.5 hours is used. Test runs have been made in order to determine the interarrival rate. The criterion during this was to provide the machines with a moderate level (~75%) utilization on the average with the maximum not exceeding 80%.

5.1.3 Size of the Batch

Batch size of the orders is assumed to be discrete uniformly distributed between 5 and 15 as is the case in the study done in [33].

5.1.4 Due Date

The goal of this study is not finding a proper due date assignment for the shop. Thus, a moderate value like 6 is used for TWK "flow allowance" parameter ([50]).

5.1.5 Number of Operations for a Part

In the study [44], the number of operations for a part was taken between 2 and 5 according to their Part-Operation table. In this study, an extended version of this table is used and number of operations is taken between 2 and 6 in order to increase competition among parts for the tools.

5.1.6 Number of Part Types

Number of different part types that the shop can produce is taken as 8 as is the case taken in [44] and [33].

5.1.7 Number of Operation Types

In the study of [44], 12 different tool types were used. The Part-Operation table of this study has been extended and number of operations is taken as 15. This extension is made in order to reduce operation overlaps between parts.

5.1.8 Part-Operation Table

The Part-Operation table used in the study of [44] has been extended in order to give every part distinctive characteristics and increase competition among parts for tools. It is also considered to keep approximately the same level of tool sharing. The tool sharing level of the [44] is 2.25 part/tool, while the tool sharing level of this study is 2.07 part/tool.

The Part-Operation table of this study is given in Table 5.1. Tools 13, 14 and 15 have been introduced to the original table of [44]. As a result, part types 1, 3 and 5 are given distinct properties.

Table 5.1 Part-Operation Table

Part Type	Sequence Position 1	2	3	4	5	6
1	O3	O7	O9	O10	O13	O14
2	O5	O8	O12	0	0	0
3	O3	O7	O10	O15	0	0
4	O2	O6	0	0	0	0
5	O4	O5	O8	O13	0	0
6	O3	O7	O9	O10	0	0
7	O1	O2	O6	0	0	0
8	O4	O5	O8	O10	O11	0

5.1.9 Operation Durations

In the simulation study of [33], the processing time for each operation was uniformly distributed from 6 to 30 minutes. In this study, the values are also taken uniformly generated between 6 to 30 minutes. Each operation has a fixed processing time and it depends only on the operation type (assumption #2 from sub-section 4.2.1). The following Table 5.2 gives the processing times of each operation type:

Table 5.2 Unit Processing Times of Operations [in hours]

Operation	Process Time	Operation	Process Time
O1	0.30	O9	0.46
O2	0.41	O10	0.32
O3	0.36	O11	0.33
O4	0.49	O12	0.42
O5	0.22	O13	0.14
O6	0.26	O14	0.45
O7	0.44	O15	0.25
O8	0.11		

Using the values from Table 5.2, Table 5.3 shows unit processing time of each part type. Total unit process time has an average of 1.24 hours with a standard deviation of 0.47 hours. Given that batch sizes also vary from 5 to 15 parts, every order introduces even more variation than this table reveals.

Table 5.3 Unit Processing Times of Part Types [in hours]

		Sequence Position						
		1	2	3	4	5	6	Total
Part Type	1	0.36	0.44	0.46	0.32	0.14	0.45	**2.17**
	2	0.22	0.11	0.42	0.00	0.00	0.00	**0.75**
	3	0.36	0.44	0.32	0.25	0.00	0.00	**1.37**
	4	0.41	0.26	0.00	0.00	0.00	0.00	**0.67**
	5	0.49	0.22	0.11	0.14	0.00	0.00	**0.96**
	6	0.36	0.44	0.46	0.32	0.00	0.00	**1.58**
	7	0.30	0.41	0.26	0.00	0.00	0.00	**0.97**
	8	0.49	0.22	0.11	0.32	0.33	0.00	**1.47**

5.1.10 Tool Slots

In the simulation study of [33], 5 slots were assumed in each magazine. It was also remarked that this value is taken to comply with other conditions and specifications of the study, and to demonstrate the effect of tool constraint.

In this experimental study, only 3 slots are assumed in each magazine. This change has been made in order to deny the release of all parts to the shop at any time. As a result, even if full process flexibility is provided, the MIP model is forced to have allocations for only a subset of parts that can be released to the shop (4 machines*3 slots=12 tools, out of 15 tool types) subject to the workload limit.

5.1.11 Production Interval

The production interval is taken as 40 hours, corresponding to a full week. At the beginning of every interval, a setup is made to change the tooling configuration of the shop at negligible time (or outside the 40 hrs.). As mentioned in sub-section 4.2.2.6, overtimes are allowed to finish the jobs that remain in the shop when the setup time is reached. Minimization of overtimes is directly related to the effectiveness of the "decreasing workload control norms" mentioned in 4.2.2.6. This condition is checked by examining "average overtimes" at the end of the pilot experiments and it has been observed that overtime lengths are around ~3-5% of normal working hours.

5.1.12 Product Mixture

Instead of having a uniform product mixture, real data from [45] have been used. There are 11 part types that can be produced by the FMS in that work. Since there are 8 part types used in this thesis, weekly production data of [45] have been randomly selected and distributed to these parts.

Table 5.4 shows weekly production data from [45], and probability of arrival for each part generated from real data of [45]:

Table 5.4 Product Mixture

Part Type	Weekly Production [# of parts]	Probability of Arrival
1	19.00	0.142
2	20.00	0.149
3	14.00	0.104
4	12.00	0.090
5	10.00	0.075
6	12.00	0.090
7	7.00	0.052
8	40.00	0.299

5.1.13 Workload Norms

Workload norms are determined in pilot runs. Moderate values are selected to fill the shop with sufficient jobs while holding congestion in the shop at a minimum level. Workload norm level-1 is set to 15 hours, workload norm level-2 is set to 25 hours and direct load norm is set to 3 hours. Considering 0.1-0.5 hours unit operation times, these correspond to 5, 8 and 1 orders of average batch size (10).

5.1.14 Load Factor

The load factor, which creates a balanced shop environment in terms of workload, has been selected to be 0.65. This has been found as the maximum level that does not cause infeasibility in the MIP model when the pool has no orders, since a setup purely for the anticipated future orders would suffice for that minimum load.

5.1.15 Penalty Factors

The penalty factors have been calculated by the trial-error approach as described in sub-section 4.1.3. The values are given in Table 5.5 for every part. The penalty factors are directly related to the number of operations of each part. This is an expected result, since more operations mean more non-linearity on the graph for routing flexibility. In order to approximate this non-linear measure, penalty factors increase as the number of points on the graph increases.

Table 5.5 Penalty Factors

Part Type	Penalty Factors
1	0.85
2	0.43
3	0.62
4	0.25
5	0.62
6	0.62
7	0.43
8	0.76

5.2 PERFORMANCE MEASURES

As mentioned in [46], Morton and Smunt [47] had indicated that minimal work in process and minimal tardiness are the indicators that show the effectiveness of an FMS. For this reason, the main performance measure of the simulation study is selected as the "Weighted Flow Time" (WFT). This is a measure of throughput and also an indication of the level of work-in-process inventory [46]. The measure is calculated as follows:

$$WFT = \frac{\sum_{n=1}^{N} (CT_n - AT_n) * BS_n}{\sum_{n=1}^{N} BS_n}$$

where CT_n is the completion time of the last part of order n; AT_n is the arrival time of order n; BS_n is the batch size of order n; N is the number of total output in terms of number of orders.

Other performance measures have also been examined to make more powerful comments on the results. These measures are “Weighted Tardiness” (WT), “Percentage of Tardy Orders” (PT), Utilization Levels, and Average Production Cycle Length.

Tardiness is the average lateness of all orders completed after their respective due dates. “Percentage of Tardy Orders” indicates the percentage of total orders completed after their due dates. According to [46], both of these measures reflect how well the shop meets customer orders. The “Weighted Tardiness” is calculated as follows:

$$WT = \frac{\sum_{n=1}^{N} \max\{CT_n - DD_n, 0\} * BS_n}{\sum_{n=1}^{N} BS_n}$$

where DD_n is the due date of order n.

The “Percentage of Tardy Orders” is calculated as follows:

$$PT = \frac{NT}{N}$$

where NT is the number of orders completed after their respective due dates.

As mentioned in [46], machine utilization is a common measure of interest in FMS studies. In this study, utilization of all machines are recorded, and also averaged to find the overall system utilization.

Average length of production cycles is used to examine the effectiveness of “decreasing workload control norms”. It is simply calculated by dividing the total run length in one replication to the number of setups performed in that length.

5.3 EXPERIMENTAL SETTING

In this section, experimental factors and levels are mentioned, and the replication parameters for the experiments are determined.

5.3.1 Experimental Factors and Levels

The subject of this experiment is to determine an optimum level for the FMF. FMF is the slope of “flexibility-entropy” line expressed in angular degrees. The appropriate level of this parameter reflects the response characteristics of the shop to the existing demand spread.

The determination for FMF is tried at 10 levels. The experiments are run ranging from 0˚ flexibility-entropy line, which reflects full routing flexibility characteristics, to the 90˚ line, which reflects full process flexibility characteristics. At these extreme levels, the resultant flexibility level is not sensitive to the entropy of the pool orders. At the intermediate levels (10˚-80˚), flexibility control is performed according to the entropy level in the pool. Figure 5.1 shows these levels on a graph. For a shop with 8 parts, the maximum entropy level is 0.903 (=log8). This means any FMF level larger than 48˚ will reflect the largest entropy achievable with full process flexibility. If the best chosen FMF happens to be less than 48˚, this may be taken as the best flexibility behavior prescribed is rather conservative as it will never allow for full process flexibility and there will always be some room (although little) for carrying alternative paths.

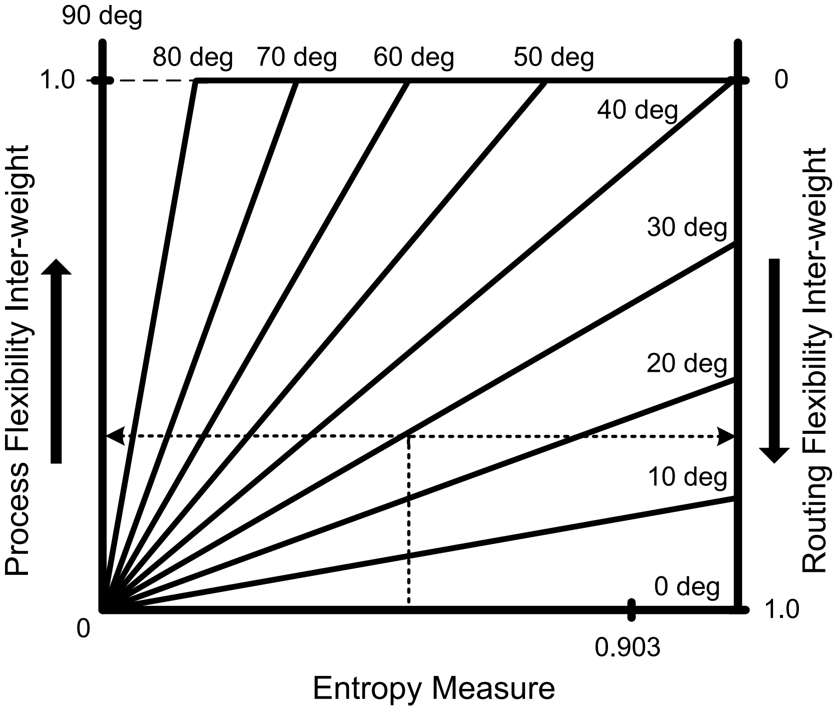


Figure 5.1 Levels for Flexibility-Entropy Line

5.3.2 Replication Parameters

The experimental study involves a non-terminating simulation. To explore its steady state behavior, a warm-up period has been determined in order to remove the transient state from the statistical data. During the pilot runs, the 90 degree flexibility-entropy line (full process flexibility) has been selected as the base case, which displayed the worst performance in terms of WFT. 10 replications were run to determine the warm-up period watching the WFT performance. Each replication had a run length of 10000 hours. The output from each replication was evaluated using a moving average of 30 observations. It was found that the first 2000 hours can be treated as warm-up period. The run length is taken to be 6000 hours, which is 3 times the length of warm-up period. Figure 5.2 shows settling of WFT for the first three replications. The replications have been performed using a computer which has Intel Core 2 Duo T8300 2.4 GHz processor and 3 GB RAM. For the base case, one replication lasts approximately

7 minutes and one cycle between GAMS and ARENA (which corresponds to one production interval of approximately 40 simulated hours) lasts approximately 2.803 seconds, which is made up of 0.132 seconds of MIP solution run, 0.185 seconds of simulation run, and 2.486 seconds of interaction between the two models.

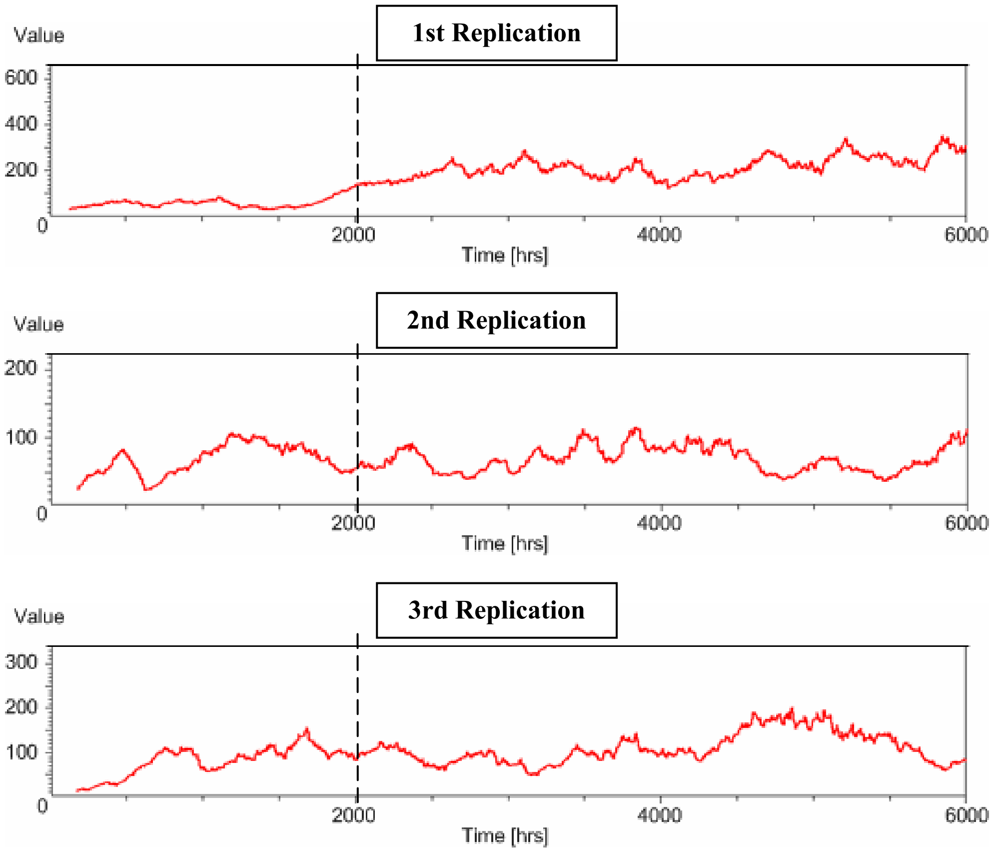


Figure 5.2 WFT Settling in the 3 Pilot Replications

The number of replications was determined in pilot runs by forming an identity in ARENA which terminates the run when the predefined precision criterion is satisfied as,

→ Half-width to mean ratio of WFT is less than 0.10 for 95% confidence interval. The formula used for confidence interval calculation ([48]) is:

$$\overline{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$$

where n is the number of replications and $t_{n-1,1-\alpha/2}$ is the upper 1–α/2 critical point for the t distribution with n-1 degrees of freedom. Using this identity the number of replications was found to be 70 with relative error safely less than 10% with more than 95% confidence for the base case. This corresponds to approximately 8 hours of run for the experiment with one factor combination setting. About 7 hours of these 8 hours would be spent in data in/out between simulation and mathematical programming.

5.4 RESULTS

In this section, the results of the experimental study are presented and discussed. The section is organized such that; firstly MIP model is run for two different job mixtures in the pool at different FMF levels. This is done in order to demonstrate the effects of chosen FMF level on the shop configuration and the behavior of MIP model during the simulation runs in detail. After that, outcomes from the main performance measure (Weighted Flow Time) are presented. Finally, other performance measures (Weighted Tardiness, Percentage of Tardy Orders, Utilization Levels, Average length of production cycle) are presented.

5.4.1 Demonstration of Flexibility Allocations in Different Pool Situations by the MIP Model

In this sub-section, two extreme pool situations in terms of entropy have been generated and used to demonstrate in detail how the MIP model suggests the shop configurations. In each pool situation, with the contents of order pool fixed, the FMF Levels have been changed. For each FMF Level, number of producible parts in the system and number of open routes for the related parts have been determined.

The first pool situation consists of 3 part types. Total critical ratios and workloads (in hours) for anticipated orders and the orders in the pool are given in Table 5.6. Intra-weights are calculated using these data.

Table 5.6 Data for 1st Pool Situation

	Anticipated Orders Data		Pool Data		Intra-weights [%]	
Part Type	**∑CRp**	**Workload [hrs]**	**∑CRp**	**Workload [hrs]**	**Process**	**Routing**
1	0.13	27.35	3	120	24.64	33.540
2	0.10	9.95	0.9	20	7.86	5.350
3	0.11	12.72	8	250	63.77	57.858
4	0.07	5.33	0	0	0.58	0.264
5	0.08	6.37	0	0	0.64	0.581
6	0.11	12.58	0	0	0.86	0.777
7	0.07	4.50	0	0	0.53	0.360
8	0.14	39.00	0	0	1.12	1.269

The entropy level for this pool situation is found to be 0.565. Table 5.7 displays the MIP solutions from FMF Levels 0 to 90. From the results, it can be seen as the FMF Levels are increased, how the inter-weight of process flexibility and inter-weight of routing flexibility changes. After Level 70, due to the entropy level in the pool, process flexibility is always preferred in full by the model. However, there are three different solutions for shop configuration. The common solution from Level 0 to Level 30 consists of 1 out of 8 part types, with 81 routes allocated for it. If FMF is increased to Level 40, 2 additional parts are included to the solution. However, the number of open routes becomes 16 for each of the producible parts. This solution extends from Level 40 to Level 50. The common solution from Level 60 to Level 90 consists of 6 out of 8 part types with only 1 route allocated to each of them.

Table 5.7 MIP Solutions at different FMF Levels for the 1st Pool Situation

FMF Level	Inter-weights		Tools in the Shop	Producible Parts	# of Open Routes
	Pr.	Rt.			
0	0	1	3x(T3)+3x(T7)+3x(T10)+3x(T15)	P3	81
10	0.099	0.901	3x(T3)+3x(T7)+3x(T10)+3x(T15)	P3	81
20	0.206	0.794	3x(T3)+3x(T7)+3x(T10)+3x(T15)	P3	81
30	0.326	0.674	3x(T3)+3x(T7)+3x(T10)+3x(T15)	P3	81
40	0.474	0.526	2x(T3)+2x(T7)+2x(T9)+2x(T10) +1x(T13)+1x(T14)+2x(T15)	P1 / P3 / P6	16 / 16 / 16
50	0.673	0.327	2x(T3)+2x(T7)+2x(T9)+2x(T10) +1x(T13)+1x(T14)+2x(T15)	P1 / P3 / P6	16 / 16 / 16
60	0.978	0.022	1x(T3)+1x(T4)+1x(T5)+1x(T7) +1x(T8)+1x(T9)+1x(T10)+1x(T11) +1x(T12)+1x(T13)+1x(T14)+1x(T15)	P1 / P2 / P3 / P5 / P6 / P8	1 / 1 / 1 / 1 / 1 / 1
70	1	0	1x(T3)+1x(T4)+1x(T5)+1x(T7) +1x(T8)+1x(T9)+1x(T10)+1x(T11) +1x(T12)+1x(T13)+1x(T14)+1x(T15)	P1 / P2 / P3 / P5 / P6 / P8	1 / 1 / 1 / 1 / 1 / 1
80	1	0	1x(T3)+1x(T4)+1x(T5)+1x(T7) +1x(T8)+1x(T9)+1x(T10)+1x(T11) +1x(T12)+1x(T13)+1x(T14)+1x(T15)	P1 / P2 / P3 / P5 / P6 / P8	1 / 1 / 1 / 1 / 1 / 1
90	1	0	1x(T3)+1x(T4)+1x(T5)+1x(T7) +1x(T8)+1x(T9)+1x(T10)+1x(T11) +1x(T12)+1x(T13)+1x(T14)+1x(T15)	P1 / P2 / P3 / P5 / P6 / P8	1 / 1 / 1 / 1 / 1 / 1

Second pool situation has been generated in order to compare the results with the first situation. In this case, the pool consists of 8 out of 8 part types. The related data for the 2nd pool situation are given in Table 5.8.

Table 5.8 Data for 2nd Pool Situation

	Anticipated Orders Data		Pool Data		Intra-weights [%]	
Part Type	∑CRp	Workload [hrs]	∑CRp	Workload [hrs]	Process	Routing
1	0.13	27.35	0.60	30	5.89	9.240
2	0.10	9.95	1.50	50	12.83	10.069
3	0.11	12.72	2	25	16.92	17.702
4	0.07	5.33	0.20	15	2.20	1.150
5	0.08	6.37	0.75	25	6.67	6.977
6	0.11	12.58	2.20	60	18.52	19.377
7	0.07	4.50	3	44	24.60	19.305
8	0.14	39.00	1.40	35	12.37	16.179

The entropy level for the second case is 0.873. Table 5.9 shows the MIP solutions. After FMF Level 50, due to the entropy level in the pool, process flexibility is always preferred in full by the model. The results show that there are five different solutions for shop configuration. The first solution identical at Levels 0 and 10 has 2 out of 8 part types loaded. The part types are P3 and P6. 27 and 54 routes are allocated to P3 and P6, respectively. The second solution at Level 20 has 4 out of 8 parts. The part types are P3, P4, P6, P7 and the number of allocated routes for them are 8, 1, 16, 1, respectively. Part type 2 is included to the solution at Level 30 leaving only one extra route for each of the part types 4 and 7. Level 40 has an alternative solution to Level 30. The only difference is the extra tool allocated to operation 2 instead of operation 6. The final solution extends from Level 50 to Level 90. It consists of 5 out of 8 parts without any additional routes allocated for them.

By comparing Table 5.7 and Table 5.9, it can be concluded that the response of MIP model depends on the pool situation. The number of solutions, the threshold FMF levels where the transition from one solution to another occurs and the resultant flexibility levels are all related to the job mix in the pool. The MIP model also takes into account the anticipated orders data. However, this data is fixed for every pool situation as we apply a fixed production cycle with stationary orders assumption.

Table 5.9 MIP Solutions at different FMF Levels for the 2nd Pool Situation

FMF Level	Inter-weights		Tools in the Shop	Producible Parts	# of Open Routes
	Pr.	Rt.			
0	0	1	3x(T3)+3x(T7)+2x(T9)+3x(T10) +1x(T15)	P3 / P6	27 / 54
10	0.154	0.846	3x(T3)+3x(T7)+2x(T9)+3x(T10) +1x(T15)	P3 / P6	27 / 54
20	0.318	0.682	1x(T1)+1x(T2)+2x(T3)+1x(T6) +2x(T7)+2x(T9)+2x(T10)+1x(T15)	P3 / P4 / P6 / P7	8 / 1 / 16 / 1
30	0.504	0.496	1x(T1)+1x(T2)+1x(T3)+1x(T5) +2x(T6)+1x(T7)+1x(T8)+1x(T9) +1x(T10)+1x(T12)+1x(T15)	P2 / P3 / P4 / P6 / P7	1 / 1 / 2 / 1 / 2
40	0.732	0.268	1x(T1)+2x(T2)+1x(T3)+1x(T5) +1x(T6)+1x(T7)+1x(T8)+1x(T9) +1x(T10)+1x(T12)+1x(T15)	P2 / P3 / P4 / P6 / P7	1 / 1 / 2 / 1 / 2
50	1	0	1x(T1)+1x(T2)+1x(T3)+1x(T5) +1x(T6)+1x(T7)+1x(T8)+1x(T9) +1x(T10)+1x(T12)+1x(T15)	P2 / P3 / P4 / P6 / P7	1 / 1 / 1 / 1 / 1
60	1	0	1x(T1)+1x(T2)+1x(T3)+1x(T5) +1x(T6)+1x(T7)+1x(T8)+1x(T9) +1x(T10)+1x(T12)+1x(T15)	P2 / P3 / P4 / P6 / P7	1 / 1 / 1 / 1 / 1
70	1	0	1x(T1)+1x(T2)+1x(T3)+1x(T5) +1x(T6)+1x(T7)+1x(T8)+1x(T9) +1x(T10)+1x(T12)+1x(T15)	P2 / P3 / P4 / P6 / P7	1 / 1 / 1 / 1 / 1
80	1	0	1x(T1)+1x(T2)+1x(T3)+1x(T5) +1x(T6)+1x(T7)+1x(T8)+1x(T9) +1x(T10)+1x(T12)+1x(T15)	P2 / P3 / P4 / P6 / P7	1 / 1 / 1 / 1 / 1
90	1	0	1x(T1)+1x(T2)+1x(T3)+1x(T5) +1x(T6)+1x(T7)+1x(T8)+1x(T9) +1x(T10)+1x(T12)+1x(T15)	P2 / P3 / P4 / P6 / P7	1 / 1 / 1 / 1 / 1

There are irresponsive FMF levels to the pool situation. However, in the 1st pool situation the irresponsive zone extends from Level 70 to Level 90; while the 2nd pool situation has an irresponsive zone that extends from Level 50 to Level 90. This is due to the difference between the entropy levels of the two situations. The irresponsiveness for the two cases is illustrated in Figure 5.3. The dashed arrows are reflected from 60 degree and 40 degree lines for the 1st and 2nd pool situations, respectively. As can be seen from the figure, these are the latest levels before the vertical rays from the related entropy levels reach the horizontal "full process flexibility" line on top part of the graph.

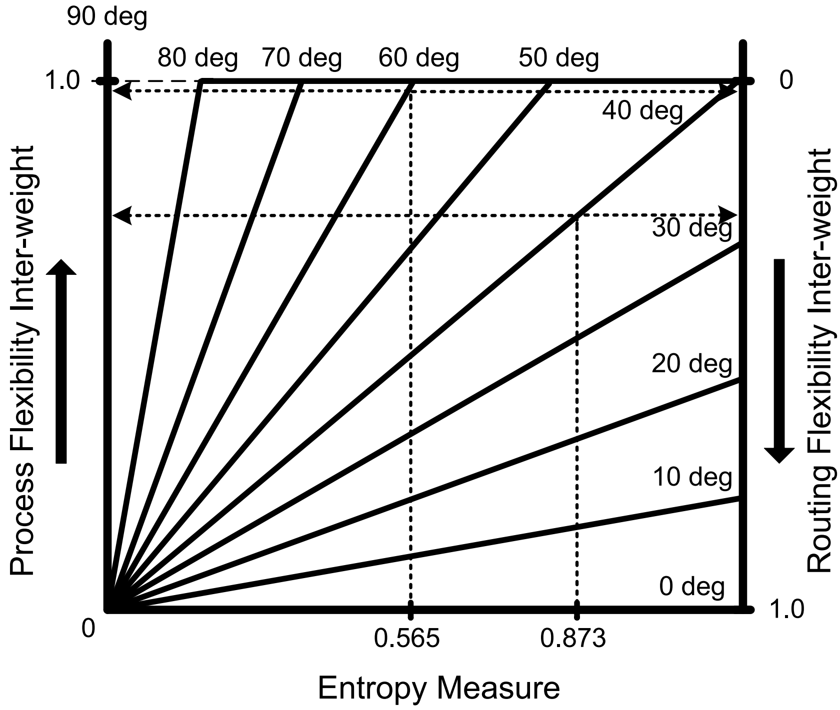


Figure 5.3 Flexibility-Entropy Lines at Different Levels for the two Pool Situations

Neglecting the irresponsive zone, it can be seen from the inter-weights of flexibilities that as the entropy level decreases, the need for routing flexibility increases. Figure 5.4 illustrates the number of producible parts and total number of routes for the two situations at different FMF levels.

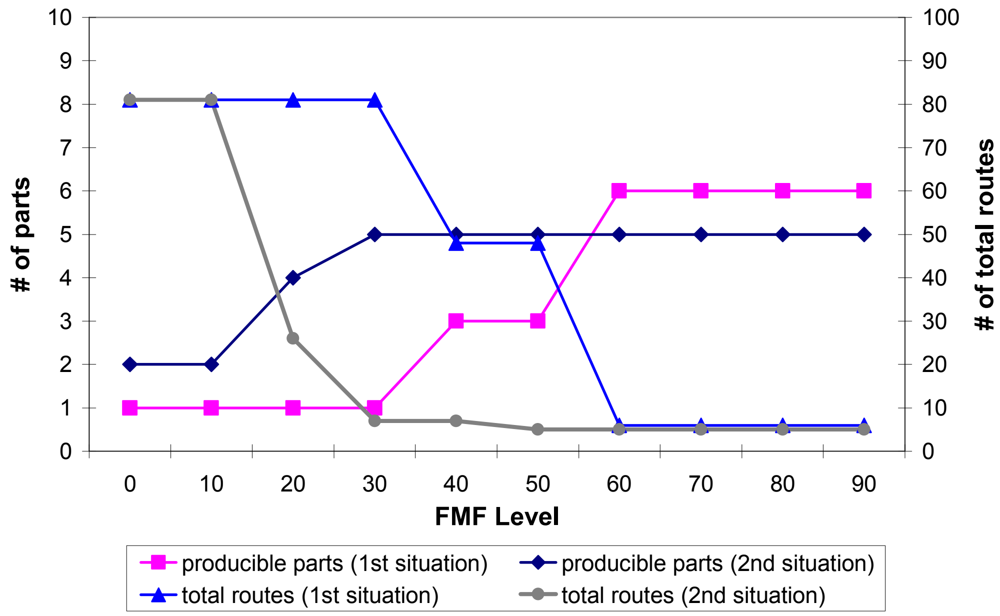


Figure 5.4 Change in Producible Parts and Total Routes for 1st and 2nd Pool Situations

5.4.2 Weighted Flow Time

The results are shown in Figure 5.5 and Table 5.10. The FMF Levels are given in terms of angular degree of Flexibility-Entropy Line. The actual FMF values are the tangent of the related angle. The average column shows the mean flow time values averaged over 70 replications. Half-width column is calculated as defined in 5.3.2. Relative error is the ratio of half-width to the average.

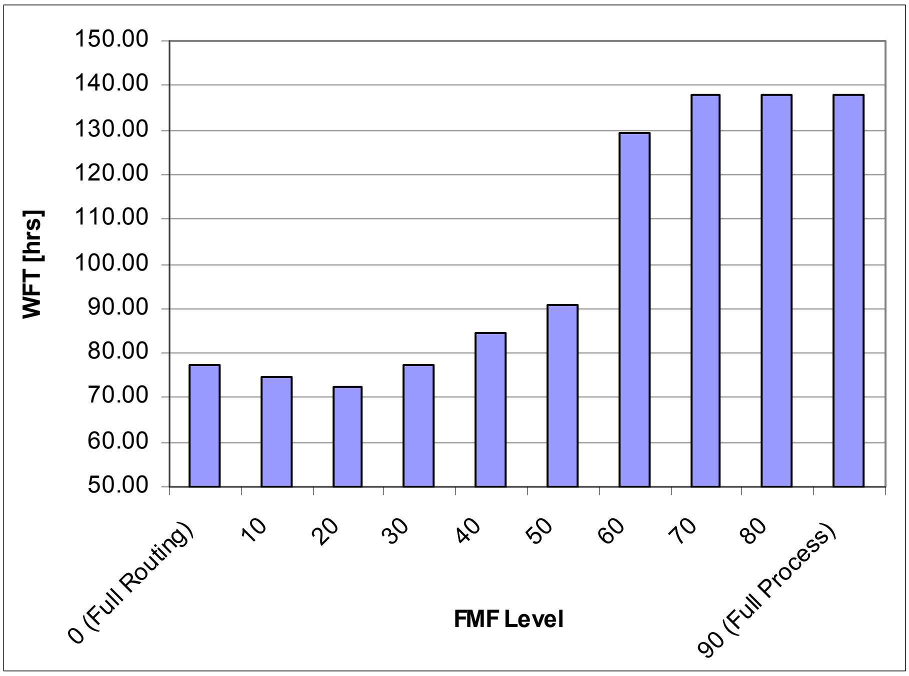


Figure 5.5 Weighted Flow Time-Graphical Results

Table 5.10 Weighted Flow Time-Tabular Results

FMF Level	Weighted Flow Time [hrs]		
	Average	Half Width	Relative Error
0 (Full Routing)	**77.14**	2.51	3.25%
10	**74.78**	2.31	3.09%
20	**72.46**	2.63	3.63%
30	**77.38**	5.15	6.66%
40	**84.59**	5.85	6.92%
50	**90.94**	7.44	8.18%
60	**129.58**	12.73	9.82%
70	**138.02**	12.74	9.23%
80	**138.02**	12.74	9.23%
90 (Full Process)	**138.02**	12.74	9.23%

The first observation is that the minimum point is located neither at the full routing flexibility nor at the full process flexibility. This is an indication of the fact that increasing only one flexibility type does not improve performance of the system. The optimum point is a mixture of different flexibility types that is controlled by FMF.

Since the confidence intervals of FMF Level 0 and FMF Level 20 intersect, a statistical test has been performed in order to show the difference between their means (from [49] pg. 250). The test has been done to strengthen the comments given above.

For the statistical test, it is assumed that the data is normally distributed for each of the two distributions. Then the hypothesis

$$H_0: \mu_1 - \mu_2 = 0$$

is tested against the alternatives

$$H_1: \mu_1 - \mu_2 = \delta < 0$$

The values of the parameters are,

$n_1 = n_2 = 70$

$\bar{x}_1 = 77.14$ (Level 0) & $\bar{x}_2 = 72.46$ (Level 20)

$s_1^2 = 110.69$ (Level 0) & $s_2^2 = 121.53$ (Level 20)

$t_{138;0.95} \sim 1.658$ (from the t tables)

Using these values, the hypothesis test shows that

$$\frac{\bar{x}_1 - \bar{x}_2}{s_w\sqrt{1/n_1 + 1/n_2}} = 2.569 > 1.658$$

This indicates that there is reason to think two means are not equal. Hence, mean flow time value for FMF Level 20 is safely assumed smaller than FMF Level 0.

As a result of the hypothesis test, it has been assumed that the minima is located at FMF Level 20. At FMF Level 20, the reduction of flow time is about 6.1% from FMF Level 0 and 47.5% from FMF Level 90.

By comparing the two extreme sides of the graph, it can be seen that routing flexibility is more effective than the process flexibility in our experimental case.

Another observation is that flow time values are irresponsive to the FMF level changes above level 70. This is related to the operational and physical characteristics of the shop that is used in the experimental study. Past level 70, the routing flexibility is not preferred in any case of the pool. If, however, the volumes of the orders were higher and they were biased in only a few part types, then it would be possible to slide the weight from process flexibility to routing flexibility after FMF Level 70.

Let us say the pool has only part type 1 with a workload of 100 hrs. Using also the anticipated orders data at FMF Level 80, the entropy value becomes 0.594 and inter-weight of process flexibility becomes 1. If the part type has a workload of 1100 hours (which is highly unlikely given the arrival pattern), then the inter-weight starts to move away from process flexibility into routing flexibility. In this case, entropy would become 0.168, and the inter-weights would be 0.95 and 0.05 for the process and routing flexibility, respectively. This shows a very slight and conservative change to preferring few parts with many routes if FMF is set high.

The irresponsiveness may also be related to the linear characteristic of the flexibility-entropy line and the scaling of the weights in the MIP model. If the scaling of the weights was changed or a non-linear flexibility-entropy line was used, the irresponsive zone might have been changed.

An average flow time is also calculated to compare the results with this reference case. Average flow time ($\overline{FT}$) is calculated using probability of arrival of each part (PA_p), unit processing time of each part (UPT_p), average batch size ($\overline{BS}$), and TWK flow allowance parameter F as follows:

$$\overline{FT} = F * \overline{BS} * \sum_{p \in P} PA_p * UPT_p$$

Inserting the related numbers $\overline{FT}$ has been found as:

$$\overline{FT} = 6 * 10 * 1.33 = 79.8\, hrs$$

The average WFT values from Level 0 to Level 30 are all less than $\overline{FT}$. If the 95% confidence intervals are taken into account, then it can be said that the means from Level 0 to Level 20 are all strictly superior to $\overline{FT}$. The highest improvement is about 8 hours (with FMF = 20˚). This means an average order is returned a full workday earlier than the desired lead time setting allows for. This shows that given the TWK flow allowance for the tested shop, the first three levels show a relatively better performance than the expected value.

In conclusion, it can be seen that the convenient level of FMF for the case in experimental study is Level 20, slope of which corresponds to the inter-weight of process flexibility to entropy ratio (tan20 = 0.364). By using this typical management factor, the tested shop can respond to the daily demand changes in a more responsive way on the average than the other levels. The transition from one flexibility type to another is automatically handled by the FMF and the corresponding entropy level.

5.4.3 Weighted Tardiness

The results are shown in Figure 5.6 and Table 5.11. Weighted Tardiness results are in line with the Weighted Flow Time results. The general behavior of the graph is the same and all the observations that are mentioned in Weighted Flow Time apply also to this case.

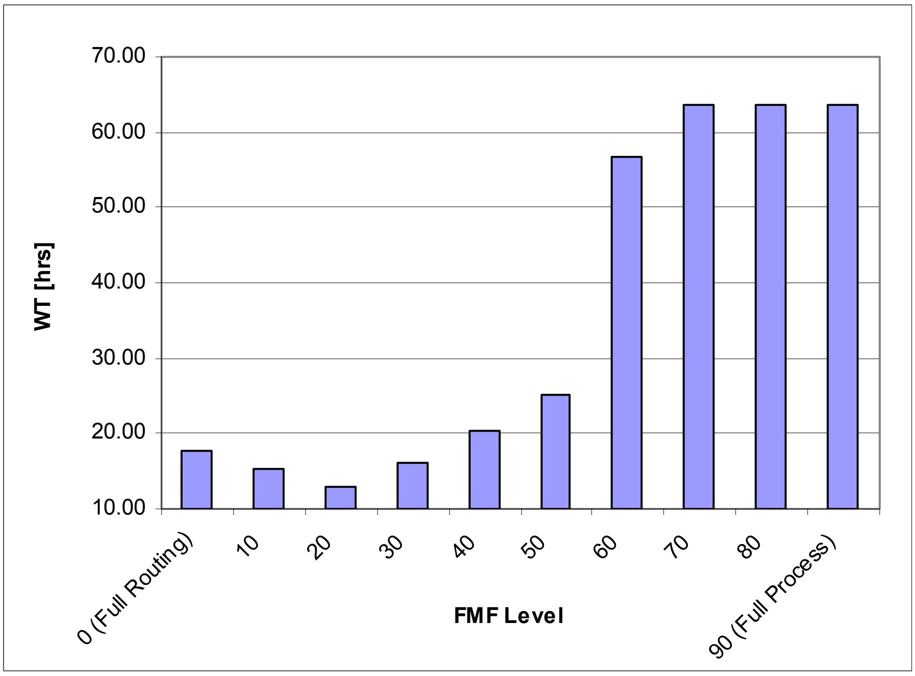


Figure 5.6 Weighted Tardiness-Graphical Results

Table 5.11 Weighted Tardiness-Tabular Results

FMF Level	Weighted Tardiness [hrs]		
	Average	Half Width	Relative Error
0 (Full Routing)	**17.79**	1.36	7.64%
10	**15.33**	1.20	7.83%
20	**13.03**	1.29	9.90%
30	**16.12**	3.43	21.28%
40	**20.42**	4.06	19.88%
50	**25.03**	5.76	23.01%
60	**56.82**	11.01	19.38%
70	**63.58**	11.27	17.73%
80	**63.58**	11.27	17.73%
90 (Full Process)	**63.58**	11.27	17.73%

The minima again appear at Level 20. At FMF Level 20, the reduction of tardiness is about 26.8% from FMF Level 0 and 79.5% from FMF Level 90. This indicates that the management of operational flexibility by choosing appropriate flexibility types and levels not only improves weighted flow time but also benefits in terms of the weighted tardiness.

5.4.4 Percentage of Tardy Orders

The results are shown in Figure 5.7 and Table 5.12. The results are in line with the Weighted Flow Time results, as is the case for Weighted Tardiness. Interestingly, all three graphs have the same characteristics. Appropriate flexibility control positively affects the performance of the shop. Hence, the importance of flexibility management is again highlighted.

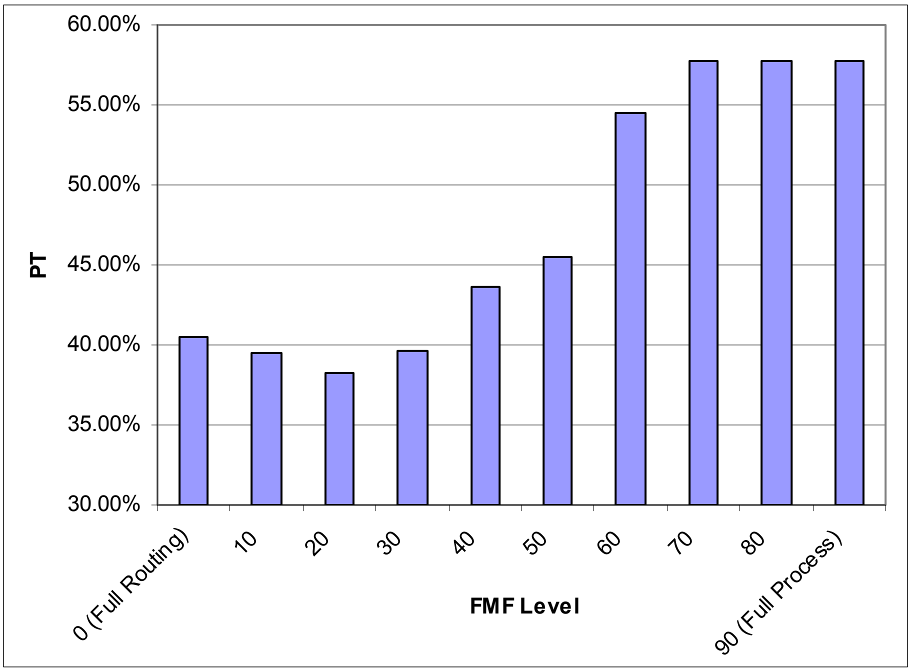


Figure 5.7 Percentage of Tardy Orders-Graphical Results

Table 5.12 Percentage of Tardy Orders-Tabular Results

FMF Level	Percentage of Tardy Orders		
	Average	Half Width	Relative Error
0 (Full Routing)	**40.54%**	0.01	2.47%
10	**39.47%**	0.01	2.53%
20	**38.31%**	0.02	5.22%
30	**39.61%**	0.03	7.57%
40	**43.61%**	0.03	6.88%
50	**45.46%**	0.03	6.60%
60	**54.46%**	0.04	7.34%
70	**57.70%**	0.04	6.93%
80	**57.70%**	0.04	6.93%
90 (Full Process)	**57.70%**	0.04	6.93%

The minima again appear at Level 20. At FMF Level 20, the reduction of percentage of tardy orders is about 5.5% from FMF Level 0 and 33.6% from FMF Level 90.

5.4.5 Utilization Levels

The results are shown in Figure 5.8. The graph shows the utilization of each machine for each FMF Level. Moreover, average system utilization is found by taking arithmetic mean of machine utilizations.

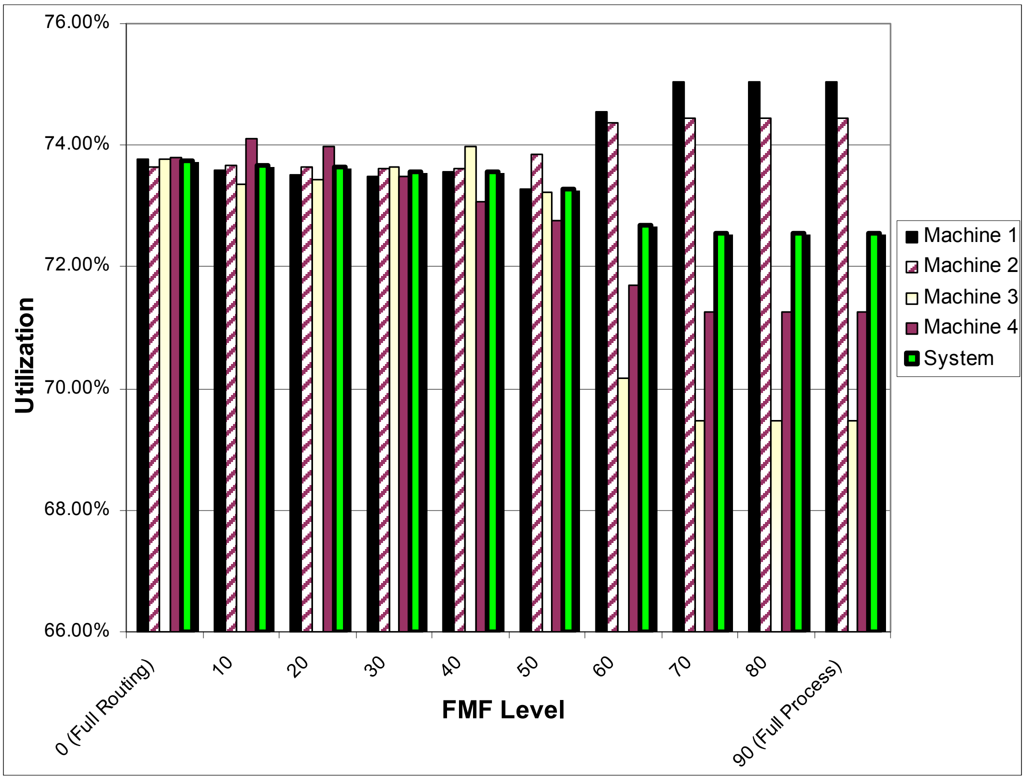


Figure 5.8 Utilization Levels-Graphical Results

It can be observed from the results that the machine utilizations are approximately at the same level from FMF Level 0 to FMF Level 50. There is a slight decrease after FMF Level 60. However, the important point is the reduction in the uniformity of workload distribution among machines after FMF Level 60. These results indicate that the congestion on the shop floor increases as the setups allow larger variety of part types on the shop. Increasing the weight of process flexibility adversely affected the workload distribution on the shop floor; as a result overall system utilization decreased slightly.

5.4.6 Average Length of Production Cycle

The results are shown in Table 5.13. The table shows the average, minimum and maximum values of average length of production cycle over 70 replications at each FMF level. The results show that the cycle lengths are approximately at the same level. However, there is a slight increase from FMF Level 0 to Level 90. This indicates that as the process flexibility is increased, the overtime tendency increases. However, the overtime lengths are around ~3-5% of normal working hours, which can be acceptable. Hence, it can be said that the "decreasing workload control norms" concept is effective on controlling the release of jobs to the shop floor.

Table 5.13 Average Length of Production Cycle-Tabular Results

FMF Level	Average Length of Production Cycle [hrs]		
	Avg.	Min.	Max.
0 (Full Routing)	**40.84**	40.54	41.10
10	**40.84**	40.54	41.40
20	**40.96**	40.54	41.38
30	**41.24**	40.82	41.67
40	**41.41**	40.82	41.96
50	**41.55**	41.10	42.55
60	**41.92**	41.38	42.55
70	**41.97**	41.67	42.55
80	**41.97**	41.67	42.55
90 (Full Process)	**41.97**	41.67	42.55

CHAPTER 6

CONCLUSION

The aim of this thesis study is to highlight the benefits of using a flexibility oriented shop management approach in make-to-order companies and show how the order release techniques can be combined with a flexible shop environment in order to replace detailed scheduling by planning for demanded flexibility only. We propose to affect the “releasability” of pool orders through configuring the flexibility mix. This adds a dimension to output control previously not fully addressed in make-to-order manufacturing.

In this study, we have developed a flexibility management approach that can deal with the inherent uncertainty contained in make-to-order manufacturing policies. It has been mentioned in the related literature that effective use of flexibility can be realized by specification of an FM Policy. This policy is made simply dependent on only a single parameter, namely Flexibility Management Factor (FMF). Two main tools have been used in connection to determine the proper FMF setting for a shop which are the mathematical model and the simulation model. The mathematical model prepares a shop environment according to the order data of the pool and expected orders at the setup moments. With the simulation model, the performance of the chosen FM policy is monitored and analyzed.

Two flexibility types are addressed in response to short-term demand fluctuations in the orders. These are process and routing flexibilities. As these flexibility types can be adjusted dynamically by changing only operation-machine assignments in the shop, they are suitable for a routine operational management control concept. We have proposed measures to quantify process and routing flexibilities in the system. In summary, process flexibility of a shop is measured by the ratio of number of part types that can be realized in the shop with the current setup to the universe of part types that the shop can produce with the proper setup; routing flexibility of a part is measured by the ratio of number of available routes for the part with the current setup to the maximum number of possible routes that can be opened for the part.

The demand for different types and levels of flexibility, and the distribution of flexibility among parts have been tied to the existing entropy level in the order pool and criticality of the orders. The higher the current entropy, the higher the demand for process flexibility; the higher the criticality of a part type order than all the other orders, the more that particular part type will benefit from the allocated flexibility.

We also proposed an extension to the workload control approach found in [4] and [5] by introducing the decreasing norms concept that is used to cease the order releases in a controlled manner as the new setup moment approaches.

An experimental study is performed in order to analyze the impacts of flexibility management approach on a flexible shop environment. In this study, workload control techniques have been combined with the flexibility management approach for a make-to-order shop. The results show that the flexibility management can be an effective tool for compensating uncertainty in demand. Determining an appropriate FM Policy for a shop improved the performance of the system in all types of measures.

The results show that the process and routing flexibilities place antithetical requirements on the system as [24] has also noted and consistent distribution of flexibility between flexibility types and within each flexibility type provides the highest gains. Although weighing the routing flexibility seems to be favored in the choice of flexibility-entropy line slope, this should not be misleading. A balanced mix in the pool still leads to higher values of FMF. This in turn will affect the resulting weight.

Another outcome of the experimental study is that while process and routing flexibilities are suitable for short-term planning decisions, increasing one type of flexibility to its upmost limits does not better the performance. It was highlighted that a mixture of the flexibility types can respond to daily demand changes in a more efficient way if physical and operational characteristics of the shop, the orders in the pool and expected orders are simultaneously taken into account. Thus, we propose to provide room for variety (process) and redundancy (routing) in an overall sense rather than trying either to schedule an uncertain environment or to infer all possible future flows among the machines with certainty.

It should be noted that the findings of the experimental study are specific for the stated conditions. Further studies can be performed by changing number of machines in the system and related parameters, adjusting expected utilization rates for tighter conditions, trying various dispatching rules, generating different conditions for number of part types, batch sizes, probability of arrival of each part, part-operation table, and unit processing times of operations.

The mathematical programming and simulation models can be further extended to include partial loading of orders for the same part, changeover dependent setup times and varying sizes in production intervals. For instance, an additional flexibility type can be added to the mathematical model that is "Product Flexibility". Introduction of product flexibility into the model can extend the capability of the mathematical model to decide on the proper setup moment and affect the tool variety, duplication of the tools and tool allocations to the machines. This can be an interesting extension to the approach.

Possible future works can include relaxation of “Full Machine Flexibility” and tool capacity assumptions. With these assumptions are relaxed, it can be examined how the behavior of the model changes when creating a shop environment with different types of machines. Moreover, “Operation Flexibility” can be provided to some of the part types in order to generate different types of routings for the same parts. This may give the parts the opportunity to pick the appropriate route in different pool and shop situations.

Inclusion of tool availability and tool life can constitute an interesting addition to the model. In this case, the mathematical model may decide to allocate several copies of the same tool on the same machine, or combined with “Operation Flexibility”, it may help analyze assignment of different routings for the same part type.

Other assumptions such as “one slot for one tool” assumption for tool magazines and “material handling system” assumption can easily be relaxed to better approximate real shop characteristics. The restrictions on minimal load may be lifted to a certain extent to create a more tolerant solution.

In this study, WLC concept has been combined with the FM approach. Other ORR strategies can be combined with our approach to test the effectiveness of different ORR-FM combinations.

Another possible future work is the integration of the alternative method for approximating FFRFM (4.1.3.1) with the mathematical model. This may improve the performance of the shop as the maximum deviation between approximation line and FFRFM will be minimized.

The elements for the flexibility allocation can be further studied. The flexibility-entropy line can be converted to a convex, concave or an S-shaped curve. With proper scaling of the weights, the parameters of the curves can be changed to search for an optimum FM Policy. These policies can be used to investigate the proper shape of the flexibility-entropy graph.

The calculations for estimated order arrivals is a major research topic that may affect the performance measures in a more favorable way. Anticipations based on different order arrival assumptions (pool content dependent or some orders with periodicity), forecasts or service level requirements can be included into the model.

REFERENCES

[1] Wisner, J.D., "A Review of the Order Release Policy Research", International Journal of Operations & Production Management, vol. 15, no. 6, pp. 25-40, 1995

[2] Bergamaschi, D., Cigolini, R., Perona, M., Portioli, A., "Order Review and Release Strategies in a Job Shop Environment: A Review and a Classification", International Journal of Production Research, vol. 35, no. 2, pp. 399-420, 1997

[3] Wight, W.O., "Input/Output Control: A Real Handle on Lead Times", Production and Inventory Management, vol. 11, pp. 9-30, 1970

[4] Land, M., Gaalman, G., "Workload Control Concepts in Job Shops: A Critical Assesment", International Journal of Production Economics, vol. 46-47, pp. 535-548, 1996

[5] Land, M., "Parameters and Sensitivity in Workload Control", International Journal of Production Economics, vol. 104, 625-638, 2006

[6] Cigolini, R., Portioli, A., "An Experimental Investigation on Workload Limiting Methods within ORR Policies in a Job Shop Environment", Production Planning & Control, vol. 13, no. 7, pp. 602-613, 2002

[7] Sabuncuoglu, I., Karapinar, H.Y., "A Load-based and Due-date-oriented Approach to Order Review/Release in Job Shops", Decision Sciences, vol. 31, no. 2, pp. 413-447, 2000

[8] Kingsman, B., Hendry, L., "The Relative Contributions of Input and Output Controls on the Performance of a Workload Control System in Make to Order Companies", Production Planning & Control, vol. 13, no. 7, pp. 579-590, 2002

[9] Sethi, A.K., Sethi, S.P., "Flexibility in Manufacturing: A Survey", International Journal of Flexible Manufacturing Systems, vol. 2, pp. 289-328, 1990

[10] Stecke, K.E., "Planning and Scheduling Approaches to Operate a Particular FMS", European Journal of Operational Research, North Holland, vol. 61, pp. 273-291, 1992

[11] Lingayat, S., Mittenthal, J., O'Keefe, R.M., "An Order Release Mechanism for a Flexible Flow System", International Journal of Production Research, vol. 33, no. 5, pp. 1241-1256, 1995

[12] Newman, W.R., Maffei, M.J., "Managing the Job Shop: Simulating the Effects of Flexibility, Order Release Mechanisms and Sequencing Rules", Integrated Manufacturing Systems, vol. 10, no. 5, pp. 266-275, 1999

[13] Chan, F.T.S., Bhagwat, R., Wadhwa, S., "Increase in Flexibility: Productive or Counterproductive? A Study on the Physical and Operating Characteristics of a Flexible Manufacturing System", International Journal of Production Research, vol. 44, no. 7, pp. 1431-1445, 1 April 2006

[14] Calvo, R., Domingo, R., Sebastian, M.A., "Operational Flexibility Quantification in a Make-to-Order Assembly System", International Journal of Flexible Manufacturing Systems, vol. 19, pp. 247-263, 2007

[15] Nomden, G., van der Zee, D.J., "Virtual Cellular Manufacturing: Configuring Routing Flexibility", International Journal of Production Economics, vol. 112, pp. 439-451, 2008

[16] Daniels, R.L., Mazzola, J.B., Shi, D., "Flow Shop Scheduling with Partial Resource Flexibility", Management Science, vol. 50, no. 5, pp. 658-669, May 2004

[17] Wahab, M.I.M., Stoyan, S.J., "A Dynamic Approach to Measure Machine and Routing Flexibilities of Manufacturing Systems", International Journal of Production Economics, vol. 113, pp. 895-913, 2008

[18] Chang, A.Y., "An Attribute Approach to the Measurement of Machine-group Flexibility", European Journal of Operational Research, 2008

[19] Henrich, P., Land, M.J., Gaalman, G.J.C., "Semi-interchangeable Machines: Implications for Workload Control", Production Planning & Control, vol. 18, no. 2, pp. 91-104, March 2007

[20] Corsten, H., Gössinger, R., Wolf, N., "Flexibility-driven Order Releases in Job-shop Production", Technovation, vol. 25, pp. 815-830, 2005

[21] Corsten, H., Gössinger, R., "Opportunistic Coordination of Operations in Job Shop Production", Technovation, vol. 24, pp. 219-234, 2004

[22] Beach, R., Muhlemann, A.P., Price, D.H.R., Paterson, A. Sharp, J.A., "A Review of Manufacturing Flexibility", European Journal of Operational Research, vol. 122, pp. 41-57, 2000

[23] Groover, M.P., "Automation, Production Systems and Computer-Aided Manufacturing", Prentice Hall, 1980

[24] Stecke, K.E., Raman, N., "FMS Planning Decisions, Operating Flexibilities, and System Performance", IEEE Transactions on Engineering Management, vol. 42, no. 1, February 1995

[25] Jaikumar, R., "Postindustrial Manufacturing", Harvard Business Review, vol. 64, pp. 69-76, November-December 1986

[26] Ettlie, J.E., "Implementation Strategies for Discrete Parts Manufacturing Technologies", Final Report, The University of Michigan, Ann Arbor, MI, April 1988

[27] Kumar, V., "Entropic Measures of Manufacturing Flexibility", International Journal of Production Research, vol. 25, no. 7, pp. 957-966, 1987

[28] Smith, M.L., Seidmann, A., "Due Date Selection Procedures for Job-Shop Simulation", Computer and Industrial Engineering, vol. 7, no. 3, pp. 199-207, 1983

[29] Akçalı, E., Uzsoy, R., Hiscock, D.G., Moser, A.L., Teyner, T.J., "Alternative Loading and Dispatching Policies for Furnace Operations in Semiconductor Manufacturing: A Comparison by Simulation", Proceedings of the 2000 Winter Simulation Conference, 2000

[30] Chiang, T., Fu, L., "Parameter Tuning of Production Scheduling Rules by an Ant System-Embedded Genetic Algorithm", Proceedings of the 2004 IEEE Conference on Robotics, Automation and Mechatronics, Singapore, 1-3 December 2004

[31] Chiang, T., Fu, L., "Solving the FMS Scheduling Problem by Critical Ratio-Based Heuristics and the Genetic Algorithm", Proceedings of the 2004 IEEE International Conference on Robotics and Automation, New Orleans, April 2004

[32] Rardin, R.L., "Optimization in Operations Research", Prentice Hall International Inc., 1998

[33] Shanker, K., Tzen, Y.J.J., "A Loading and Dispatching Problem in a Random Flexible Manufacturing System", International Journal of Production Research, vol. 23, no. 3, pp. 579-595, 1985

[34] Ro, I.K., Kim, J.I., "Multi-Criteria Operational Control Rules in Flexible Manufacturing Systems (FMSs)", International Journal of Production Research, vol. 28, no. 1, pp. 47-63, 1990

[35] Melnyk, S.A., Ragatz, G.L., "Order Review/Release: Research Issues and Perspectives", International Journal of Production Research, vol. 27, no. 7, pp. 1081-1096, 1989

[36] Ragatz, G.L., Mabert, V.A., "An Evaluation of Order Release Mechanisms in a Job-Shop Environment", Decision Sciences, vol. 19, pp. 167-189, 1988

[37] Oosterman, B., Land, M., Gaalman, G., "The Influence of Shop Characteristics on Workload Control", International Journal of Production Economics, vol. 68, pp. 107-119, 2000

[38] Land, M., "Workload Control in Job Shops, Grasping the Tap", PhD. Thesis, University of Groningen, 2004

[39] Hendry, L.C., "A Decision Support System to Manage Delivery and Manufacturing Lead Times in Make-to-Order Companies", PhD. Thesis, University of Lancaster, 1989

[40] Tatsiopoulos, I.P., "A Microcomputer-based Interactive System for Managing Production and Marketing in Small Component-Manufacturing Firms Using a Hierarchical Backlog Control and Lead Time Management Methodology", PhD. Thesis, University of Lancaster, 1983

[41] Stevenson, M., Hendry, L.C., "Aggregate Load-oriented Workload Control: A Review and a Re-classification of a Key Approach", International Journal of Production Economics, vol. 104, pp. 676-693, 2006

[42] Chan, F.T.S., Bhagwat, R., Wadhwa, S., "Taguchi's Method Analysis of an FMS Under Review Period Based Operational Controls: Identification of Control Periodicity", IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, vol. 37, no. 2, March 2007

[43] Dellaert, N.P., "Production to Order: Models and Rules for Production Planning", Lecture Notes in Economics and Mathematical Systems, vol. 333, Springer-Verlag, Berlin, 1989

[44] Goswami, M., Tiwari, M.K., Mukhopadyay, S.K., "An Integrated Approach to Solve Tool-Part Grouping, Job Allocation and Scheduling Problems in a Flexible Manufacturing System", International Journal of Advanced Manufacturing Technology, 2006

[45] Montazeri, M., Van Wassenhove, L.N., "Analysis of Scheduling Rules for an FMS", International Journal of Production Research, vol. 28, no. 4, pp. 785-802, 1990

[46] Amoako-Gyampah, K., Meredith, J.R., "A Simulation Study of FMS Tool Allocation Procedures", Journal of Manufacturing Systems, vol. 15, no. 6, 1996

[47] Morton, T.E., Smunt, T.L., "A Planning and Scheduling System for Flexible Manufacturing Systems", Flexible Manufacturing Systems: Methods and Studies, edited by A. Kusiak, pp. 151-164, Elsevier Science Publishers, Amsterdam, 1986

[48] Law, A.M., Kelton, W.D., "Simulation Modeling and Analysis", 2nd Edition, McGraw-Hill Inc., 1991

[49] Guttman, I., Wilks, S.S., Hunter, J.S., "Introductory Engineering Statistics", 2nd Edition, John Wiley & Sons Inc., 1971

[50] Ragatz, G.L., Mabert, V.A., "A Simulation Analysis of Due Date Assignment Rules", Journal of Operations Management, vol. 5, pp. 27-39, 1984

APPENDIX A

GAMS CODES FOR THE MATHEMATICAL MODEL

```
$Title  Flexibility Allocation Problem

Sets
oper             / o1*o15 /
mac              / m1*m4 /
part             / p1*p8 /
seq              / seq0*seq6 /
wflexintra       / wpintra, wrintra /
workload         / wl /
uptopers         / uptoper /
uptparts         / uptpart /
;

Alias(oper,operp)
;
Alias(mac,macp)
;
Alias(seq,seqp)
;

Table partseq(part,seq) part-operation table

      seq0     seq1     seq2     seq3     seq4     seq5     seq6
p1    14       3        7        9        10       13       14
p2    12       5        8        12       0        0        0
p3    15       3        7        10       15       0        0
p4    6        2        6        0        0        0        0
p5    13       4        5        8        13       0        0
p6    10       3        7        9        10       0        0
p7    6        1        2        6        0        0        0
p8    11       4        5        8        10       11       0
;
Table operparttab(oper,part) operation types that are applicable to
part types

      p1      p2      p3      p4      p5      p6      p7      p8
o1    0       0       0       0       0       0       1       0
o2    0       0       0       1       0       0       1       0
o3    1       0       1       0       0       1       0       0
o4    0       0       0       0       1       0       0       1
o5    0       1       0       0       1       0       0       1
o6    0       0       0       1       0       0       1       0
o7    1       0       1       0       0       1       0       0
o8    0       1       0       0       1       0       0       1
o9    1       0       0       0       0       1       0       0
o10   1       0       1       0       0       1       0       1
o11   0       0       0       0       0       0       0       1
o12   0       1       0       0       0       0       0       0
o13   1       0       0       0       1       0       0       0
o14   1       0       0       0       0       0       0       0
o15   0       0       1       0       0       0       0       0
;
```

```
Table uptopertab(oper,uptopers) tabular unit processing times of
operations

         uptoper
o1       0.30
o2       0.41
o3       0.36
o4       0.49
o5       0.22
o6       0.26
o7       0.44
o8       0.11
o9       0.46
o10      0.32
o11      0.33
o12      0.42
o13      0.14
o14      0.45
o15      0.25
;
Table uptparttab(part,uptparts) tabular unit processing times of
parts

         uptpart
p1       2.17
p2       0.75
p3       1.37
p4       0.67
p5       0.96
p6       1.58
p7       0.97
p8       1.47
;

Parameter count(part) number of operations for a part;
         count(part) = sum(seq, (partseq(part,seq)>0))-1;
         display count
;
Parameter dmax(part) maximum number of additional routes for a part;
         dmax(part) = count(part)*(card(mac)-1);
         display dmax
;
Parameter PF(part)   penalty factor / p1   0.85, p2   0.43, p3
0.62, p4   0.25, p5   0.62, p6   0.62, p7   0.43, p8    0.76/
;
Parameter weightpart(wflexintra,part) intraweights of different
parts
;
Parameter wprocinter interweight of process flexibility
;
Parameter wroutinter interweight of routing flexibility
;
Parameter workloadpart(workload,part) total workload status of the
parts
;

Variables
x(oper,mac)
b(oper)
c(part)
ctot
y(oper,mac,part)
t(seq,part)
epos(seq,part)
eneg(seq,part)
d(part)
dtot
wl(oper,mac)
z
```

```
*used for RUN transition between GAMS and ARENA
argams
;

Positive Variable y, t, epos, eneg, d, ctot, dtot, wl;
Binary Variable x, b, c;

Equations
opt
* process flexibility
minonemac(oper)
partshow(part)
ctotal
* routing flexibility
assignopermac(part,seq,oper,mac)
assignpart(part,seq,oper,mac)
countoprt(part, seq)
diffseqtoseq(part,seq)
counttotaddrt(part)
dtotal
* maximum tool capacity
maxtool(mac)
* workload
workloadballow(mac)
workloadshrok(oper,mac)
workloadshr(oper)
* used for RUN transition between GAMS and ARENA
argamseq
;

opt .. z =e= wprocinter*ctot + wroutinter*dtot;

minonemac(oper) .. sum((mac), x(oper,mac)) =g= b(oper);

partshow(part) .. sum((seq,oper)$((partseq(part,seq) > 0) and
(ord(seq) > 1) and (ord(oper) = partseq(part,seq))), b(oper)) =g=
count(part)*c(part);

ctotal .. ctot =e= sum((part), weightpart('wpintra',part)*c(part));

assignopermac(part,seq,oper,mac)$((partseq(part,seq) > 0) and
(ord(seq) > 1) and (ord(oper) = partseq(part,seq))) ..
y(oper,mac,part) =l= x(oper,mac);

assignpart(part,seq,oper,mac)$((partseq(part,seq) > 0) and (ord(seq)
> 1) and (ord(oper) = partseq(part,seq))) ..  y(oper,mac,part) =l=
c(part);

countoprt(part, seq)$(partseq(part,seq) > 0) .. t(seq,part) =e=
sum((oper, mac)$(ord(oper) = partseq(part,seq)), y(oper,mac,part));

diffseqtoseq(part,seq)$((partseq(part,seq) > 0) and (ord(seq) > 1))
.. t(seq-1,part) - t(seq,part) =e= epos(seq,part) - eneg(seq,part);

counttotaddrt(part) .. sum((seq)$((partseq(part,seq) > 0) and
(ord(seq) > 1)), t(seq,part)) - count(part)*c(part) -
PF(part)*sum((seq)$((partseq(part,seq) > 0) and (ord(seq) > 1)),
eneg(seq,part)) =g= d(part);

dtotal .. dtot =e= sum((part),
weightpart('wrintra',part)*d(part)/dmax(part));

maxtool(mac) .. sum((oper), x(oper,mac)) =l= 3;

workloadballow(mac) .. sum((oper), wl(oper,mac)) =g= 0.65*40;

workloadshrok(oper,mac) .. wl(oper,mac) =l= 100000*x(oper,mac);
```

```
workloadshr(oper) .. sum((mac), wl(oper,mac)) =l=
sum((part)$(operparttab(oper,part)=1),
uptopertab(oper,'uptoper')*workloadpart('wl',part)/uptparttab(part,'
uptpart')*c(part));

*used for RUN transition between GAMS and ARENA
argamseq .. argams =e= 1;

Model flex /all/ ;

*===Import from Excel PART INTRAWEIGHTS & FLEXIBILITY INTERWEIGHTS
*===UNLOAD
$CALL GDXXRW.EXE partweight.xls par=weightpart rng=sheet1!A2:I4
par=wprocinter rng=sheet1!B13 Dim=0 par=wroutinter rng=sheet1!B14
Dim=0
par=workloadpart rng=sheet1!B23:J24 Cdim=1 Rdim=1
*===IMPORT
$GDXIN partweight.gdx
$LOAD weightpart
$LOAD wprocinter
$LOAD wroutinter
$LOAD workloadpart
$GDXIN
display weightpart;
display wprocinter;
display wroutinter;
display workloadpart;

flex.reslim = 1000;
flex.nodlim = 1000000;
option iterlim=1000000,domlim=100000;
option limrow=30;
option optcr=0.00;

Solve flex using mip maximizing z ;

Display x.l, x.m;

*===Export to Excel
*===UNLOAD
execute_unload 'macoper.gdx',x,argams;
*===EXPORT
execute 'gdxxrw.exe macoper.gdx o=macoper.xls SQ=N var=x
rng=Sheet1!A2:E17 rdim=1 cdim=1'
execute 'gdxxrw.exe macoper.gdx o=exchange.xls SQ=N var=argams'
```

APPENDIX B

VBA CODES FOR INTERACTION OF GAMS & ARENA

B.1 GAMS TO ARENA

```
Private Sub Worksheet_Change(ByVal Target As Range)

        AppActivate "Arena"

        ' delay
        Start = Timer    ' Set start time.
        PauseTime = 0.1  ' Set duration.
        Do While Timer < Start + PauseTime
         DoEvents    ' Yield to other processes.
        Loop

        SendKeys "{F5}", True

        AppActivate "gamside"

End Sub
```

B.2 ARENA TO GAMS

```
Private Sub VBA_Block_1_Fire()

AppActivate "gamside"

        ' delay
        Start = Timer   ' Set start time.
        PauseTime = 1  ' Set duration.
        Do While Timer < Start + PauseTime
         DoEvents   ' Yield to other processes.
        Loop

        SendKeys "^{TAB}", True

          ' delay
        Start = Timer   ' Set start time.
        PauseTime = 0.1  ' Set duration.
        Do While Timer < Start + PauseTime
         DoEvents   ' Yield to other processes.
        Loop

        SendKeys "^%{TAB}", True

           ' delay
        Start = Timer   ' Set start time.
        PauseTime = 0.1  ' Set duration.
        Do While Timer < Start + PauseTime
         DoEvents   ' Yield to other processes.
        Loop

        SendKeys "{F9}", True

End Sub
```