1991

Development of theory of constraint-based framework for determining process batch, transfer batch and buffer size in shop floor

Tajinder P.S. Vohra
Lehigh University

Recommended Citation

Vohra, Tajinder P.S., "Development of theory of constraint-based framework for determining process batch, transfer batch and buffer size in shop floor" (1991). *Theses and Dissertations*. 5398.
https://preserve.lehigh.edu/etd/5398

Development of Theory of Constraint-based Framework for determining process batch, transfer batch and buffer size in shop floor

by

Tajinder P. S. Vohra

A Thesis
Presented to the Graduate Committee
of Lehigh University
in candidacy for the Degree of
Master of Science
in
Manufacturing Systems Engineering

Lehigh University
1990

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

Roger N Nagel

JUNE 26, 1990

Dr. Roger N. Nagel,
Advisor

Date

6/28/90.

Dr. Keith M. Gardiner,
Director,
Manufacturing Systems Engineering

Date

M U Thomas

6/29/90.

Dr. Marlin U. Thomas,
Chairperson,
Industrial Engineering

Date

Acknowledgments

I wish to express my sincere gratitude to my advisor, Dr. Roger N. Nagel, for his guidance and encouragement that helped me in preparing this thesis.

I would also like to acknowledge Lehigh University's Intelligent Systems Laboratory for the financial support which made my graduate studies possible.

Table of Contents

List of Figures

Abstract

In a highly dynamic and competitive manufacturing world, need for new and innovative management and operating philosophies is patent. Traditional methodologies have proved their irrelevance to existing milieu in the numerous companies that failed to change to new philosophies. Customers demand high quality, low costs and wider variety of products. This, along with stiff competition from competitors is forcing many organizations to rethink their operating goals. It has become obvious that local optima may, and in fact often, conflict with global optima. The only way to achieve the global optima, which is what a company should be aiming at, is through viewing the plant operations with a systems thinking approach.

Theory of Constraint is one such emerging philosophy that could create a more rational perspective for managing the shop floor operations. The basic premise of this theory is that it is logical to focus on the constraint or the bottleneck in an organization instead of attempting to manage everything. The simplicity of the technique is in fact what makes it so relevant and effective. Over the years, manufacturing activities have become highly specialized and localized within a plant. This has led to local optimizations that have often infringed upon global

optimization. The reasons for such contradictions have been a lack of systems thinking and archaic operating procedure, specially the accounting protocols.

Theory of Constraint aims at regulating the part flow in a production system according to the production rate of the constraint. The underlying reason being that the production rate of a line cannot exceed that of the constraint and any excess workload at the other non-constraint machines will lead to additional work in process inventories.

It has also been realized that smaller process and transfer batches lead to shorter lead times, which subsequently get translated into greater flexibility of the system. An ideal system can be visualized as that which has the process as well as the transfer batch equal to a single unit. In practical cases, the capacity constraint at the bottleneck and the setup time and cost involved for the process batches forces a compromise to be made to balance the advantages of smaller process batch size and increased setup costs involved for greater number of setups.

This theory was used as a basis for developing a framework for determining the process batch, transfer batch and buffer sizes. The capacity at the constraint machine is used as the basis for formulating a model which uses the expected demand for different products for a given time horizon. The model then generates suggested lot sizes for each product with an intent to minimize the average lot processing time. The reason for following such an objective function is to induce greater flexibility in the system by cutting down the flow time.

The second section of the thesis presents simulation analysis of a production line with the purpose of defining guidelines for determining buffer sizes. The production line was simulated for different operating conditions and the performance parameters were analyzed. This analysis was then used to form guidelines.

Chapter1

INTRODUCTION

Thesis Objective

The objective of this thesis is to define a framework for determining the process batch, transfer batch and buffer sizes in a shop floor with the intent to minimize the response time or, conversely, to increase the flexibility of the line. The flexibility has been defined as the average time taken to introduce and process a new lot in the system. The framework is based on "Theory of Constraint" concepts whereby the batch sizing decisions can be based on the production constraint only. The framework will generate suggested process and transfer batches and buffer sizes for the machines based on the expected demands for a given time horizon. The impact of the suggested approach on a production environment has been elucidated by examples of such implementations in the thick-film print facility at an electronics firm.

The first chapter of the thesis describes the new manufacturing environment. A description of the changing priorities and the new performance measures that are required is given. The second chapter describes the basic concepts of the theory of constraint and its relevance to the changes in the manufacturing milieu. The third chapter is concerned with description of the framework for determining the process batch, transfer batch and the buffer sizes at the machines. The fourth chapter presents a detailed analysis of simulation results for a line to determine the buffer sizing guidelines for the constraint machine. Finally the proposed framework has been explained in the light of the actual operating strategy at the thick film line.

The New Manufacturing Environment

A rapidly changing milieu in manufacturing and stiff competition is forcing most of the manufacturing concerns to reconsider their operating strategies and priorities on the shop floor. It has become imperative for the organizations to produce high quality goods of a wide variety with extremely short lead times and all this has to be performed in a highly cost efficient environment. Within this dynamic scenario for manufacturing have emerged numerous management techniques like JIT, MRP, TQC, SQC, Group Technology etc. Each of these techniques aims at improvement in the organizations operations [21]. JIT is focused at eliminating waste through reduction in work in process inventory. MRP attempts to create an infrastructure for shop floor control by managing the materials movement. TQC and SQC were aimed at improving quality through scientific application of statistical techniques to the processes. Group technology was developed to create more focused factories and by reconfiguring the traditional shop floor into cells for easier supervision and control [25].

Most of the above mentioned approaches and philosophies tended to be localized to specific aspects of the shop floor. JIT, though, was a philosophy based on total systems thinking and viewed the entire organization as interlinked and interacting set of components each of which had to be considered [22]. Advent of JIT in the US manufacturing environment heralded the concept of systems thinking. The conflict between local and global optimization was acknowledged and the invalidity of many traditional approaches to assess shop floor performances were comprehended [8]. With so many organizations trying to improve their flexibility, it is expected that a measure to gauge the flexibility of a system will be needed. It has also been comprehended that such measures have to be based on more rational concepts based on systems thinking.

Flexibility

The need for measuring flexibility is heightened by the fact that it is often difficult or sometimes impossible to specify a benchmark for flexibility. Yet it is possible to quantify the flexibility in some areas of an organization by considering the impact that flexibility has on the system. Flexibility in a manufacturing enterprise's shop floor can be quantified as the time it takes to setup and start producing the first part after a changeover. Another measure will be the number of different products that can be simultaneously produced for a given configuration since it directly affects the product mix. System flexibility can also be compared by comparing the ratio of the protective capacity needed by the two. It is possible that better scheduling might reduce the need for total amount of protective capacity required. Time still is of paramount importance since it can quantify the speed with which the system can respond to demand changes.

It is also necessary to comprehend the significance of flexibility. An appropriate method to view the significance of flexibility is to see flexibility as one of the means to reach the goals of the organization as a whole. Often, stiff competition and economy threaten the very existence of a company. In such a case flexibility will be a means for survival of the company. For other companies, even though well established in the market, a need to respond quickly to the customer demands is imperative for the company to stay ahead or at least at par with the competition. At a larger perspective, the fact that the global economic environment is in flux, is obvious and so is the need for an organization to adapt itself to those changes or risk obsolescence. There are other equally important issues that flexibility can address. For example, a need for a better working conditions has led many companies to rethink the traditional way of operating and many have introduced flextime and telecommuting. This type of operating flexibility allows the working environment to be made more attractive to people.

The flexibility issue addressed in this thesis is related to the reduction of lead times on the shop floor. The framework developed in this thesis aims at an increased flexibility on the shop floor by decreasing the average lot processing time on the floor. The inherent premise is that reduction in the lot processing time enables a wider spectrum of parts to be produced without complicated and involved scheduling.

The second section of the thesis that presents the simulation analysis for buffer management, also aims at identifying the minimal buffer sizes before the constraint so as to keep lead times to a minimum.

Chapter 2

Theory of Constraint

One of the philosophies that propounded the relevance of aiming at global optimization is "Theory of Constraint". The basic premise of this theory is that it is logical to focus on the constraint or the bottleneck in an organization instead of attempting to manage everything [10]. The apparent simplicity in this concept belies the tremendous implication that it carries. Over the past decades, departmentalization of manufacturing activities has led to an isolation that has bred a quest to optimize locally. A classic example is placement of buffers before each and every machine in a production line to protect against statistical fluctuations. The logic holds only if each machine is viewed separately and hence a need to protect it from becoming idle is perceived and the need is satisfied by a buffer. Yet, when the line is considered in its entirety, it is often realized that not all machines have to be buffered because all machines do not operate at an equal production rate. Hence only the slowest machine has to be protected by buffering.

Secondly, the traditional tendency to increase the efficiency of each machine causes the shop floor to be overwhelmed by work in process inventory that was released to provide work for the machines. Theory of constraint realizes that the production rate cannot exceed that of the constraint and hence the material release should be coupled with the production rate of the constraint. This apparently simple concept is repeatedly ignored in most manufacturing concerns.

It will be pertinent to point out the similarities and differences between theory of constraint and JIT. Both of them are excellent examples of systems thinking. Both try to eliminate excess inventory from the production systems and hence improve lead times. The difference lies in the fact that JIT assumes that the constraint lies in the market and hence considers it unnecessary to buffer the machines at all [11]. Theory of constraint, on the contrary presumes that the constraint can either lie in the market or in the system. In case the constraint is in the system, it has to be buffered and secondly, the constraint can move in and out of the system and should be managed accordingly.

TOC Performance measures

TOC defines three operating measures. These measures are throughput, inventory and operating expenses. Throughput is the rate at which the organization generates money through sales. It must be understood that the reference here is to sales and hence throughput is counted not just when the parts come out of the system but when they are sold.

Inventory is the monetary value of the material introduced in the system for final conversion into finished goods. Inventory includes value of raw material, work in process, finished goods, and auxiliary production supplies.

Operating expenses are the money that is spent in converting the inventory into throughput. This includes the salaries, wages, advertisement, rework etc.

Chapter 3

Framework for Batch Sizing and Buffering

One of the classical problems facing the shop floor manager in any concern is that of lot sizing, scheduling and buffering of the machines in a production line. An extensive array of existing algorithms to solve this reflect the variety of approaches and objectives perceived by the researchers. The plethora of solutions and perspectives can be attributed to the vulnerability of this problem to mathematical modeling. The solutions that emerged in this quest were accompanied by simplification and modification of the problem. Notably, most of these OR based approaches considered a given shop floor configuration and attempted to optimize some local performance measures.

Since flexibility is the issue that has shot into prominence in the recent years, it was considered pertinent to approach this issue more analytically. In this thesis, the suggested framework aims at analyzing the factors that effect flexibility and attempts to quantify it. The objective was to to improve the flexibility of a shop floor by decreasing the response time.

The framework suggested in the thesis is driven by the demand and the product mix for a given time horizon. The input for the model is the demand for each product, the processing time on each machine and setup times. The model outputs the suggested process batch sizes. It must be understood that the output is not a schedule in terms of lot sizes and product sequence to be produced for a given time horizon but instead, a list of suggested process batch size which can satisfy the demand for the given time horizon and reduce the response time. The input data concerning the product mix and the demand are considered to be *estimates* for future demands and are thus used as guidelines for establishing the batch sizes. Here the approach differs from other classical techniques which generally consider the given weekly or daily demand as unalterable and generate a schedule for satisfying that particular demand. The following sections describe the significance of the process and transfer batch sizes.

Transfer and process batch

Process batch is defined as being composed of units that undergo processing on a machine for one setup. Conversely, it the number of units after which the setup of a machine is changed for next batch. Transfer batch is defined as the number of units that are transferred between machines.

Traditionally, the process batches were determined using the economic order quantities which sought to seek a balance between the setup costs and holding costs. The objective was to keep the costs to a minimum possible and this approach was, notably, very localized and myopic. Transfer batches generally equaled the process batches and as a batch finished processing on one machine, it was moved to the succeeding machine.

The effect of these two types of batches on the average inventory and flow times has been demonstrated in the following two examples.

Consider the manufacturing of a part that requires five processing steps. Let the demand for this part be 1000 units. It is possible to produce the entire demand in one lot of 1000. Also, the entire lot can be processed at one machine before being moved to the next machine. Hence in this case both the process and transfer batch equal 1000 units. The processing time for each unit on each machine and the setup time for each machine is as follows.

Step #	Processing time minutes	Setup time minutes
1	2	80
2	3	90
3	2	40
4	5	50
5	4	120

For such an operation, which incidentally is very common in many shop floors, the total time to complete the order will be as follows:

Time to complete the order = 1000 (2+3+2+5+4) + (80+90+40+50+120)

=16380 minutes

Now consider the case where, instead of transferring the complete lot from one machine to the succeeding machine, a single unit is transferred and the machines are blocked by the succeeding machines. Blocking means that if the succeeding machine is still processing and its buffer is full, the preceding machine stops processing. Transfer batch here is one unit while the process batch is of 1000 units. Now if no buffers are allowed before the machine, the total time to process this lot will now be as follows:

Time to complete the order = 999 (5) + (2+3+2+5+4) + (80+90+40+50+120)

= 5391 minutes

The reason for this significant reduction in lot processing time is that parts do not spend a lot of time waiting for processing. The processing approximates an assembly line where the throughput rate is dictated by the slowest machine which in this case is machine 4. Hence the parts are produced at the rate of 1 every five minutes. The first part, after introduction in the system, will exit after (2+3+2+5+4) + (80 +90+40+50+120) = 386 minutes. Each succeeding part will exit after a regular interval of 5 minutes. It is also assumed here that transfer time is negligible.

Now consider the case where the process batch size is cut down to half, i.e. 500 units. The total time to process a lot of 500 units will now be as follows:

$$\text{Time to process 500 units} = 499\ (5) + (2+3+2+5+4) + (80+90+40+50+120)$$
$$= 2891 \text{ minutes}$$

The time to process the entire demand is 5782 minutes. Hence by decreasing the process batch size, the average time to process a batch goes down but if setup times are significant, the total time to process the lot increases. The advantages of smaller batches is the greater flexibility in scheduling the floor and faster response time. The smaller those process batches are, the greater the flexibility is going to be. As the process batch sizes decrease, the setups become more significant hence a tradeoff has to be established in terms of the size. This has been elucidated in the case of print room operations at an electronics firm in the succeeding sections.

Hence it is obvious that both types of batches impact the flow times and inventory in a system. Even for the same process batches, i.e. keeping the setup costs constant, it is possible to reduce the flow times and inventory by cutting the transfer batch size to a single unit. The counter-argument can be the increase in the material handling costs due to more frequent transfers but the fact is that most of the present day material handling systems are automated and they do not display a significant increase in operating expense for higher volumes.

It thus becomes apparent that, on a plant floor the efforts should be directed towards cutting the batch sizes.

Existing Approaches

Recent developments in the theory of computational complexity indicates that except for a limited number of special situations, it is extremely unlikely that optimal solutions to scheduling problems will ever be possible except by partial enumerative methods such as branch and bound [27]. Hence most of the approaches are sub-optimal methods to provide acceptable near optimal solutions and it may be argued that this is really what the industry is primarily concerned with anyway [12]. The most popular objective function of these algorithms has been minimizing the maximum tardiness. Branch and bound methods have been effectively used in algorithms developed by Mcmahon and Florian to solve such models [17]. Larson and Dessouky also developed a heuristic for single machine problem to minimize maximum lateness using branch and bound algorithm [14]. Branch and bound algorithm was also used by Sen and Gupta to solve a bicriterion scheduling problem [23]. Potts and Wassenhove solved a similar model to minimize the total weighted tardiness [20]. Chen and Bulfin used the classical assignment model to solve the scheduling problem with multicriteria objective [5]. Another approach used to solve this problem involved response surface methodology by Larson and Devor [15]. Other approaches included minimizing flow times and maximum penalty [29] or minimizing number of late jobs [24]. Some of the heuristics were designed for the sole purpose of solving very large sized problems in a short time [4]. The approach defined in this thesis aims at reduced lot processing time with an intent to improve flexibility. The approach uses integer programming models and uses the expected demand for the entire product line to generate the lot sizing decisions.

Suggested Approach

As described in previous sections, for a given line it is desirable to have small process and transfer batches. The reason is to improve the lead time performance of the system. An ideal case, obviously is the situation where the transfer and the process batch equal one unit. In most of the practical situations, it is not difficult to have transfer batch of single unit size. The currently available material handling systems can afford to operate for such small batch sizes. The problem arises for the process batches when setups are involved. The setup costs and the time involved in setting up increases. Thus a feasible process batch size has to be developed which can attain a tradeoff between small batch sizes and the setup time and cost.

It is also understood that in most of the production lines, there exists a production constraint that gates the throughput of the system. This is the machine that is most heavily loaded. In the approach presented in this thesis, the constraint machine is used for generating the "optimal" batch sizes. The constraint machine will define the smallest batch sizes that can be used. The approach will use the available time on the constraint machine and the processing time and setup time for different products to determine the batch size for each different product in such a manner so as to minimize the average lot processing time. The batch sizes that are thus generated for the constraint machine can be used for other non-constraint machines. The reason is that if lots from the non-constraint machines are larger than the size defined for the constraint machine, there will be an accumulation of parts before the constraint machine. On the other hand if the lots from the non-constraint machines are smaller than those defined by the constraint machine, then the non-constraint machines will be setup for a greater number of time without any benefit of lead time reduction. It must also be understood that minimal lead time will be realized only if transfer batches are reduced to single unit.

Algorithm

The first step is the demand estimation for each product. The expected daily, weekly or monthly demand for each product is estimated. These demands may appear as:

Product	Day				
	1	2	3	4	5
A	-	100	200	-	100
B	50	50	300	100	50
C	-	-	50	200	-
D	-	50	-	-	100
E	50	-	100	50	-

These daily demands for five products over a weeks time are then consolidated so as to yield the average weekly demand. These demands may appear as:

Product	Demand
A	400
B	550
C	250
D	150
E	200

These weekly demands are then used as the input for the model. The next step is determining of the work load on each machine. The work load is the ratio of the amount of processing time needed to the total time available:

workload ratio for machine n = W_n / A_n

where $W_n = d_i \times p_i + a^n_i$ (d_i is demand for product i and p_i is processing time for product i on machine n and n=1,2..k and i=1,2..m)

A_n = total available time on machine n

k = number of machines

m = number of different products

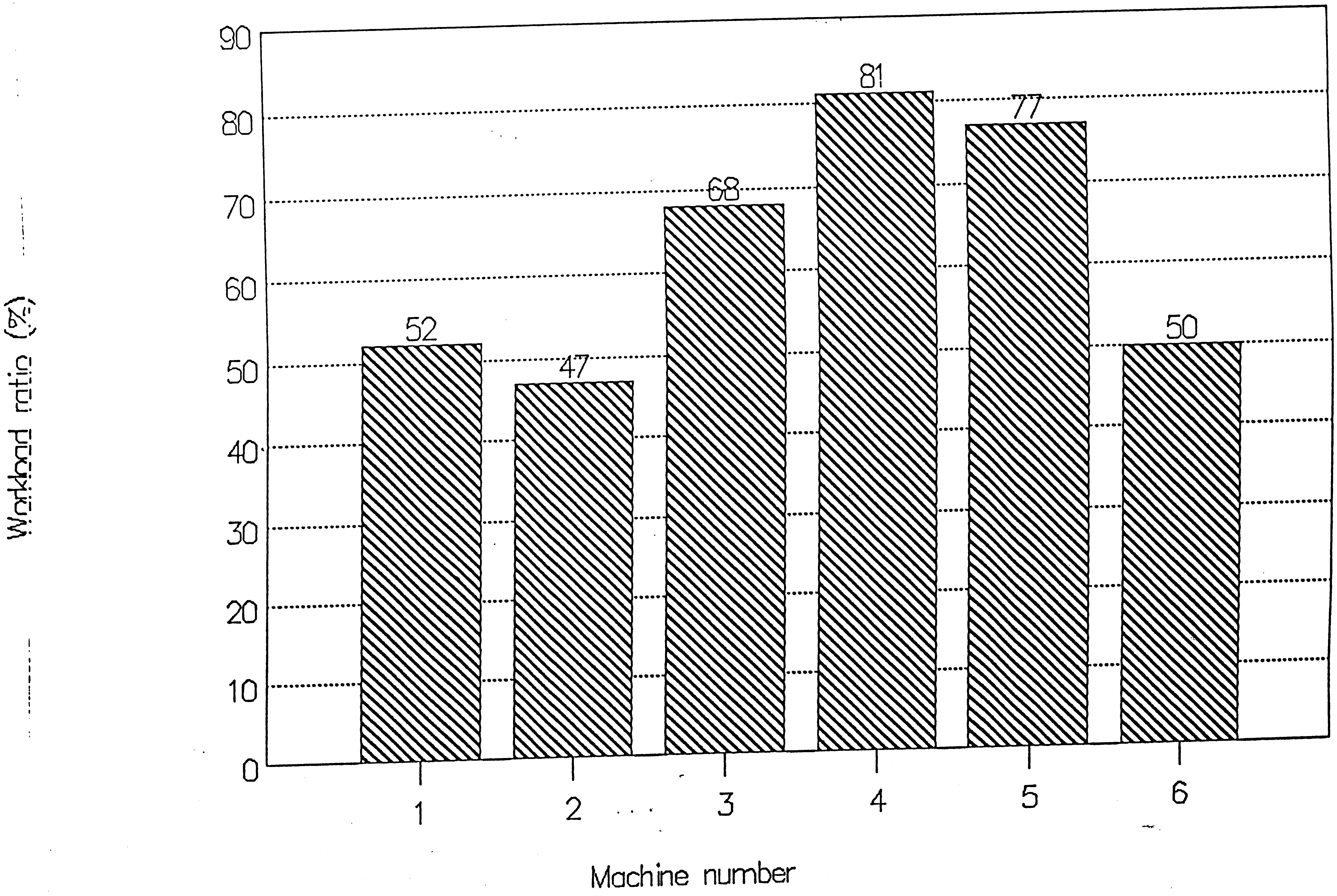
a^n_i = setup time for product i on machine n

Constraint Identification

The next step is identification of the *constraint* machine or the critical machine on which the process batch sizing decisions are to be made.

Consider a hypothetical plot of workload ratio versus machine numbers as shown in figure 15. In the given case it can be observed that the production system can produce the given demand with the available

Workload profile



capacity i.e. in the strictest sense, there is no *capacity constraint* (assuming that each product is setup only once). Yet machine number 4 has the maximum workload and in case there is some loss of production time at this machine, the throughput might be effected. Hence machine 4 can be considered as a constraint machine which will dictate the batch sizing decisions. It will be shown that only this machine has to be considered for such a decisions and the remaining machines should be loaded with batches determined at this machine. The approach to solve the problem if there is a *capacity constraint* in the system i.e. the workload ratio for a machine is greater than one, has been discussed in the later sections.

In the given example, it has thus been observed that the "constraint" machine has some extra time. Since the total processing time will be constant, the remaining time can be used for additional setups. The reason why additional setups are desirable is because then more and consequently, smaller lots can be produced. This in fact is what the model drives at. The model tries to break up the demand quantity into smaller lots by trying to utilize the extra time for setups and thus the average time to process any given lot will be decreased.

The following model will generate the list of number of setups for each different product type which will result in minimized response time.

It has been assumed that each product type will have a same process batch size.

The Model

Consider the constraint machine on a shop floor. Let x_i be the number of times product i is setup for the given time horizon. The process batch size for product i will then be d_i/x_i where d_i is the demand for product i. The average batch size for the entire product mix at the constraint machine will be:

$$1/m \sum (d_i/x_i)$$

The time taken for any lot to be completely processed at the constraint machine is directly proportional to this quantity. Now consider the case where the transfer batch equals one unit. In that case the time to process a lot through all the machines will be gated by the batch size defined for the constraint machine. The reason is that the lot size at the non-constraint machines can be equal to the lot size at the constraint machine. Hence the average time that a lot will take to be introduced in the system and be processed will depend upon this quantity.

The upper bound for the maximum time to process the entire lot through the system will be:

$$n\,(1/m \sum (d_i/x_i))$$

Hence the objective function will be:

$$\text{minimize} \quad n\,(1/m \sum (d_i/x_i)) \qquad \text{for } i=1,2,\ldots.m$$

The constraint equations will be:

$$\sum x_i a_i + \sum d_i t_p < K$$

$$x_i > 0 \text{ and } x_i \text{ is an integer}$$

where K = total available time

t_p = processing time for each unit

The model has to be solved as an integer programming problem since x_i can only have integer values. The solution will yield different values for x_i which will indicate the number of setups for product i thus if x_2 = 5, then the average lot size for product B will be 110 and this product will be setup 5 times during the given time interval. The trivial solution for the problem will be that all x_i are equal to 1 i.e. each product type will be setup only once and the entire demand will be produced in one lot.

The problem can be simplified by solving it as a linear programming model. The values of x_i can be then rounded off to the nearest integer and the lots can be balanced again. This can be done by relaxing the assumption that each product type has same batch size for different setup eg. a product may have a smaller process batch in the last setup. At this point it must again be stressed that unlike conventional operations research models, the proposed model attempts to establish guidelines for lot sizes instead of specifying the exact values.

Model with a capacity constraint

The second possible case in the configuration described above is when a capacity constraint exists. In such a case it will not be possible to produce the entire quantity demanded. Hence a decision has to be made concerning the product mix and the quantities that should be produced. The factors affecting this decision is the penalty of not identifying a particular product, the processing time for that product and setup times. A linear programming model similar to the one described above can be used to create such a model. Theory of Constraint defines control measures like throughput-dollar-days and inventory-dollar-days to quantify such cases.

Determination of transfer batch sizes

A common approach in most job shops or production lines for processing and transferring lots was to treat the transfer batch size and the process batch as same. The factor commonly attributed to this approach was the ease of tracking the batches through the system. Yet the transfer and process batch sizes effect the lead times and the need to rationalize the process of determining them is obvious.

In most of manufacturing facilities, some definite cost is associated with transferring each unit from one machine to another. Even if the exact unit cost is difficult to obtain, an estimate can be made by considering the cost that has to be incurred for additional traffic on the floor (eg. cost of adding additional AGV or carts on the floor). This cost will tend to restrict the transfer batches to a larger size.

On the contrary, smaller transfer batches will lead to smaller lead times and subsequently, lesser holding costs. These factors will tend to favor smaller transfer batches. Smaller process batches will also lead to smaller lead times but these batch sizes will be dictated by the sizes obtained for the constraint machines.

Hence the most practical method to determine the transfer batches is to consider the cost involved in installing a material handling system capable of handling smaller batches. The second important factor to consider is the traffic patterns created on the shop floor because of the higher number of transfers. Often high traffic levels, because of reduced transfer batch sizes, cause operational problems which might force the transfer batches to be increased in size. Thus determination of transfer batches is often a function of operational constraints and logistics on the shop floor and development of a mathematical model can be an extremely difficult or even irrelevant since quantification of tradeoffs into dollar value will be impossible.

Chapter 4

Buffer Management

The model suggested in the previous sections assumes deterministic processing and setup times and no machine breakdowns. Often, on a shop floor this is not so and statistical fluctuations are rampant which often cause disruption in production. This disruptive effect can be countered by two approaches. The first is to reduce the fluctuations [21]. Systems like JIT in fact aim at reducing the fluctuations which can often take years. This approach involves a thorough comprehension of each process and its rationalization, setup reduction and extensive worker training. This approach is certainly the logical method of operating the shop floor but it is also patent that most systems do have fluctuations which are often unavoidable and even generated outside the system [30]. In this case, it is necessary to devise a system that can protect itself or at least lessen the effect of fluctuations. Thus the second approach is by maintaining buffers at critical points in the system [7].

Theory of constraint is the ideal framework for designing these buffers. This theory can be used to determine the placement of buffers in a production system and devising an operating strategy for the given environment. The suggested approach by the theory is called the Drum-Buffer-Rope (DBR) approach [6] and [11]. This approach has been explained briefly in the following sections.

Drum, Buffer and Rope

The drum is production pace set by the constraint. This constraint can be a resource, raw material supply, management policy or market damned. Consider a production system where the slowest machine produces at the rate of 8000 parts/month. If the market demand is just 6000 part/month then the drum in this case is the consumption rate of the market. If the market demand is now 12000 parts/month, then the constraint and consequently, the drum shifts to slowest machine. The drum beat or the production rate is now limited to 8000 parts/month. The production schedule for the machines will now be a function of this drum beat.

Buffers are the protection time given to particular machines to decouple them from the fluctuations at the preceding machines. This protection time is provided by placing some parts before the machines i.e. these parts are planned to reach the machine some time before they are planned to be processed. The fluctuations and disruptions can be due to machine breakdown, fluctuations in setup times, fluctuations in processing times etc. Such protection for machines is seen on most shop floor as the buffers before them. The most significant deviation of theory of constraint from traditional approaches is observed here. Traditional approaches tend to provide protection for every machine by placing buffers before them. The premise being that fluctuations are present everywhere and thus each machine is susceptible to disruptions at the preceding machine. Theory of constraint, on the contrary, strives towards protecting only the constraint. The reason is that time lost at the constraint is irrecoverable hence it should be protected while the non-constraints, by the virtue of their faster processing rate, can cover up for lost time. Any protection or placement of buffers before them will add to the lot makespan time without effecting the throughput.

Rope is the mechanism that forces all the parts of a system to work according to the pace set by the drum. Thus it is a schedule that prevents the non-constraints from producing more than what can be handled by the constraint. The rope can be physical or logical. The most common example of physical rope is use of conveyer belts in auto assembly plants. The kanban cards are examples of logical ropes. The rope can also be built in the controlling software for the material handling system which will prevent a part from being sent to the succeeding machine if the succeeding machine has a full buffer. This strategy is also called blocking. The ropes, like buffers, are not spread all over the floor but are used only at the gating operations.

In the following section, the method to devise the buffer size and the ropes are discussed using simulation results.

Design of DBR System

A hypothetical line was simulated for different operating conditions to determine the buffer requirements for the constraint machine. Data from over 400 simulation runs was obtained and analyzed to determine the relationship between the constraint placement, machine failure rate and the performance parameters like throughput, time in system and average work in process inventory. The purpose of the simulation was to examine the trends in the performance parameters for different operating conditions which could then be used as a guideline for designing the buffers.

Simulation Environment

A production line with sixteen workstations was simulated. Each workstation had a buffer of infinite capacity before it. All the machines had the same processing time except one which took longer to process a part. The processing times were considered to be normally distributed around their means. The machines also had a downtime profiles. The downtime was simulated as being exponentially distributed. This was done to simulate the unavailability of machines due to setups and failures.

Each part was sequentially processed by the machines and the routing for all parts was fixed.

Operating Strategy

The part release to the system was coupled to the constraint machine. A part was introduced in the system only when a part finished processing at the constraint machine. In the terminology of theory of constraint, the drum beat here was controlled by the output rate at the constraint machine.

Experiment Setups

The following sections describe the variables and the performance parameters used to make the deductions.

Simulation Variables

The following parameters were varied in the simulation runs:

(1) Position of constraint machine.

(2) Processing time means for the non-constraint machines.

(3) The downtime profile of the machines.

(4) The number of parts initially placed in the constraint buffer.

(5) Random number seeds. (for multiple runs).

For different runs the placement of the constraint machine was changed. The positions used were 1st, 5th, 11th and 16th. This was done to analyse the impact of constraint position on the system's performance parameters.

The processing time for the constraint machine was taken as normally distributed with mean of 4 min/part and a standard deviation of 10% of mean. The processing time for the non-constraint machines was taken as normally distributed with mean of 2 min/part and 3 min/part. These two different values were used for different sets of runs to analyse the impact of the processing time differential between the constraint and the non-constraint on the system's performance parameters.

The downtimes were modeled using mean time between failures (MTBF) and mean time to repair (MTTR) values. The different MTBF values used were 240 min., 120 min., and 90 min. The MTTR value was kept constant at 10 min. The distributions used for both the MTBF and MTTR values was taken as exponential.

For different runs, the simulation was started with different number of parts in the constraint buffer. These numbers were 4, 6, 8, 10, 15 and 20. This initial constraint buffer size was varied to observe its effect on the throughput and the time in system for a part.

For better accuracy, runs for similar conditions were replicated using different random number streams. SLAM was used for the simulation purposes and it has 10 different random number streams. For the simulation under consideration, 4 different streams were used.

Performance Parameters

Following performance parameters were used for analysis.

(1) Throughput.

(2) Time in system.

(3) Average work in process inventory.

The throughput in terms of parts/day was used as a measure to determine the performance of the system.

Time in system was the average time taken by a single part to be completely processed by the system. This parameter was measured in minutes.

Average WIP was determined by counting all the parts in the system at regular intervals of time.

Simulation Run Lengths

For each condition, the simulation was run for 12 days where each day had 480 minutes. The results for the first day were discarded to exclude the warm-up period.

Simulation Runs Configurations

The tables 1 and 2 show the simulation run configuration. A total of 24 simulation sets were used as shown in the figure. The corresponding values of the variables are also shown. In each simulation set, the number of parts initially placed in the constraint buffer were varied from 4 to 20. A typical output for one set is shown in figure 3. In figure 3, row number 3 corresponds to the simulated results obtained when the initial number of parts in the constraint buffer were kept as 8. These figures for throughput , time in system, and work in process inventory were obtained by simulating the system for 12 days for four different random number streams and taking their average.

SET #	MTBF/MTTR (minutes)	Processing time for constraint (minutes)	processing time for non-constraint (minutes)	constraint machine #
1	240/10	4	3	1
2	120/10	4	3	1
3	90/10	4	3	1
4	240/10	4	3	5
5	120/10	4	3	5
6	90/10	4	3	5
7	240/10	4	3	11
8	120/10	4	3	11
9	90/10	4	3	11
10	240/10	4	3	16
11	120/10	4	3	16
12	90/10	4	3	16

Table 1

SET #	MTBF/MTTR (minutes)	Processing time for constraint (minutes)	processing time for non-constraint (minutes)	constraint machine #
13	240/10	4	2	1
14	120/10	4	2	1
15	90/10	4	2	1
16	240/10	4	2	5
17	120/10	4	2	5
18	90/10	4	2	5
19	240/10	4	2	11
20	120/10	4	2	11
21	90/10	4	2	11
22	240/10	4	2	16
23	120/10	4	2	16
24	90/10	4	2	16

Table 2

SET # 6

intial number of parts in constraint buffer	Throughput	Time in system	WIP
4	91	139.39	25
6	96	152.10	29
8	99	163.90	33
10	101	175.10	36
15	103	199.43	42
20	103	222.42	47

Table 3

Simulation Results

The following sections present the results of the simulation runs that were made. An analysis of results was done to identify definite trends in the performance parameters for different operating conditions.

Throughput

Variations in the throughput were analyzed for different constraint machine positions and downtime profile. Graphs 1, 2, and 3 show the daily throughput for the simulated system for different downtime profiles. The processing time for the constraint machine was kept at 4 minutes per part and for the non-constraint, it was 3 minutes. Graph 1 was generated from the data obtained from simulation set numbers 1, 4, 7, and 10. Graph 2 was generated from simulation set numbers 2, 5, 8, and 11 and graph 3 was generated from set numbers 3, 6, 9, and 12. For graph 1, the downtime profile was modeled as an exponential distribution with mean of 240 minutes. The repair times were also exponentially distributed with a mean of 10 minutes. For graph 2, the MTTR was reduced to 120 minutes while in graph 3, it was kept at 90 minutes. The MTTR was kept constant for all runs at 10 minutes. For each graph, the throughput was plotted for different constraint machine positions. The constraint machine positions used were 1st, 5th, 11th, and 16th. The legend for each graph indicates the defining pattern for the bars for different constraint positions. The x axis indicates the initial number of parts that were put in the constraint buffer. The Y axis shows the throughput as parts produced per day.

Among the three different graphs, the throughput was found to be very sensitive to the constraint position for small initial starting constraint buffer sizes. The throughput was expectedly unaffected by the initial constraint buffer size when the constraint machine was at the very beginning of the line. In graphs 1, 2, and 3, the throughput for configurations having first machine as constraint was constant along the X axis (for different value of initial constraint buffer size). It was observed that as the constraint machine position shifted down the line, the throughput got affected more profoundly for different initial constraint buffer sizes. In graph 1, when the constraint machine was at the end of the line (position 16), the maximum throughput (theoretical throughput calculated by estimating the downtime) was obtained when the initial constraint buffer size was kept at 15 parts. The throughput remained more or less constant for higher values of initial constraint buffer sizes. On the other hand, when the constraint buffer was closer to the beginning of the line, maximum throughputs were realized when the initial constraint buffer size was kept at 6. The observed trend was that the initial constraint buffer size requirements went up when the constraint machine position moved downwards to the end of the line. For graph 1, the throughputs for all the different constraint machines positions stabilized for initial constraint buffer sizes of around 15 or in other words, when the initial number of parts in the constraint buffer were kept at around 15, the constraint machine position became insignificant.

The analysis across graphs 1, 2, and 3 shows the effect of downtimes on throughput for different initial constraint buffer sizes. As the downtime profiles deteriorated (MTBF decreased), the system got more sensitive to the initial constraint buffer sizes. As seen in graph 2, for constraint position number 5, the maximum throughput is now obtained for initial constraint buffer size of around 10 (compared with a figure of 6 for MTBF of 240). When the constraint shifts to the last machine, the throughput stabilizes at initial constraint buffer size of 50 (not shown in graph). A similar trend is observed in graph 3 where the MTBF is 90 minutes. The throughput deteriorates rapidly as the constraint machine shifts down the line. It can be seen from graph 3 that the difference in throughput between the case when the constraint machine is at the beginning of the line and the case when it is at the end of the line is 30% while in graph 1 (MTBF 240 minutes) it is 9%. Thus the initial constraint buffer size requirement goes up significantly when the breakdown profile worsens.

Another important observation was the values at which the throughput stabilized for different cases. The stabilization state was determined by plotting the state of the constraint buffer at regular time intervals. This plotting was done for different initial constraint buffer sizes as shown in graphs 13 and graph 14. Steady state was identified when the plot showed a variation around a constant value as seen in graph 13. For cases where the initial constraint buffer size was insufficient, the graph of constraint buffer level showed a decline. The value of throughput was recorded when the plot of constraint buffer stabilized. It was observed that for high MTBF values, the stabilized values of throughput for different constraint machine positions were same but as the downtimes deteriorated, these stabilized throughput values tended to differ. In fact, in some cases the throughput actually increased when the constraint machine was shifted down the line. In graph 2, the stabilized throughput value for constraint machine position 1 was 110 part/day while the stabilized value for constraint machine position 5 was 113 part/day.

Throughput vs. processing time differential

Graphs 4, 5, and 6 were generated with similar operating conditions as graphs 1, 2, and 3 but the processing time for the non-constraint machine was changed to 2min/part from 3 min/part. The simulation sets for graph 4 were 13, 16, 19, and 22. Simulation sets for graph 5 were 14, 17, 20, and 23 and for graph 6, they were 15, 18, 21, and 24. The trend in the observed values for the stabilized throughput was similar to that observed in the first three graphs but in this case, the throughput stabilized for lower values of initial constraint buffer sizes. In graph 5 it can be seen that when the constraint machine position was 5 (MTBF 120 min.), the throughput stabilized at a value of 8 for initial constraint buffer position as compared to a value of 10 in graph 2 when the processing time for the non-constraint machine was 3 minutes/part. This trend was present for all three different values of MTBF in graphs 4, 5, and 6. Thus for the same processing rate for the constraint machine, a larger differential in the processing time between the constraint and non-constraint machine led to a lower initial constraint buffer requirement for cases when the constraint machine was moved down the line. The reason attributed to this is that a larger differential helped in smoothing out the perturbations in the machines before the constraint machines which incidentally are the machines that effect the throughput

Time in system and WIP

The next performance measures that were analyzed were the time in system and the average work in process inventories. Since the time in system was a direct function of the WIP, the trends observed in both of them were identical and hence only one set, time in system, has been discussed. Time in system is the average time a part spends in the system both waiting and processing.

It was considered more pertinent to compare the values for time in system for a given value of initial constraint buffer size over different operating conditions like the MTBF values or the difference in processing rates. For a smaller differential in the processing rates (3 min. and 4 min.), the time in system tended to be higher for cases where the constraint machine was at the beginning of the system. The difference in time in system for case where the machine was at the beginning and when it was at the end of the line increased when the breakdown profile worsened. This can be seen in graph 7, 8, and 9. It was also observed that for a given condition e.g. graph 8 (MTBF 120 min and processing rates of 3min/part and 4 min/part), the differential between the case where the machine was at the beginning and when it was at the end was constant for different value of initial constraint buffer sizes.

It was also observed that when the differential in the processing time increased, i.e. the non-constraint machines speeded up, the difference in the time in system for case where the constraint machine was at the beginning and when it was at the end, became negative i.e. the time in system for the first case became less than that in second case. This difference again became positive when the breakdown profile deteriorated. These trends can be observed in graphs 10, 11, and 12. Thus the time in system for cases where the constraint machine was at the end of the line tended to show a comparative decrease when the breakdown profile worsened. Stated alternatively, the time in system for case where the constraint machine was at the end showed a lesser value than the case when the constraint machine was at the beginning if the breakdown profile deteriorated. This decrease was more prominent for a smaller differential in the processing time i.e. when the non-constraint machines were comparatively slower.

Conclusions and Deductions

The simulation studies were done with an intent to develop models or guidelines for designing buffer systems on the floor. It was realized that factors which strongly influenced the design of buffers were the following:

(1) Processing times for the constraint and non-constraint machines.

(2) Breakdown profiles (MTBF and MTTR).

(3) Position of the constraint machine in the line.

(4) Operating strategies on the shop floor.

(5) Number of machines on the floor.

(6) Part routing and variety.

It was also realized that the vast number of interacting factors, like those mentioned above, and their uniqueness for each manufacturing facility will make development of a generic mathematical model, a very difficult task. Instead it was considered pertinent to identify the trends and guidelines which could be used to design the buffers for a given shop floor configuration.

Buffer sizing for the constraint machine

As explained in earlier chapters, the constraint machines need a buffer protection to decouple them from the perturbations that occur in the machines before them. Non-constraint machines, on the other hand do not need buffers before them since generally the differential in their processing rates and the processing rate of the constraint machine is sufficient to compensate for operating inconsistencies of the system.

It is important to correctly size the constraint buffer because if the constraint is overprotected i.e. the constraint buffer is too large, then the average time to process a lot or a part goes up. On the other hand, an underprotected constraint will lead to a loss in throughput. Thus an attempt was made to identify the "optimal" buffer size which will afford maximum throughput without increasing the time in system too much.

First step involves estimating the maximum theoretical output.

This is done by calculating the total available time on the machine and dividing it by the average processing time for a part.

$$\text{max throughput} = \frac{\text{hours in shift} * 60 - (\text{shift time})/\text{MTBF}*\text{MTTR}}{\text{processing time}}$$

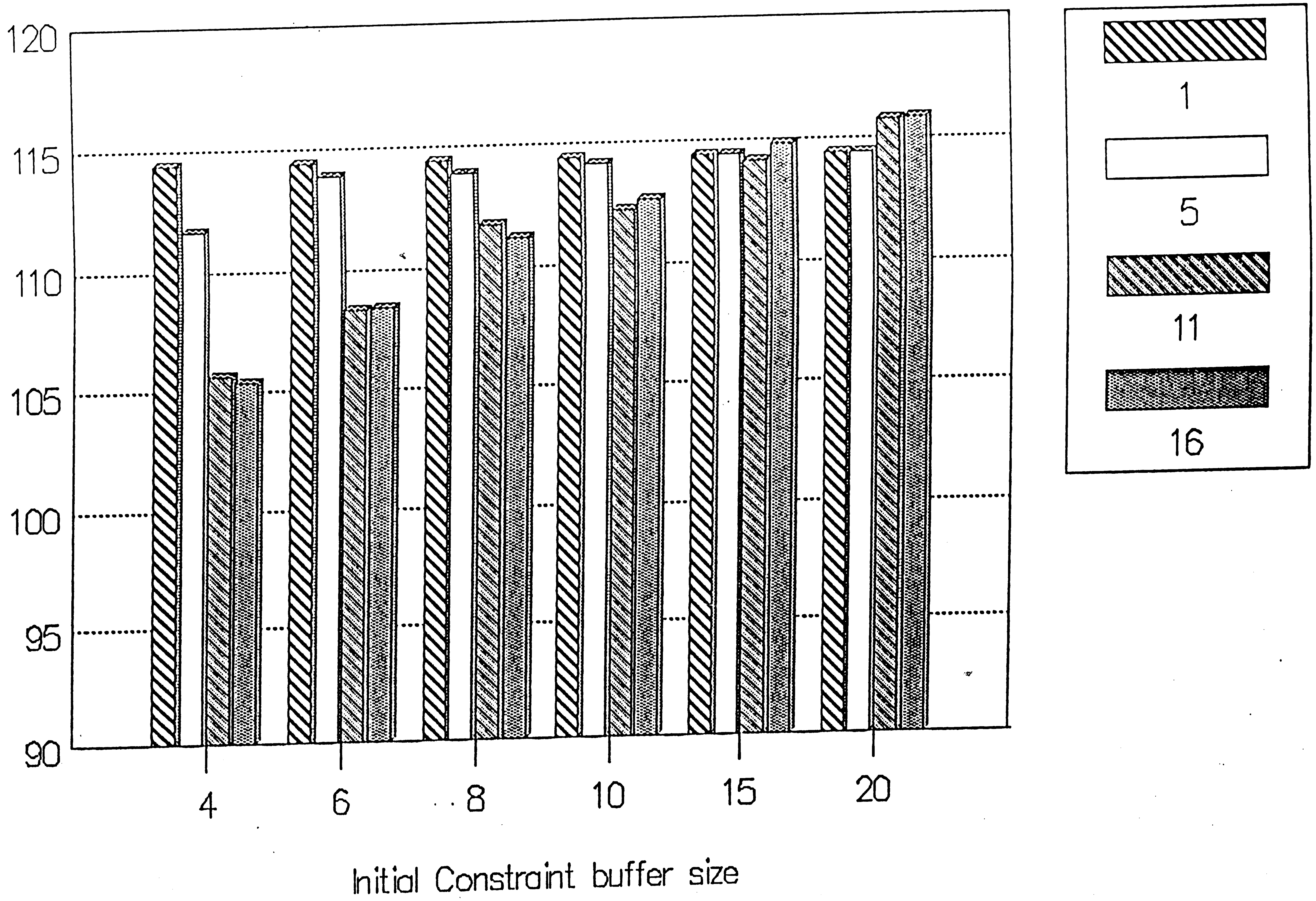
This figure indicates the maximum possible throughput that can be realized if the constraint buffer is adequately protected.

Next step is identification of the required constraint buffer size. This can be done by starting the production (or a simulation of it) with an arbitrary number of parts in the constraint machine buffer. The simulation runs performed were done in a similar manner whereby different initial constraint buffer sizes were selected. The constraint machine buffer is then monitored at regular intervals of time. The number of parts in the buffer are plotted as a function of time. An example of such a plot is given in figure 13. If the plot shows a steady depletion, the constraint buffer can be assumed to be under-protected. Thus the buffer size should be increased. If the plot looks like the one shown in figure 13, it indicates that the buffer is sufficiently protected. These two plots can thus be taken as the upper and the lower bounds for defining the buffer size. The buffer can be downsized from the size shown in figure 13 till the constraint machine starts getting starved. In fact, it might be feasible to accept occasional starving of the constraint machine if the decrease in the time in system and subsequently, the decrease in lot processing time has sufficient payback. This can be done by observing the frequency at which the constraint machine buffer depletes to zero.

Throughput vs. Constraint position

MTBF 240 MTTR 10 (3/4)

Graph 1

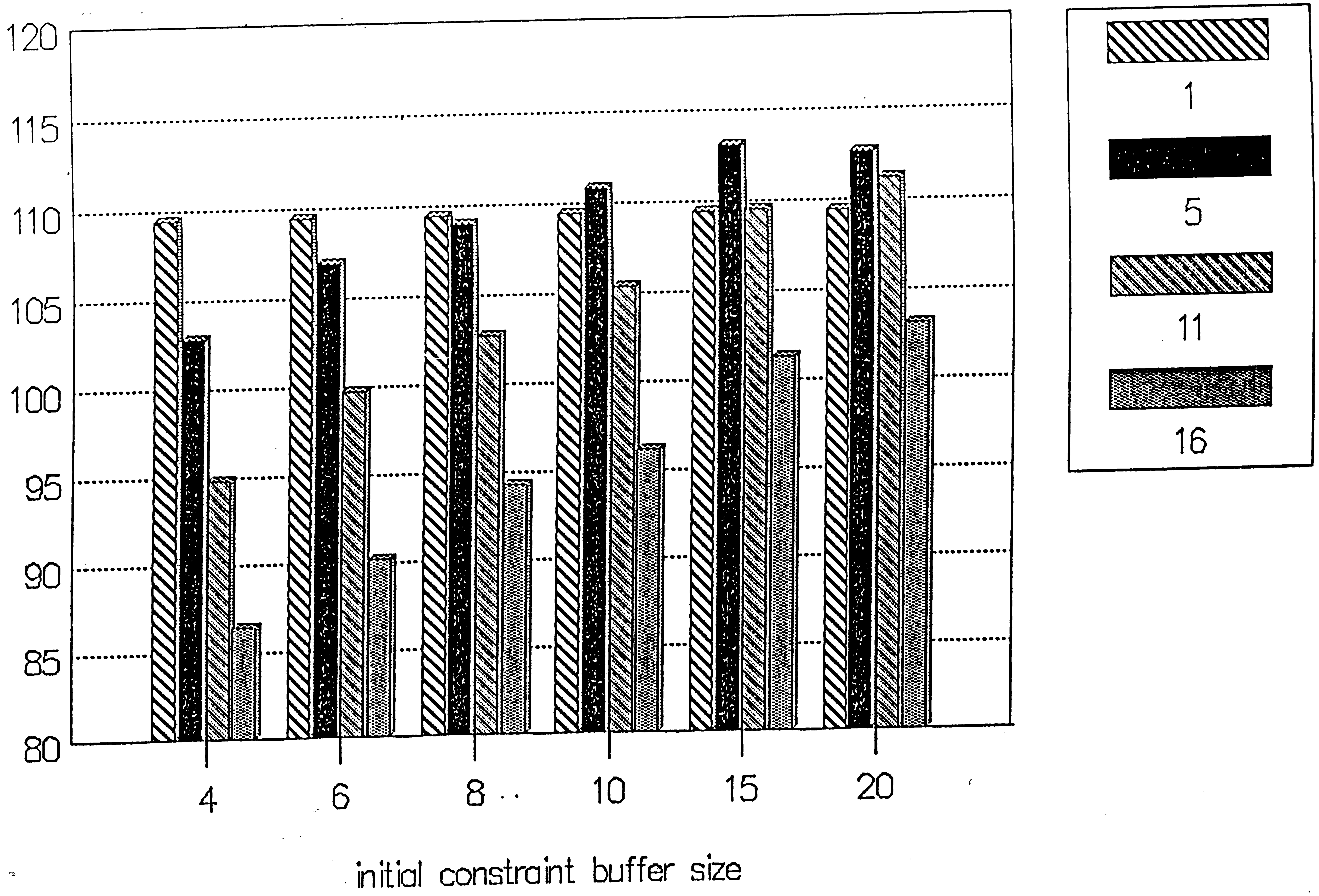


✓

Throughput vs. Constraint position

MTBF 120 MTTR 10 (3/4)

Graph 2

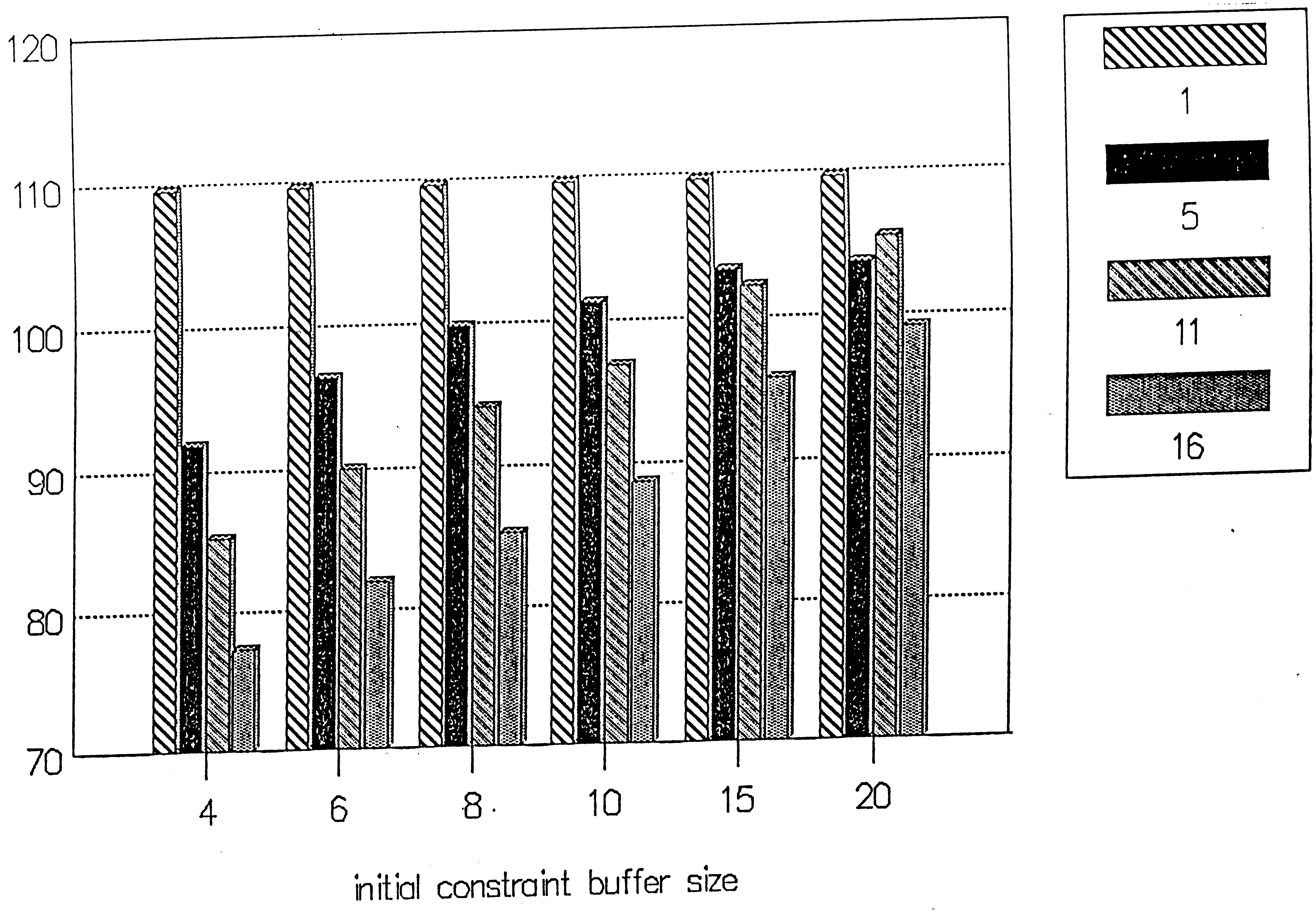


3

Throughput vs. Constraint position

MTBF 90 MTTR 10 (3/4)

Graph 3

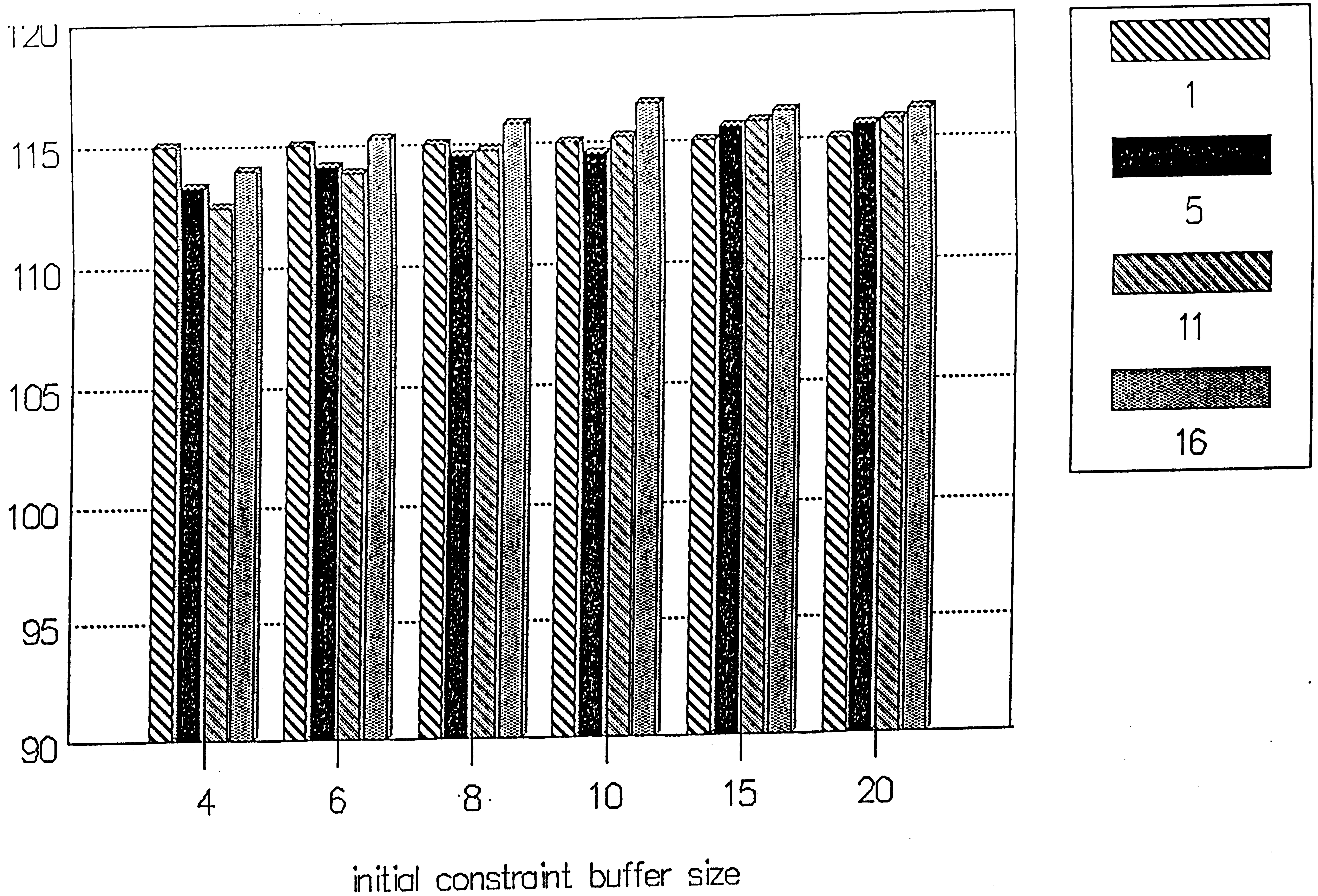


&

Throughput vs. Constraint position

MTBF 240 MTTR 10 (2/4)

Graph 4

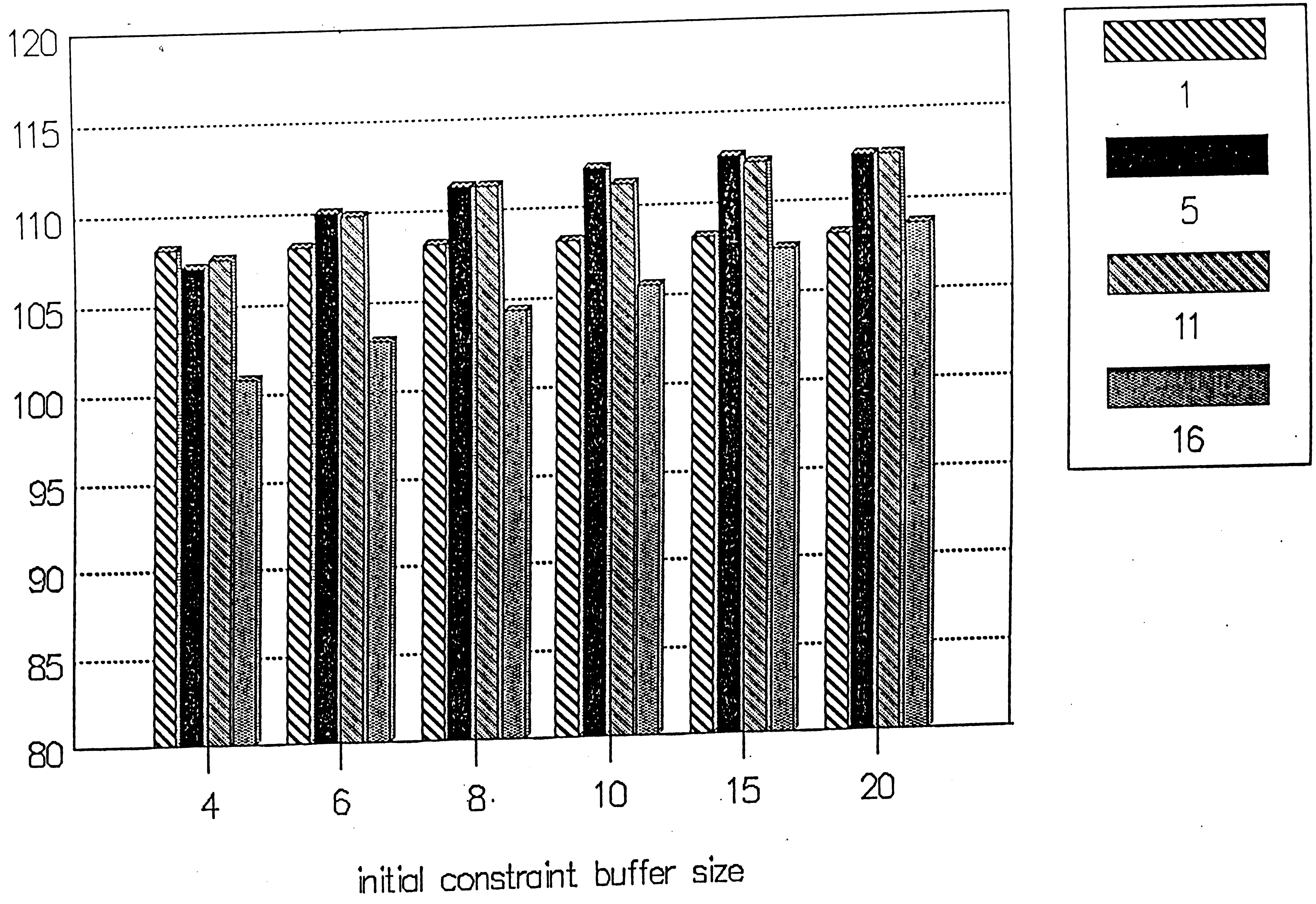


6

Throughput vs. Constraint position

MTBF 120 MTTR 10 (2/4)

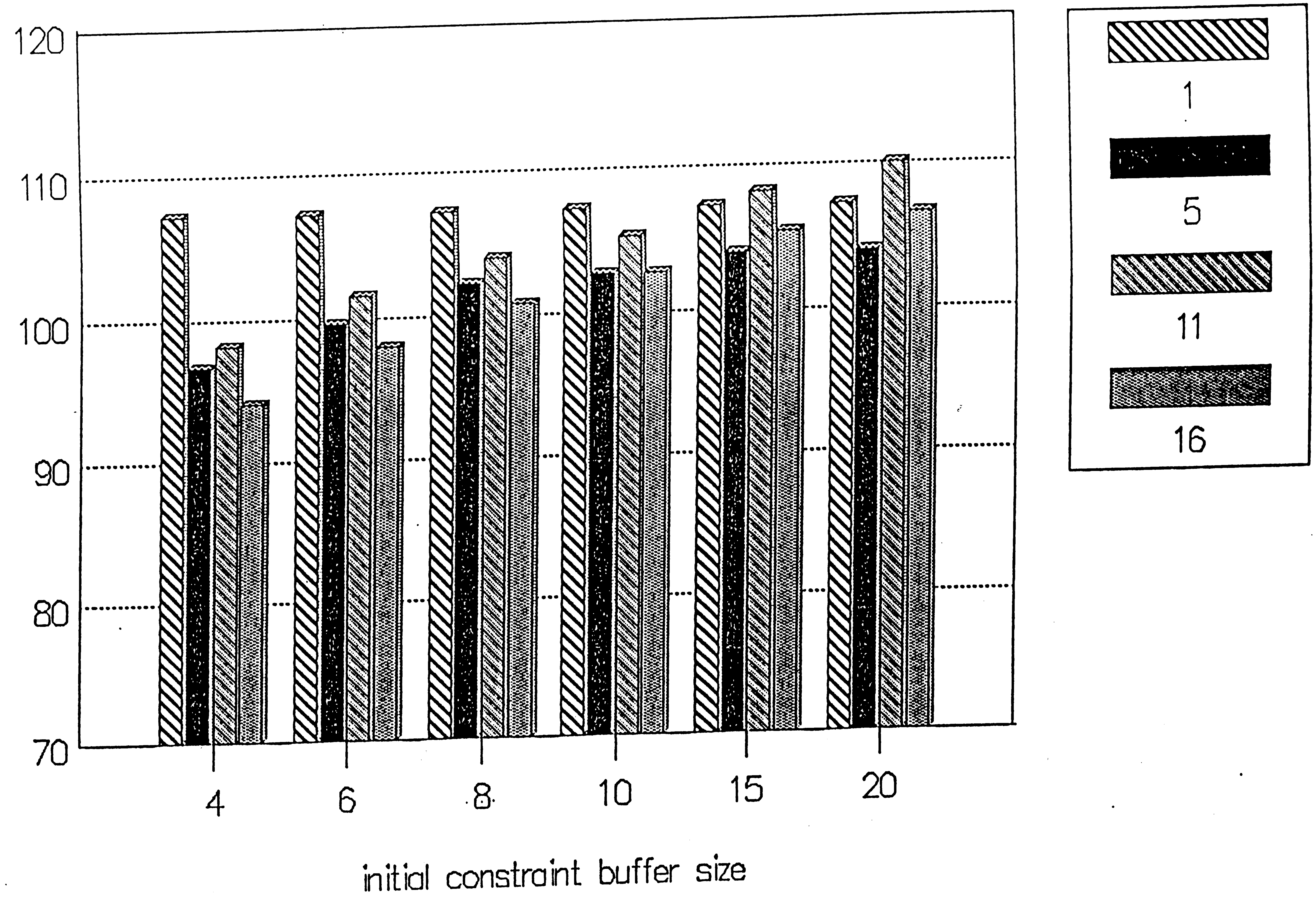
Graph 5



Throughput vs. Constraint position

MTBF 90 MTTR 10 (2/4)

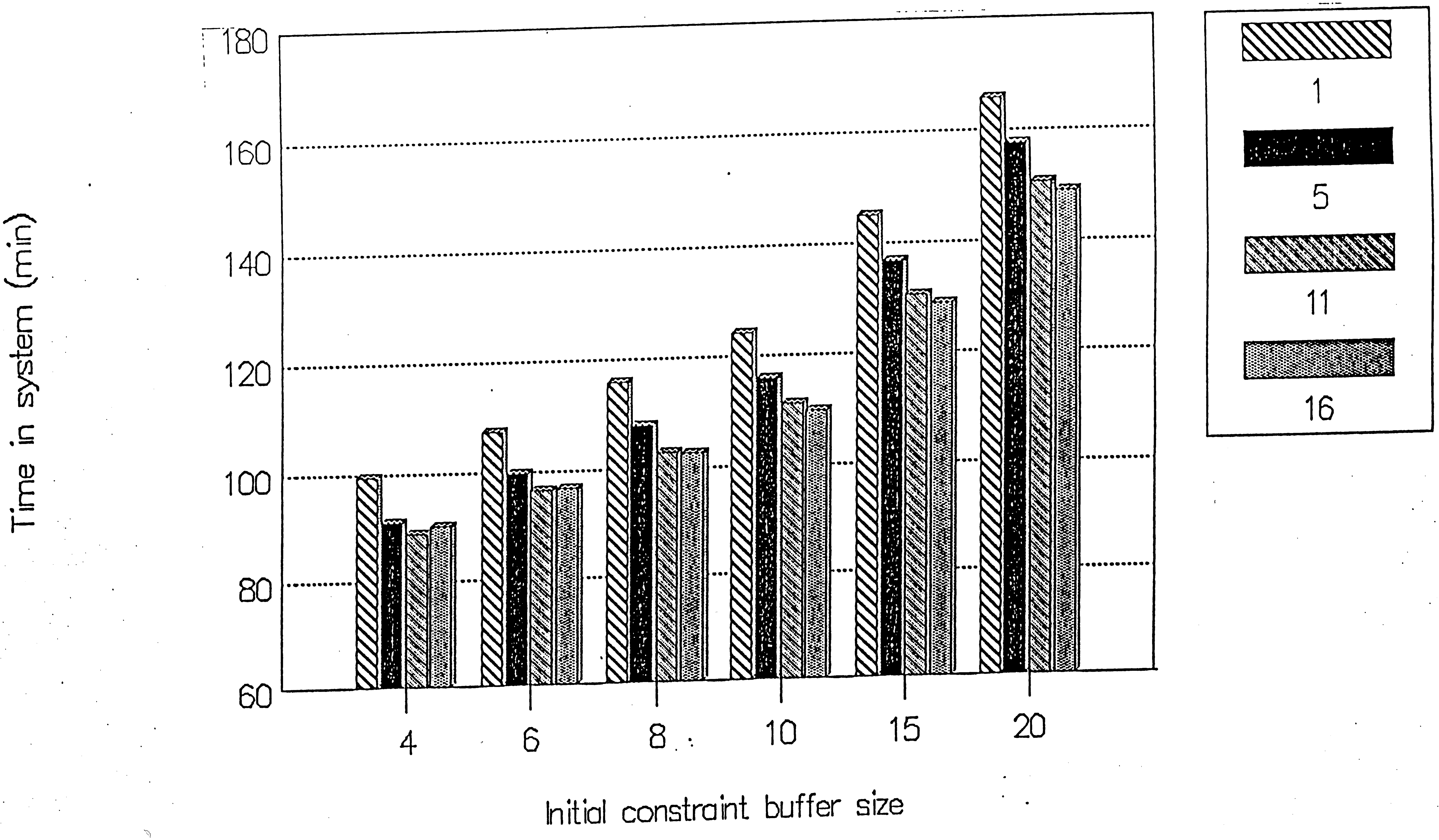
Graph 6



Time in system vs. Constraint position

MTBF 240 MTTR 10 (3/4)

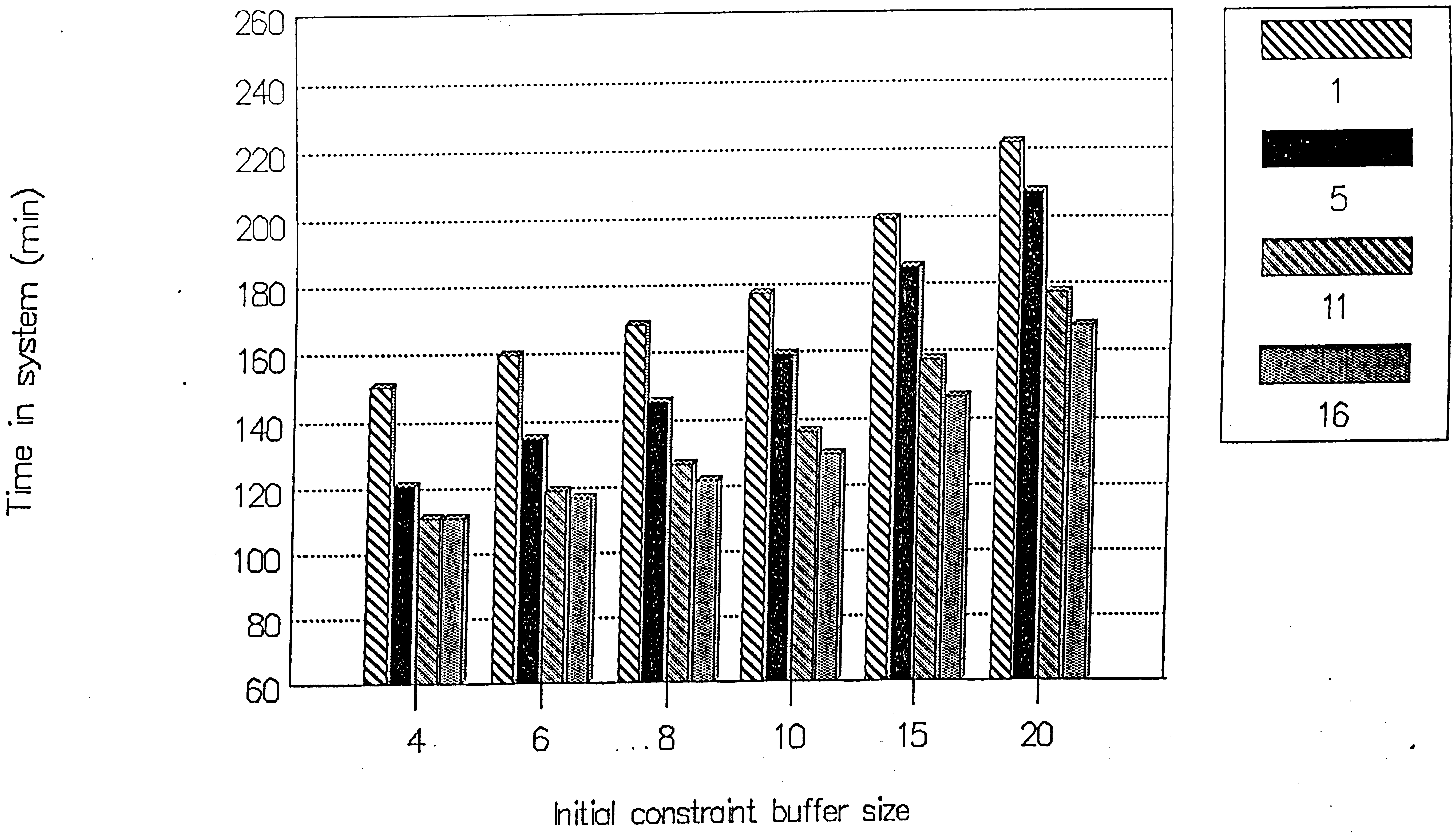
Graph 7



Time in system vs. Constraint position

MTBF 120 MTTR 10 (3/4)

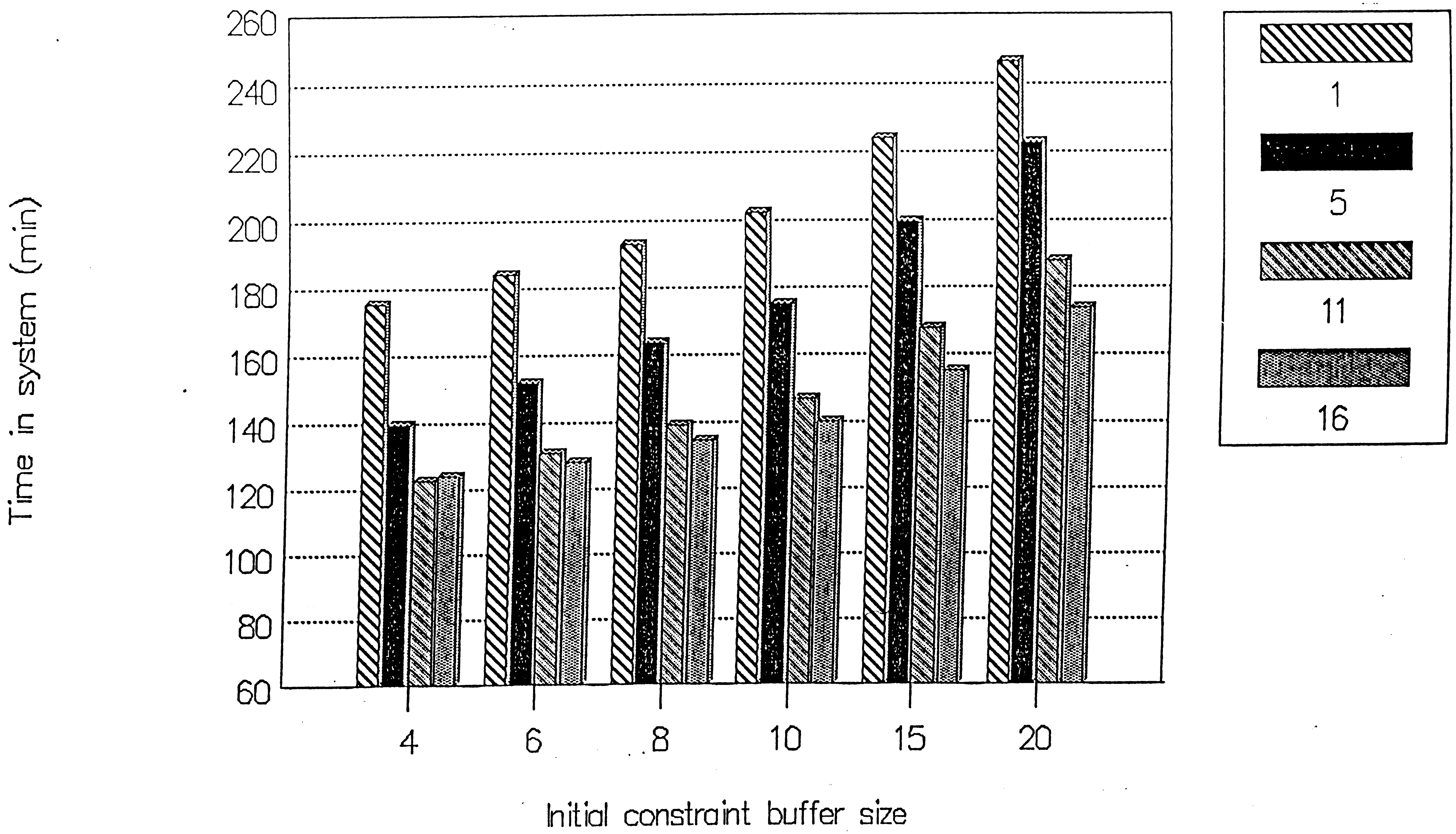
Graph 8



Time in system vs. Constraint position

MTBF 90 MTTR 10 (3/4)

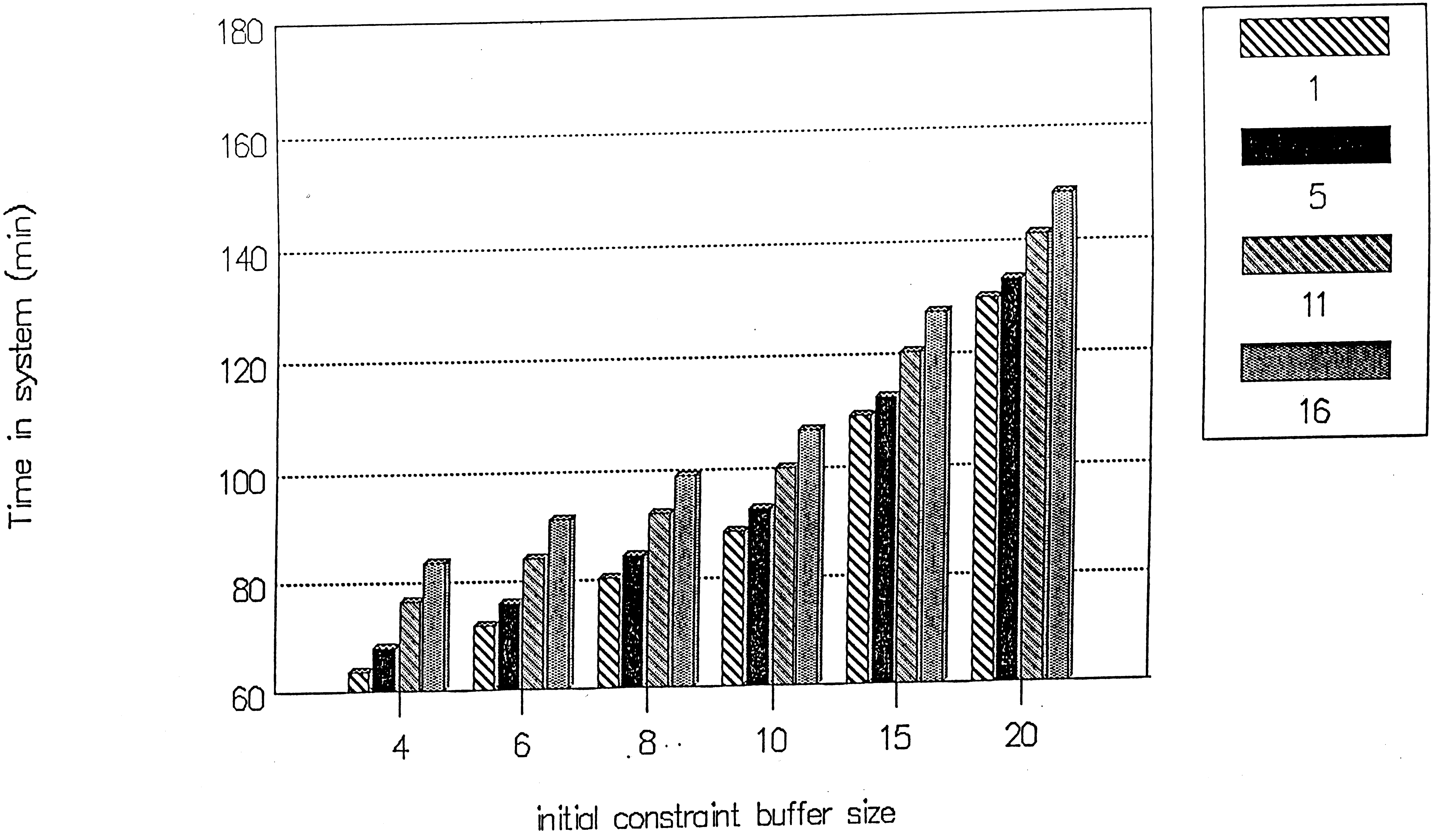
Graph 9



Time in system vs. Constraint position

MTBF 240 MTTR 10 (2/4)

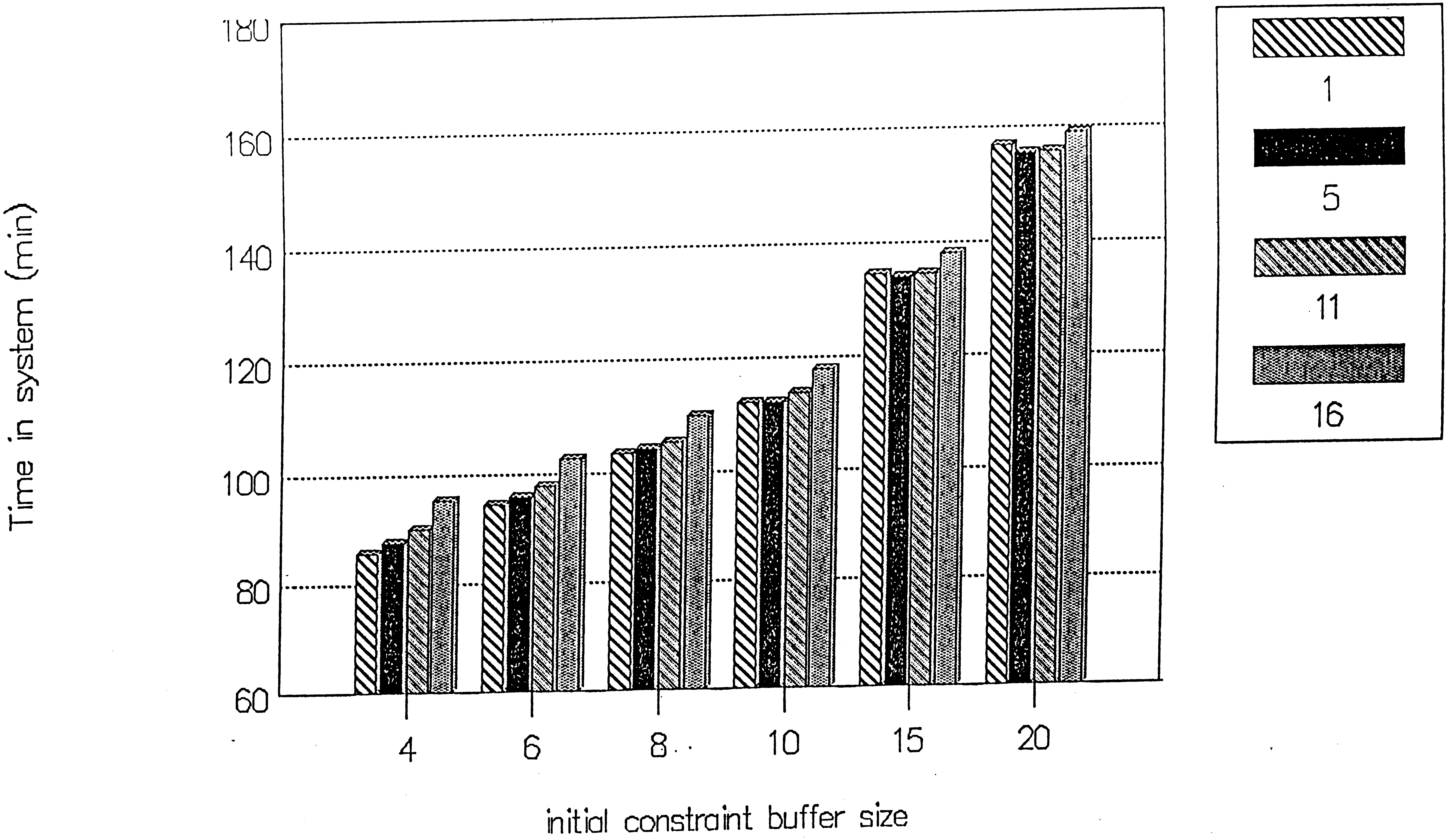
Graph 10



Time in system vs. Constraint position

MTBF 120 MTTR 10 (2/4)

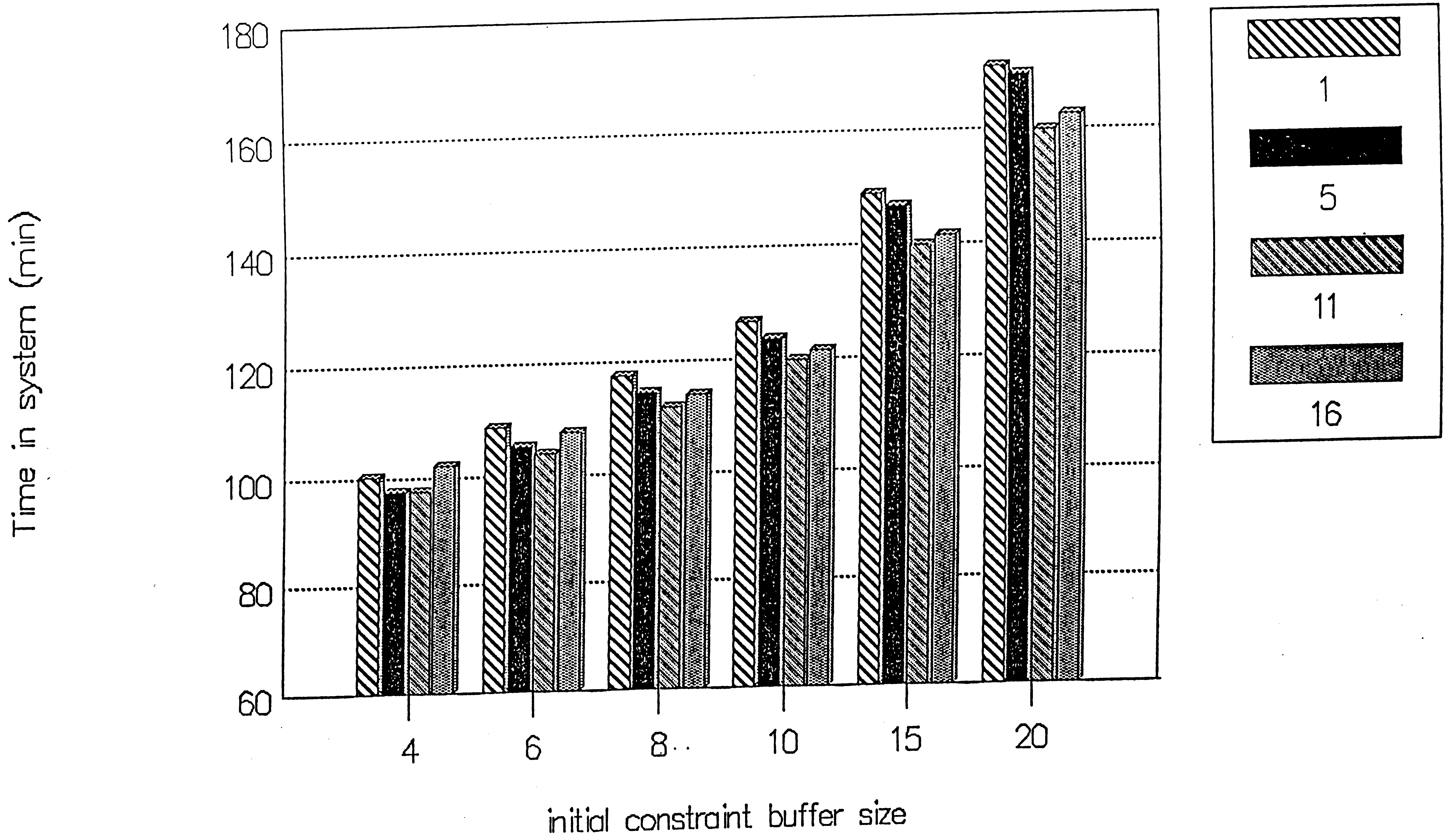
Graph 11



Time in system vs. Constraint position

MTBF 90 MTTR 10 (2/4)

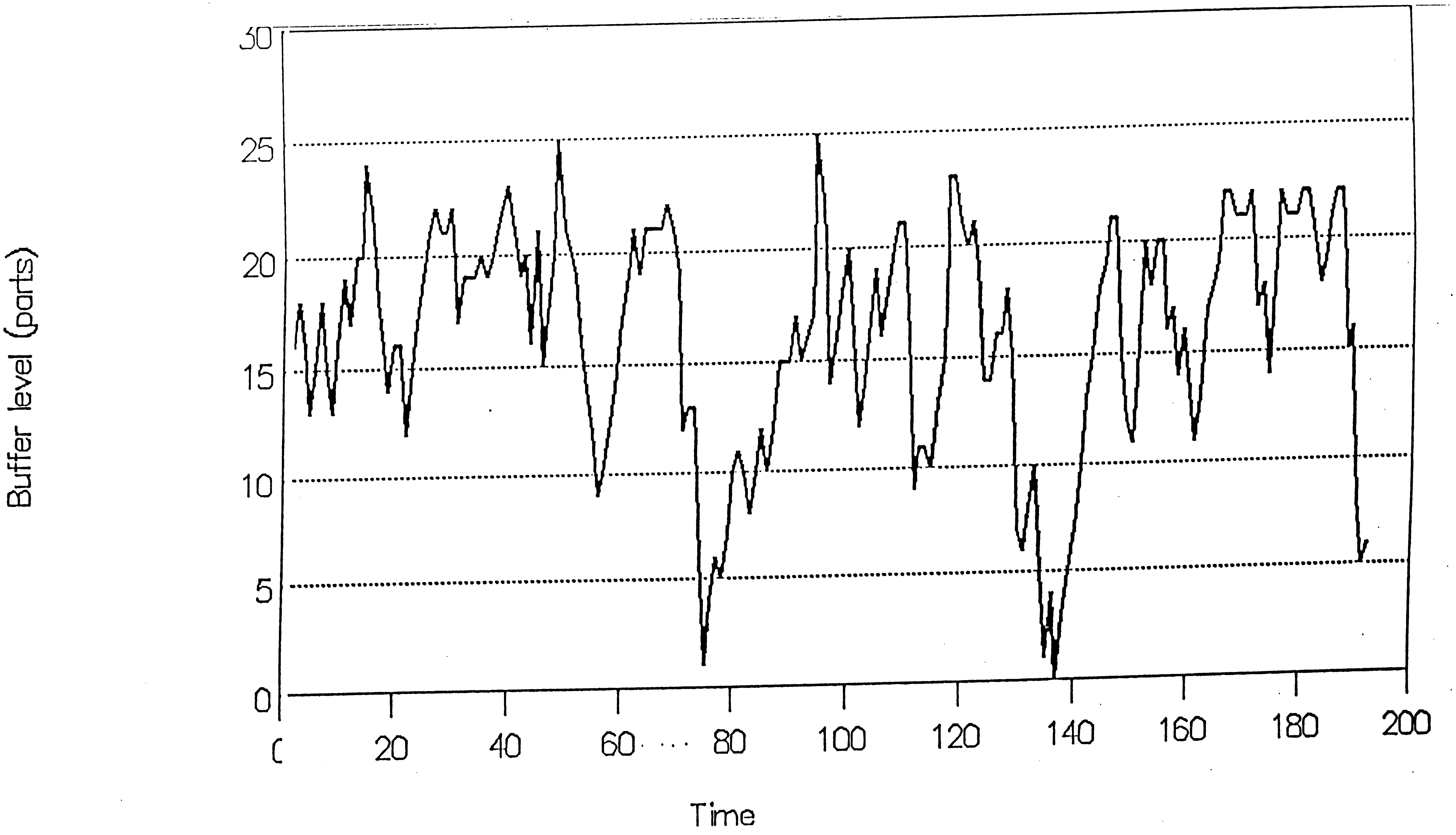
Graph 12



Buffer level before the constraint

MTBF 240 MTTR 10

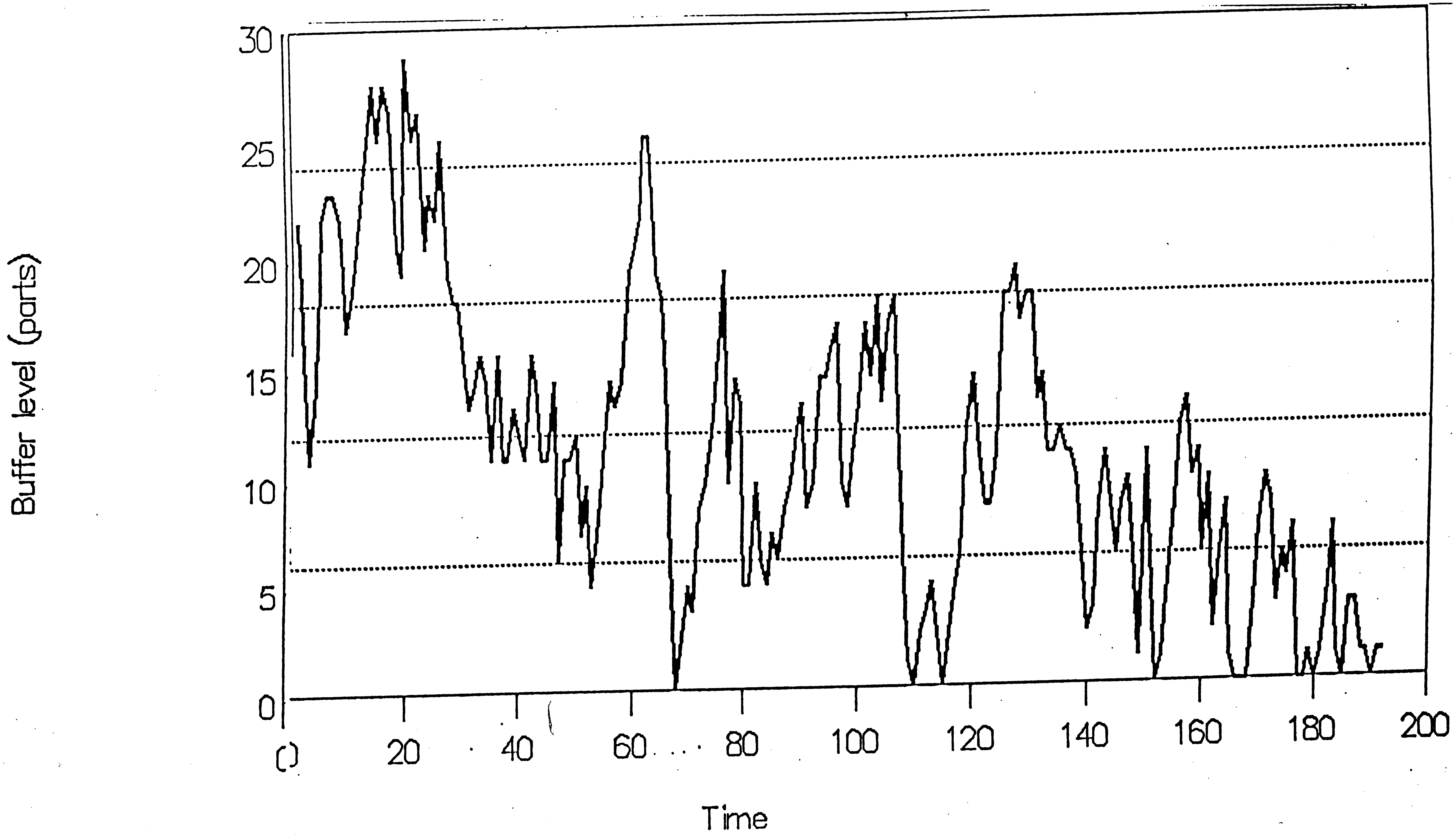
Graph 13



Buffer level before the constraint

MTBF 90 MTTR 10

Graph 14



Chapter 5

Applications

The previous chapters have described the criticality in sizing the processing and transfer batches. The impact of the sizes of these batches on the lead times and flexibility has been illustrated in chapter 3 with example. In the following sections, the impact of lot sizing at a major electronics firm has been discussed.

Thick Film Print Room at an electronics firm

The case in consideration is the proposed thick film print room at an electronics firm. Blank ceramic substrates undergo multiple screen printing steps till a complete circuit layout of resistors, dielectrics and conductors is developed on the substrates. These printed substrates are then used for electronics control for automobiles.

Layout

This area of the plant will have 11 printers/dryers and four furnaces. The material handling will be performed by an overhead monorail system comprised of self-propelled carts. Each printer/dryer and the furnaces have buffers before and after them which can hold magazines (containing 70 substrates).

Process

Blank substrates first undergo a few printing operations during which conductors or dielectric circuits are screen printed depending upon the type of substrate. The substrates are then fired in the furnaces. This process is called the first co-fire. Subsequent to the first co-fire, the substrate undergoes a few more printing operations and finally proceeds for the second co-fire at the furnaces. Of the available four furnaces, two are dedicated for first cofire and two for second cofire. The printers, on the other hand, are available in a common pool and can be used for any print step.

Current operating techniques

According to the existing operating strategy, a lot (typically comprised of about 40 magazines) is routed to a printer where the entire lot is printed. Only when the processing is complete, the lot is moved to the next printer which is then setup for the succeeding print step. The magazines are moved on push carts which can accommodate entire lot. Notable is the fact that the transfer batch equals the process batch.

At present, it generally takes 2 to 3 days for a lot to be entirely processed in the thick film room. Average work in process inventory is approximately 800 magazines. It is also common for the processing of a lot to be interrupted to accommodate a priority or *hot* lot.

Proposed operating techniques for the new facility

It was, understood that the new facility will require a new operating strategy to cut WIP and lead time. The strategy evolved in two phases. These phases have been described in detail in the following sections

Model Evolution

Phase I

The initial model assumed a conveyer as a material handling system and also included a central buffer. A central buffer was supposed to hold the magazines waiting for printing or firing. The aim in this modeling effort was to reduce the central buffer size to minimum while maintaining the required throughput.

Subsequently, a loading strategy based on the theory of constraint was developed. The furnaces were identified as the constraints in the system because of their slower processing rates. Hence the release of material into the system, in this case the magazines, was coupled with the output rate of the furnaces. Efforts were made to ensure that the furnaces never starved. The loading strategy was designed to setup two parallel *flow lines* in the system by loading only two different lot types simultaneously. These two different lot types were loaded in the system by putting one magazine of the alternating type at a given time interval. This time interval was the function of the processing rate of the constraint, namely the furnaces. The strategy recognized the fact that the furnaces were the constraint and loading the magazines into the system at a faster rate would increase the central buffer size. A new lot was loaded only when the last magazine of one lot had entered the system.

The simulation showed such an operating technique to be feasible and the required throughput was obtained while maintaining the central buffer size equal to approximately the lot size used.

Phase II

A decision regarding elimination of central buffer and use of overhead monorail system forced the model to be restructured to take into account these changes. The reason for elimination of the central buffer was in fact driven by the results obtained in the first phase of simulation which indicated very small size for central buffer. The second factor was the inability of the standard translogic software to handle the existence of the central buffer.

The new model had to account for the fact that a magazine waiting for printing or firing could wait only in the buffer lanes of the equipment since no central buffer was available. This required a new operating and loading strategy.

The new strategy recognized the fact that the furnaces were still the constraint and had to be protected. The difference between the old and new strategy was that now the printers or the furnaces could be *blocked* i.e. they can be stopped if the succeeding printer or furnace has no space in its buffer. Hence the new strategy was designed to ensure that the furnaces were continuously fed by the printers.

The number of different types of lots being loaded in the system simultaneously was varied for different experimental runs. This heuristic approach was used to determine the *optimal* number of lots to be simultaneously loaded so as to not to starve the furnace at any time. This number was determined to be 5.

The second important factor in devising the loading strategy that was recognized was the simplicity of the loading. Consider the beginning of the operations on the plant. Five different lots will be loaded, magazine by magazine, on five different printers. Once this is established, the loader has to keep sending the respective magazines to their assigned printers. A new lot is loaded only when two conditions are fulfilled:

(a) The last magazine of the lot has been loaded to the printers.

(b) The first magazine of the same lot has gone for the second firing.

The logic behind imposing such conditions is that the above two conditions will ensure that the lot will complete processing in near future without getting blocked. Consider a lot that has been introduced into the system. The first magazine will seize a printer and start its first printing operation. This printer will not be allocated to any other lot till the last magazine of the lot has arrived here. Meanwhile the first magazine, after finishing the first printing operation, will try to seize the next printer that is available. This process will continue till the first magazine proceeds for the first cofire. At such an instant a string of printers and furnaces will exist dedicated to one lot. Hence a equivalent of a flow line will be formed which will push out the magazines of that lot without getting blocked. When the last magazine is introduced in the system, it is an indication that as it travels down this line, it will free up the printers and the furnaces for the next lot. The results indicated that for tracking each lot, only the first and the last magazine has to be identified.

For the blocking strategy, it was realized that the lot size was critical for smooth operation. Too many lots of large sizes could block each other leading to an impasse. But if the lot sizes were limited to 40 magazines, then three printers were sufficient to perform all the necessary printing operations as shown in the attached figure. In such a case, the lot could loop back to the previous printers for the succeeding operations.

The simulation results also showed such a strategy to be feasible. The required throughput was obtained. It was also observed that the average WIP on the floor was around 70 magazines which should be compared with the present figures of around 1000 magazines. Second important result was the average makespan time for the lots. With the new strategy it was found to be less than a day as opposed to the existing figures of 3 days.

The advantage of such a low makespan time will be the possibility of giving equal priority to all lots instead of prioritization of some lots as *hot*.

The reason for this high makespan time in the existing setup has been identified to be because of the *policy constraints* rather than physical constraints. At present the transfer batch equals the process batch while in the new strategy, the transfer batch will be reduced to one. A lot is currently processed at a printer or a furnace and this entire lot is then moved to the next printer. In the new strategy it was suggested that multiple printers be setup for succeeding processes for one lot and thus each lot can simultaneously be processed at more than one printer. This was made possible by cutting the transfer batch to one magazine.

Summary

The new manufacturing environment is characterized by a strong international competition. Companies are being forced to produce a wider variety of products at a more economic costs and with better quality. Even though product design is consider to play an important role development of a superior product, yet it is imperative for the companies to produce those products more efficiently. The intensity for the need to produce efficiently is heightened by the fact that today business survival depends upon ability to compete on a national and international level. Also, it has been realized that few US manufacturers are competitive on a global scale. Japanese and the German manufacturers have been able to produce high quality products at a lower costs thus capturing a significant portion of the world market.

It is also comprehended that many of failures in the US business enterprises can be attributed to their failure to adopt modern management techniques. Most of the traditional manufacturing and management philosophies, adhered to by most US organizations, have focused on local optima instead of global optima. The drastic changes in the operations within the manufacturing organizations specially in the case of labor contents and product proliferation has led to invalidation of traditional operating methodologies. Only companies capable of manufacturing sophisticated products at a low cost at high quality and being able to provide competent customer service could survive. These companies adopted newer and more pertinent management philosophies which could deal with short product life cycles and increased international competition. The most outstanding features of most of these new management philosophies was the ability to focus on long term goals instead of short term goals. Also these philosophies tended to view the organization as a whole rather than discrete independent units.

The most popular of these management techniques and tools are JIT, MRP, SQC, TQC, Group Technology, Theory of Constraint etc. Those techniques were designed to improve the organization's operations. JIT, which has been implemented globaly with great success, is focused at reducing waste through reduction in the work in process inventory. JIT strives towards a flexible and efficient manufacturing outfit by reducing all production buffers from the shop floor. It's success is strongly dependent upon the interaction and the degree of rapport of the company with its vendors. The second critical feature of this technique is the workforce training. The system relies on the ability of its workforce to be sufficiently cross trained so as to be able to perform most the jobs within a team. MRP is a shop floor control techniques used extensively for material management specially in companies with a wide product spectrum and complicated routings. TQC and SQC were designed to approach quality problems more scientifically using statistical techniques to monitor processes. Group technology was created to reconfigure shop floor into cells based on the premise that smaller independent cells are easier to manage as compared to a large and often cluttered shop floor. Theory of constraint was another such systems thinking approach whereby the suggested approach to manage an organization was by successively concentrating on the systems bottleneck or the constraint. In the present manufacturing environment, with its extensive and complicated shop floor operations, this approach appeared to be an extremely simple and effective technique for management.

As stated earlier, archaic management philosophies had led to departmentalization of most production activities. This approach led to a drive for local optimization. The most glaring example of such an operating technique was the buffering or maintaining of work in process inventories before every machine or workcenter. This was done to provide a safeguard for the machines against fluctuations in the system. Such an approach led to very high levels of WIP on the shop floor. This consequently, led to high lead times, lack of flexibility and intractable quality problems. Another reason for the high WIP on the floor was attributed to the tendency to increase the efficiencies of all the machines-on a shop floor. This caused the supervisors to maintain high workloads on all machines even if the output was not immediately needed.

Theory of constraint operated with the understanding that the production rate cannot exceed that of the constraint and hence material release should be coupled with the constraint machine. Subsequently, the excess inventory before most of the machines could be eliminated since the faster machines will still be forced to produce at the rate at which the constraint produces.

On a shop floor, process and transfer batches play an important in determining the lot processing time. Shorter process and transfer batches lead to smaller lead times. The limit on reduction of process batches is imposed by the excessive cost or time involved in setting up more batches when the process batch size goes down. Hence a compromise between the size of process batch and the setup cost and tome involved has to be reached. Transfer batches are more dependent on the capability of material handling systems. Though with the existing systems often it is feasible to have transfer batches of single units.

In the framework described in this thesis, The production capabilities of the constraint machine were used as the basis for defining the batch sizes. A case was considered where the most heavily loaded machine had enough processing time available to process all the quantity needed for the given horizon with some time to spare. The model attempted to minimize the processing lots so as to minimize the average lead time for any given lot. The model constraints (not to be confused with the production constraint) were the setup times involved for each different part type. The model used the expected or forecasted demand for each different product type and generated a suggested lot size for each type.

In the second section of the thesis, a guideline for designing buffers before the constraint machine has been presented. Extensive simulation studies were done on a production line with sixteen machines in series. A wide range of operating conditions were simulated to analyze the performance characteristics of the system. The results were then discussed and suggestions for the buffer design were presented.

Finally, the application of the techniques and approaches,developed in the thesis, at an electronics firm were discussed.

BIBLIOGRAPHY

1. Berry, W.,"Lot sizing procedures for requirements planning systems:a framework for analysis", Production and Inventory Management, Vol. 13, No. 2, pp 19-34.

2. Blackstone, J. H., Capacity Management, Cincinnati: South-Western Publishing Co., 1989.

3. Callahan, J. M.,"GM's Plant Saver",Automotive Industries, March 1989.

4. Campbell, H. G., Dudek, R. A., and Smith, M. L.,"A heuristic algorithm for the n job, m machine sequencing problem", Management Science, Vol. 16, No. 10, 1970, pp 630-637.

5. Chuen-Lung Chen and Bulfin, R. L.,"Scheduling unit processing time jobs on a single machine with multiple criteria", Computers and Operations Research", Vol. 17, No. 1, 1990, pp 1-7.

6. Cohen, O.,Chapter 4 "Drum-Buffer-Rope Approach to Logistics,"Computer aided Production Management edited by A. Rolstadas, New York: Springer-Verlag, 1988.

7. Cox, J. F. and Finch, B. J.,"Effective placement of Manufacturing Inventories," American Production and Inventory Control Society 32nd International Conference proceedings, Orlando, FL., Oct. 23-27 1989, pp 658.

8. Goldratt, E. M., "Hierarchical Management-The inherent Conflict," The theory of Constraint Journal, Vol. 1 No. 1, Oct-Nov, 1987a.

9. Goldratt, E. M."Computerized Shop Floor Scheduling," International Journal of Production Research, Vol. 26, No. 3, March 1988

10. Goldratt, E. M. and Cox, J. The Goal. Revised edition, Croton-on-Hudson, NY: North River Press, 1986.

11. Goldratt, E. M. and Fox, R. E. The Race. Croton-on-Hudson, NY: North River Press, 1986.

12. Gupta, S. K., and Kyparisis, J.,"Single machine scheduling research", Int. Journal of Management Science, Vol. 15, No. 3, 1987, pp 207-227.

13. Groover, Mikell P., Automation, Production Systems, and Computer Integrated Manufacturing. Prentice-Hall, Engelwood Cliffs, New Jersey, 1987.

14. Larson, R. E., and Dessouky, M. I.,"Heuristic Procedures for the single machine problem to minimize maximum lateness", IIE Transactions, Vol. 10, No. 2, 1978, pp 176-181.

15. Larson, R. E., Dessouky, M. I., and Devor, R. E., "A forward-backward procedure for the single machine problem to minimize maximum lateness", IIE Transactions, Vol. 17, No. 3, 1985, pp 252-259.

16. Maxwell, W., Muckstadt, John., Thomas, J.,"A modeling framework for planning and control of production in discrete parts manufacturing and assembly systems", Interfaces, Dec. 1983, pp 92.

17. Mcmahon, Graham and Florian, Michael, "On Scheduling with ready times and due dates to minimize maximum lateness", Operations Research, Vol. 23, No. 3, 1975, pp 475-482.

18. Phillips, Don T., Garcia-Diaz Alberto, Fundamentals of Network Analysis, Prentice Hall, 1981, New Jersey

19. Plossl, G. W., and Wight, O. W., Production and Inventory Control Principles and Techniques. Prentice-Hall, Engelwood Cliffs, New Jersey, 1967.

20. Potts, C. N., and Wassenhove, L. N.,"A branch and bound algorithm for total weighted tardiness problem", Operations Research, Vol. 33, No. 3, pp363-377.

21. Schonberger, R. J., Japanese Manufacturing Techniques: Nine hidden lessons in Simplicity. Free Press, New York, 1982.

22. Schonberger, R. J.,World Class Manufacturing: Lessons of Simplicity Applied, New York: The free Press, 1986.

23. Sen, T., and Gupta, S. K.,"A branch and Bound Procedure to solve a bicriterion scheduling problem", AIIE Transactions, Vol. 15, No. 4, pp 84-88.

24. Shanthikumar, J. G.,"Scheduling n jobs on one machine to minimize the maximum tardiness", Computers and Operations Research, Vol. 10 No. 2, pp 255-266.

25. Skinner, W.,"The Focused Factory,"Harvard Business Review, May-June 1974.

26. Solberg, J.,"Capacity planning with a stochastic Workflow Model", AIIE Transactions, June 1981, Vol. 13, No.2, pp116.

27. Spachis, A. S., and King, J. R.,"Job-shop scheduling heuristics with local neighborhood search" Computers and Operations Research, Vol. 13, No.1, 1989, pp21-27.

28. Thitzy, J. M., and Wassenhove, L. N.,"Lagrangean Relaxation for the Multi-item capacitated lot-sizing problem: A heuristic Implementation" IIE Transactions, Vol. 17, No. 4, Dec. 1985.

29. Thuruthickara, C. J.,"Tradeoff solutions in single machine production scheduling for minimizing flow times and maximum penalty", Computers and Operations Research, Vol. 16, No.2, 1989 pp 471-479.

30. Umble, M. and Srikanth, M. L., Synchronous Manufacturing: Principles for World Class Excellence. Cincinnati: South-Western Publishing Co., 1990.