1990

An approach for designing protective capacity based on the tradeoffs between "work-in-process", "manufacturing lead time", and "capacity"

Ashish Masih
Lehigh University

Recommended Citation

Masih, Ashish, "An approach for designing protective capacity based on the tradeoffs between "work-in-process", "manufacturing lead time", and "capacity"" (1990). *Theses and Dissertations*. 5278.
https://preserve.lehigh.edu/etd/5278

An Approach for Designing Protective Capacity based on the Tradeoffs between "Work-in-process", "Manufacturing Lead Time", and "Capacity"

by

Ashish Masih

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
1989

An Approach for Designing Protective Capacity based on the Tradeoffs between "Work-in-process", "Manufacturing Lead Time, and "Capacity"

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science in Manufacturing Systems Engineering

July 7, 1989

Date

Roger N Nagel

Dr. Roger N. Nagel
Professor in charge

Dr. Keith M. Gardiner
Director, MSE Program

Marlin U Thomas

Dr. Marlin U. Thomas
Chairman, Industrial Eng. Dept.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Roger N. Nagel for his time, help, and encouragement all through my association with him. I specially wish to thank him for sharing his vast experience and motivating me to come up with new ideas for the research.

I am grateful to the faculty and staff of MSE for making my academic and work life at Lehigh a very pleasant experience and for their help on numerous occasions. For, providing me with valuable help and encouragement during all stages of my studies and work, I would like to thank all my friends at Lehigh and the MSE class of 1988.

Finally, I wish to thank my parents and my sister for their help and support.

TABLE OF CONTENTS

CHAPTER PAGE NUMBER

LIST OF FIGURES

ABSTRACT

Capacity decisions and capacity planning/management are vital to a manufacturing organization's competitive position. Apart from the nominal or needed capacity, adequate "protective" capacity is required for protection against the uncertainties within the organization, in the market and in its environment. Also, traditional organizational slacks such as excess work-in-process inventory and long lead times, that were supposed to protect the orgnization, have other significant drawbacks. Thus a case has been presented for designing "protective" capacity in a production system.

There exist strong interdependencies and tradeoffs between capacity, work-in-process and manufacturing lead time. These tradeoffs are the basis of a framework for designing "protective" capacity both at the workstation and production system level, as well as for long and short term purposes. The frame work has, then, been discussed with respect to relevant issues such as flexibility, amount and type of "protective" capacity, relationship to forecasting, and the conversion of a buffer (wip) protection into equivalent "protective" capacity.

CHAPTER I

INTRODUCTION

1.1 Long term competitive strategy and manufacturing strategy of a manufacturing organization

The long term competitive strategy of any company must have manufacturing as its integral component to obtain a sustainable competitive advantage.[11] It should not just assume the performance of the manufacturing function, but involve it in the strategy formulation process. This is because the manufacturing strategy includes crucial aspects of the company which affect its position in terms of capacity/growth possibilities, value added at various stages of the manufacturing process, and similar factors which determine the key areas of advantage and disadvantage for the company and its competitors. Buffa identifies six basics of manufacturing strategy; positioning the production system, capacity decisions, product & process technology, work force & job design, operating decisions, and suppliers & vertical integration.

1.2 Relationship of capacity planning with manufacturing strategy

As mentioned above, capacity related decisions are a part of the manufacturing strategy and deals with the issue of balancing predicting future demands and impact of new technology with physical capacity requirements, alternate plans, and the related economic effects. This implies building protection from changes in the above-mentioned factors. Poor capacity decisions can have a disastrous effect on the manufacturing strategy. If adequate capacity protection is not available, then the response to the customers will not be adequate due to missed due dates, long lead times, etc. On the other hand excessive overcapacity can also ruin a product cost advantage. In the case of new product introduction, the capacity planning must also be done simultaneously for needed and "protective" capacity to be able to successfully penetrate the market. The business plan may be right but if adequate capacity does not exist to manufacture the right products at the right time, the favorable market conditions would be lost forever. This is a case of mismatch of marketing strategy and manufacturing capability and capacity.

1.2.1 Concept of needed and "protective" capacity

We can observe that the success of any strategic corporate plan very much depends on the capacity and potential capability of the manufacturing facility. This capacity is dependent on factors such as technology, product life cycle, product complexity, equipment costs, equipment lead times and nature and composition of the work force.

It is very important to be able to measure, plan and manage/control this manufacturing capacity so as to be able to achieve a satisfactory match between demand and available capacity. Due to the uncertainty in the customer demand, there will be periods with demand levels less than nominal capacity leading to underutilization and there will also be periods when the demand exceeds the capacity requiring extra or "protective" capacity to satisfy it. Thus a feasible capacity plan should be able to design both the needed and "protective" capacity in the production system. Another important factor is the increasing sophistication of technology, and it has allowed the manufacturers to obtain flexible production capacity which is less sensitive to the effects of product design, customer modifications and schedule changes and enables them to be more competitive. This also implies that the ability of the manufacturing facility to adapt to changes in the customer demands (volume and variety), both long term and short term has improved.

1.3 Slacks in a manufacturing organization

Any manufacturing organization has slacks inherent in its organization. These are manifested in various forms at different stages in the manufacturing and order cycle.[82] Some of these are:

- excess inventory in the raw material, work-in-process and finished goods areas
- excess direct and indirect labor
- overtime costs
- long manufacturing lead times and poor due date performance
- long new product development cycles
- lack of responsiveness to changing business environment

The traditional management attitude and perception is that these slacks serve to protect the system from the uncertainties and the unknown. But on the contrary, they do more harm than good to the system and the management strategy should be towards a systematic reduction of most of these slacks so as to enable better performance of the manufacturing organization. The better performance will result in improved organizational productivity, reduced overtime, reduced inventory investment, reduced obsolescence, improved inventory turnover, lower purchasing, manufacturing and distribution costs, and improved customer service levels; all these resulting in better manufacturing capability/capacity and improved bottom line performance measurements. It must, however be understood that certain types of slacks at particular locations in the organization may be very beneficial towards improving the performance measurements by protecting the critical resources of the production system. It is therefore very important to identify the types, locations and values of these slacks to be able to obtain the proper benefits rather than detrimental effects which may be quite prominent if the types, locations and values of the slacks are not right.

1.4 Scope of this research

The objective of this research is to provide a framework for designing "protective" capacity both, for long term and short term purposes based on the tradeoffs between work-in-process, manufacturing lead time and capacity. Conventionally, manufacturing organizations have protected themselves against the uncertainties by providing excess work-in-process and quoting long lead times. Excess or idle capacity, has on the other hand, been regarded as more of an evil and cost burden. A rationale for using "protective" capacity has been presented for protecting a manufacturing organization from the uncertainties & variations and methods to compute/manage it are also described.

Chapter II looks at "protective" capacity through the issues of capacity , work-in-process, and manufacturing lead time. The measures of capacity, the need for and current techniques of capacity planning are described very briefly along with the drawbacks and advantages of each. Other relevant issues such as effect of capacity unbalance in a plant, bottleneck & non-bottleneck resources, and the consequences of high capacity utilizations and efficiencies are discussed. The importance of manufacturing lead time, the causes of variability in it, and methods of its control are then illustrated. Finally, the causes and functions of work-in-process and its importance and relevance are described. In each case, the issue of "protective" capacity, as and when relevant, is related to the discussion.

In chapter III the tradeoffs between capacity, work-in-process and lead time are demonstrated. It is done at the workstation level, first by the graphical method and then by the relations developed as a part of the capacity requirements planning model of Karni. The capacity planning methodology at the production system level is explained through three distinct models, and in each of these, the tradeoffs, as they are apparent through the relevant inputs and results are depicted. In each of these discussions, the method of designing "protective" capacity at workstation and production system levels is emphasized.

Chapter IV presents a framework for designing "protective" capacity into the production system. The framework is based on the tradeoffs (both graphical and relations) and the capacity estimation model presented in a previous chapter. Issues and questions relevant to "protective" capacity are discussed in detail based on the framework.

Chapter V presents a summary and conclusions.

CHAPTER II

OBSERVATIONS ON PROTECTIVE CAPACITY THROUGH THE CONCEPTS OF CAPACITY PLANNING/MANAGEMENT, MANUFACTURING LEAD TIME (MLT), AND WORK-IN-PROCESS (WIP)

2.1 Capacity planning / management

The APICS dictionary defines capacity as "the highest reasonable output rate which can be achieved with the current product specification, product mix, work force, plant and equipment". It is actually a constraint on the manufacturing system. The available capacity of any facility is dependent on a number of factors which can be classified as planned factors and monitored factors. The planned factors are land, space, constant labor force, machines, technology, shift and overtime decisions, subcontracting and learning curves. The monitored factors are unplanned orders, scrap and rework, material shortages, absenteeism, labor problems, and machine breakdowns. An issue related to capacity is "load" and is the amount of work scheduled to be done by the manufacturing facility.[2]

2.1.1 Measures of capacity

Capacity of a manufacturing facility is measure of output, and is expressed in number of hours of production available over a period of time period such as day, shift, week, month or quarter. If the facility is manufacturing only a few products requiring very similar resources per unit then the capacity can be stated in terms of units manufactured per period. The choice of the measure of capacity should be chosen based on what most affects the actual capacity to fulfill the production plan.[82] Thus the measure may be based on the key resources of the manufacturing system. Thus, if labor is a key resource, then labor hours may be an appropriate measure of capacity; likewise it may be hours available in a particular work area. Some very relevant issues such as the effect of current trend of reduced labor content of the product cost and the changing nature of manufacturing technology (e.g. becoming capital intensive) must be understood before selecting the measure of the capacity. Finally, the choice should represent the understanding of all the people responsible for monitoring and planning capacity of the production system.

The load on the system is typically expressed as hours of production or units of production per period. It is the work input to a resource and the capacity of the resource determines how much time will be required to complete the work.

2.1.2 Need for and current techniques of capacity planning

Any manufacturing plan that exceeds capacity (based on the "capacity" and "load" comparison) is not feasible and cannot be achieved. The role of the capacity planning/ management function is to change capacity over time so as to meet the short and long term production goals. Thus the capacity requirements for the future time periods must also be known along with the currently available capacity. The required capacity plan is derived from a combination of business plan, production plan, master production schedule and material requirements plan depending on the length and extent of the planning horizon. Thus, the aim of any capacity planning methodology should be to enable the fulfillment of the production plan by planning for the required capacity. This planned capacity is needed in the right time periods; if not then the production plan has to be modified and the capacity planning cycle repeated again resulting in an iterative methodology.

The capacity planning techniques currently in vogue are known as Resource requirements planning, Rough-cut capacity planning, Capacity requirements planning and finite capacity loading in the order of decreasing level of aggregation.[2] The techniques used for determining the capacity requirements are known as capacity planning using overall planning factors (CPOF), resource profiles, capacity bills and capacity requirements planning. CPOF can be performed with standard accounting data while resource profiles needs somewhat detailed end-product information.

The technique of capacity bills also provides information about capacity requirements according to time periods whereas capacity requirements planning is a very comprehensive method and utilizes the total manufacturing resource planning database for obtaining the results.

The different capacity planning methods are used at different levels in the hierarchy of a manufacturing resource planning (MRPII) system to plan at varying levels of aggregation. Resource requirements planning is the tool that is used at the highest level to identify the aggregate level of major resources required to meet the production plan. The critical resources are also identified and included in the "resource profile". This allows the management to compare the production plan to the critical resources in a realistic manner. Rough-cut capacity planning constructs resource profiles for each item in the master production schedule, and provides a more detailed breakdown of resources as compared to the first method. In the method of capacity requirements planning, the level of detail is highest and time horizon shortest (generally the planning horizon of the MRP system). It determines the amount of labor and machine resources necessary to meet the material plan over the planning horizon.

2.1.3 Effect of capacity unbalance in a plant

The strategies for capacity management are have traditionally been oriented towards planning and control of production as if the plant has balanced capacity.[27] In reality, no production facility is, or can be, balanced and there are a number of valid reasons for it.

The production of any unit in a plant is composed of a series of processes, each with a standard processing time. But, the actual processing time is slightly different from this standard time; this is the because the processes are inherently, stochastic or non-deterministic in nature. The effect of this stochastic nature is magnified by the fact that the processes are in series and interdependent because of technological reasons (process routings). The deviations of these sequential processes get accumulated and get magnified by the time all the processing on the product is completed. Hence, the combined effect of stochastic and the phenomenon of interdependence generates a pattern of accumulated delays and which increase downstream in the processing sequence.

The management emphasis has generally been on balancing the plant and also on balancing the plant capacity to the production level that is required, and the effect these efforts can be quite harmful. The balancing is achieved by putting work-in-process between unbalanced processes and this results in ever increasing inventory levels. The increased inventory levels will also increase the cycle time of the product to a large extent. Also, the throughput will get reduced due to the increased inventory levels. Thus, it seems that the ideal approach should not be to balance the plant but to manage the imbalances in a better manner by balancing the flow of the product through the system.

2.1.4 Concept of bottleneck and non-bottleneck resources

Based on the discussion in the previous section, it is obvious that any typical plant is not a balanced one. This implies that some of the workstations, or resources will be and some will not be running at full capacity due to the inherent differences and variations in the processing times. Thus there will be some resources that will act as bottlenecks (a resource with capacity less than the demand placed on it) in the system and the rest will be non-bottlenecks. Another important concept relevant here is of activation and utilization.[52] It should be understood that the ability of the system to produce is constrained by the bottlenecks and hence, to activate a resource when the resulting output is not able to go through the bottleneck only creates excess work-in-process inventory. In this case this particular resource is only being activated and not really utilized and the actual level of utilization of non-bottleneck resources is dictated by the bottleneck resource.

The concept of a capacity constrained resource is also related to the issue of bottlenecks, and is defined as a resource whose utilization is close to capacity and it could become a bottleneck due to bad scheduling. This can be caused by changing batch sizes or excessive machine downtime.

Another important issue is that whatever time that is lost at the bottleneck resource directly affects the throughput of the system and the system loses throughput worth that time. But this is not the case for the non-bottleneck resources. Following the same logic, time saved on the bottleneck resource directly benefits the system throughput and the same time saved at a non-bottleneck resource is actually a mirage. Thus one needs to "protect" the bottleneck resource so that it does not lose any time and hence the system does not lose the associated throughput (production capacity). This protection prevents the system from losing capacity on this bottleneck resource and losing valuable throughput. The above mentioned protection may be done by placing a buffer in front of the bottleneck resource or by providing significant "protective" capacity in the non-bottleneck resources so they may be able to always generate enough work to keep the bottleneck busy and utilized to full capacity or by some other suitable method. But the bottleneck has to be protected against losing any valuable throughput by providing the right amount of capacity, work-in-process, or lead time. This protection should also be augmented by providing extensive management attention on the bottlenecks to minimize downtime and other related factors which may cause the loss of throughput.

2.1.5 Myth of high capacity utilizations and efficiencies

The traditional cost accounting measures have forced the evaluation of machines and workstations based on efficiencies and utilizations. These efficiencies are computed by comparing actual performances with time-standards. As discussed in the previous section, 100% utilization is not necessary at all the workstations and is in fact harmful at non-bottleneck stations due to the negative effects of high work-in-process inventory. It should be noted that this rule implicitly assumes that the production system is perfectly balanced which is never the case.

The above mentioned attitude of high utilizations is based on the strong emphasis by traditional cost accounting systems on costs associated with workstations and hence the need to "fully" utilize the station to financially "justify" it.[88] This emphasis has developed over the years and was a result of considerable high labor content of product cost.[21] But the current reality and trend is very clearly oriented towards decreasing labor costs and increasing material costs. These trends have made the paradigm of 100% utilization obsolete and calls for minimizing inventory. Actually high inventory is bad not just due to the carrying cost reason but more so because of its effect on lead time and response time. To complete the argument, "protection" by work-in-process does not seem to be the best method.

2.2 Manufacturing Lead Time (MLT)

Time is one of the most critical resources relevant to a manufacturing system. The relevant time measurements are manufacturing lead time and also various other lead times (purchasing, order, etc.) defined in that context. As with other resources, it needs to be managed well in so as to improve the bottom line measurements.[63] There are two misconceptions which are sometimes prevalent in the manufacturing world and these originate from the design of manufacturing planning and control systems that are in vogue; that control of purchasing and processing time is not possible leading to manual adjustment (i.e. increase the lead time to achieve protection) of planned lead times in these system databases; and that a task can be done better by taking more time leading to acceptance of always increasing lead times.

2.2.1 Definitions

At this stage the definition of manufacturing lead time and related factors should be stated and properly understood.[63]

"Manufacturing lead time" is the total elapsed time from the determination of the need for an item made in the factory until it is available to the customer".

"Purchasing lead time" is the total elapsed time from the determination of need for an item procured from an outside vendor until it is available for use.

"Order lead time" is the time required after the receipt of a customer's order to ship the ordered items.

Both the manufacturing and purchasing lead times have components that can be classified as follows:

1. preparation or paperwork time for the order
2. setup time
3. run time to process the order or the batch
4. move time to transport the batch between the workstations
5. waiting time spent in the queue

It is an accepted fact that the actual set up and run times compose only a very small fraction of the total order lead time and that the queue or wait time is its highest fraction. Also, there may be a number of sub-assemblies coming into an end product in parallel; this implies that the manufacturing lead time will be the sum of the critical path activities required to produce the particular product.

2.2.2 Importance and relevance of manufacturing lead time in today's business environment

The trend of increasing product variety and decreasing product life cycle have led to smaller lot sizes and demanded high flexibility from the current manufacturing systems. This is also coupled with increasing worldwide competition leading to short delivery times. The combination of these phenomena have demanded shorter and more reliable order lead times from the manufacturing organizations of today's businesses. Since manufacturing lead time is a prime component of the total order lead time, the objective should be to facilitate actions which result in its reduction.[4]

The effect of decreased manufacturing lead time and hence the order lead time is very pronounced on the bottom line performance measurements of an enterprise. This kind of response will ensure adequate and continuing business from the customers resulting in increased throughput. The decreased inventory level which is the consequence of shortened manufacturing lead time results in some very sizable benefits to these bottom line measurements and they will be discussed in detail in a later section on work-in-process inventory.

Long manufacturing lead times (which seemingly are supposed to protect the system) also affect the manufacturing planning and control system of a company. Together with the fast data crunching and replanning ability of today's computer based systems, the assumption that longer planning and execution of manufacturing plans is better can yield in disastrous results. This is specially true because of another assumption in these systems, particularly in MRP systems; i.e. the assumption of fixed manufacturing lead times in the preparation of manufacturing plans. Manufacturing lead times are actually a result of planned schedules and the way that they are executed and should not be an assumption which drives the schedules.

Another important effect of short manufacturing lead times is to reduce the manufacturing planning and control system nervousness.[42] If the lead time is longer, the chances of the customer changing the quantity and the due dates of the open orders are higher. This leads to reshuffling of priorities and changing schedules resulting in waste of valuable capacity which could have been used to process the products that were really needed. Thus, shorter lead times insulate the system from these changes and decrease system nervousness.

2.2.3 Variability in manufacturing lead time

Apart from the fact that long lead time is harmful, it is the variability in that lead time that also causes significant problems. As was discussed in an earlier section, the process variations at each stage get accumulated and magnified as the product moves downstream. Another form of variation is early completion of orders on the shop floor.[34] This has two different effects. Firstly, capacity devoted to completing unneeded work cannot be used to complete the needed orders. Secondly, early completions go into finished goods inventory for no good.

The source of lead time variability is Murphy and ineffective manufacturing engineering practices. Among the manufacturing engineering practices, poor plant layout, processing technology, and setup/tool/fixture design are prominent. Another reason is ineffective practices in following production schedules. In many cases, due to the long setup time, a worker may run batches of similar parts back to back so as to share the setup. What is also important is that longer the lead time, the greater the possible variation in it.

2.2.4 Manufacturing lead time control

The manufacturing planning and control systems (particularly, systems such as MRP) are operated by using some assumed value of lead time for any end product, subassembly or a component. This assumption is probably based on experience, or on estimates of various components of lead time or a combination of such factors and also some extra time added to "protect" it from uncertainties. When this lead time is used to explode the bill of material and if there are a significant number of levels, the effect on the cumulative lead time of an end product as estimated and "planned" by the MRP system can well be imagined.

2.2.4.1 Vicious cycle of constantly increasing lead times

There is also a vicious cycle of constantly increasing lead times which can happen if they are not managed properly.[63] This happens when orders start getting piled up and a backlog develops. In such a situation the load exceeds the capacity and due dates start getting missed. The planning system, upon sensing this, increases the planned lead time for increased "protection" which is used to develop the material requirements and schedules. This action automatically releases more orders to the shop floor and hence work centers get loaded more increasing the queue lengths. Thus, due to increased queue lengths actual lead times get longer and more delivery dates will be missed and the cycle repeats itself. The end effect of this method of managing lead times is higher work-in-process, less valid schedules, shortages and increased overloads. We have observed that any effort to manage lead times by bringing planned and actual values closer by changing the planned time is very harmful. The preferable method is to manage the capacity to take care of extra orders, if they are, and keep the queues in control. So, it seems, that a probable and better method may be to have the right amount of "protective capacity" to handle the situation.

2.2.4.2 Indicators of lead time control - flow time and allowance time

Following the idea of using actual times as the manufacturing lead times, relationship has been developed between this actual value (L_a) and the planned lead time (L_p).[42] This will provide a method for determining the offset that the two values are from each other to enable proper control of the actual lead time. Let:

d = order due date

r = order release date

c = order completion date

Thus we have:

$$L_p = d - r$$

$$L_a = c - r$$

It should be understood that management has real and direct control over L_p and it can be changed by a simple managerial decree to the manufacturing planning and control system. But the situation is different in the case of L_a which is actually a combined effect of a number of management actions undertaken on the shop floor.

Whenever a production plan is established and compared to the existing capacity scenario, the implicit assumption is that L_p will be equal to L_a for each order. The deviation of L_a from L_p is an indication of the effectiveness of the capacity planning method and subsequent shop floor control system to execute it. A useful value to compute at this stage is order lateness (OL) which defined as:

$$OL = L_a - L_p$$

This value of OL for each open order on the shop floor is a very good measurement to monitor the manufacturing lead time and the effectiveness of the control methodology being used.

2.2.4.3 Methods of control

From the previous sections it is clear that manufacturing lead time is a very critical parameter and is valuable as a "protective parameter" and as a control value to monitor responsiveness and inventory levels. The value of the lead time also drives the MRP routine and determines the effective length of planning horizon.[87]

The aim should be to effectively manage and control the manufacturing lead time by viewing it as an important and controllable resource. As mentioned previously, rather than fall into the trap of constantly increasing lead times by changing its planned value, suitable capacity management must be observed. To achieve the above, some ordered set of guidelines may need to be followed.[9]

The most significant factor that should be controlled is the length of queues at any workstation and a good method to achieve this input-output control. This method ensures adequate visibility by keeping track of queue lengths with respect to the work input and actual output in each time period. This implies computing and making sufficient capacity available through the complete planning horizon; thus the need for "protective" capacity for effective lead time management & control and to be able to keep them short.

2.3 Work-in-process (WIP)

Inventory in any manufacturing plant is present in three different forms: raw material, work-in-process, and finished goods. APICS defines work-in-process (work-in-process) as "product in various stages of completion throughout the plant including raw material that has been released for initial processing, upto completely processed material awaiting final inspection and acceptance as finished product. Many accounting firms also include the value of semifinished stock and components in this category". The control of raw material inventory is probably in the hands of the purchasing section and that of finished goods with sales and marketing group. The control of work-in-process is more difficult compared to the other two types of inventories mainly due to the complex nature of the movement of parts in various stages of semi-finished states on the shop floor. The state of work-in-process is very dynamic in nature and somewhat real time information has to be maintained about its location so as to control it. The issue of inventory record maintenance and record accuracy also becomes very relevant and is crucial in achieving the above mentioned control.

2.3.1 Causes and functions of work-in-process

The omnipresent work-in-process is an integral part of any production environment whatever be the kind of manufacturing planning and control system that is being used. The causes and functions of work-in-process are very much interrelated and are worth investigating into. There are good and bad effects of work-in-process; may be more bad reasons; but definitely a trade-off has to be achieved which will be dealt with in greater detail later on. Wip has always been "protection" against the unknown; it enhances the comfort level and fulfills a psychological need; it fills up the pipeline and keeps everybody busy; it is there just in case the customer needs the part. But does it always serve that purpose i.e. to satisfy the customer whenever he wants the part at the right time?

The causes and reasons for keeping work-in-process or for work-in-process to be present on its own are many.[84] A very prominent cause for the existence of work-in-process is the use of batch sizes for the processing and movement of parts on the shop floor. It is common knowledge now that the models of computing batch sizes such as Economic order quantity (EOQ) look at optimizing a very micro level problem and may cause large batch sizes and hence significantly large work-in-process. It may indeed minimize combination of setup and carrying costs but the harmful effects on the manufacturing lead time are not at all considered by the model.

The other very widely prevalent cause of high work-in-process is the use of buffer stocks between different workstations for "protection" purposes. It offers protection against unreliability in processes, and also from the differing cycle times of processes which may be operating in series. The more different these cycle times are from each other and lower the reliability of the equipment, the greater will the need for these buffers. Thus, a buffer also decouples a stage of a production system from another and protects one stage from another. Another important although unnoticed reason is a situation where one component may be required in many end products. This type of situation causes highly lumpy and seemingly uncertain demand for such components, specially in a MRP kind of planning environment and causes one to carry larger "safety" stocks.

Management policies are also largely and often responsible for increased work-in-process. These policies emphasize the performance measurements such as high utilization of labor and equipment which increase work-in-process. Thus line management is only concerned with high labor utilization reports and no one person may be responsible for overall flow of the parts through the system. This again reemphasizes up all the arguments presented in section 2.1.5 about the myth of high capacity utilizations and efficiencies and all them are valid here in increasing the work-in-process levels.

The design and layout of any manufacturing facility also contributes towards increasing work-in-process levels. This problem is not manifested in a flowline kind of production system. In the case of workcell kind of layout, the flow of parts in intermediate stages of processing is rapid within the cells. But the flow may be rather slow in the case of inter-cell work transfer leading to work-in-process creation at these stages. In the traditional job shop kind of layout, the flow problems are maximum and the so is the work-in-process level. The type and effectiveness of the shopfloor scheduling system will greatly determine the work-in-process levels that will be created. This ties with the management policies and the manufacturing planning and control system in use. The effectiveness of the scheduling system impacts the work-in-process level most in a job shop kind of environment and least in the flow line. The scheduling is also more complex in this kind of environment and bad schedules can really clog the shop floor by wrong sequencing of the jobs.

2.3.2 Relevance and impact of work-in-process in today's business environment

It was mentioned in section 2.2.2 about the trend of increasing flexibility leading to increasing product variety, decreasing product life cycle, and smaller lot sizes. All this has led to customers demanding shorter delivery times. It is quite apparent that when the work-in-process level is high, the corresponding lead time is also longer resulting in a delayed response to the customer. Also, if the work-in-process is high, the ability to respond to changes in customer specifications (engineering change orders) and to push through new or different products is reduced. This is obvious, since if the shop floor and the machines are already busy and loaded with jobs with significant queues, then any change is difficult to be effected even and will only be possible through considerable expediting. This would again, complicate the schedules with the associated increase in work-in-process and lead time. Thus an important impact of reduced work-in-process is shortening of the lead time and reducing the opportunity costs due to lost sales or market share.

The level of work-in-process affects directly the state of the cash flow in any company by reducing the investment in material thereby reducing the inventory carrying costs. Thus it directly affects the return on investment and cash flow situation of the company by reducing the cash investment. But through the method of carrying cost, it also indirectly, affects the operating expense of a company which in influences all the bottom line measurements i.e. net profit, return on investment and cash flow.[29]

The measures of quality and associated scrap are also influenced by level of work-in-process. With reduced work-in-process, the cost of storage, insurance, obsolescence, and scrap are reduced. More important of all lower the inventory, the higher the end quality is going to be. If the inspection is done at the end, then in case of a quality problem it is very difficult to trace the cause of the defect.[29] This is specially true because the lead time would also have been long and thus the root of the problem will never be traced. Also, a big batch of parts may have to be scrapped because of this defect. If the batch sizes were small with the associated low work-in-process levels, the cost of scrapping will not be that high and also the root of the problem can be traced to prevent the defect from occurring again. Thus it seems that work-in-process also "protects" the defects and problems from surfacing and being detected.

Another consequence of work-in-process level is related to its direct correlation with manufacturing lead time and this equivalence will be discussed and explored in more detail later. But this effect on lead times influences the forecast validity also. In any forecast there is a frozen and reliable segment of forecast the length of which depends on the customer demands and expectations.[29] If the lead time is long, then this is also comparatively long. Hence, this forecast reliability is dependent on work-in-process level and the benefits of a reliable and valid forecast are well known and need not be repeated.

CHAPTER III

TRADEOFFS BETWEEN WORK-IN-PROCESS (WIP), MANUFACTURING LEAD TIME (MLT), & CAPACITY AND THEIR RELATIONSHIP TO CAPACITY PLANNING

3.1 Tradeoffs between work-in-process, manufacturing lead time and capacity for workstation - graphical depiction

The individual factors of work-in-process, manufacturing lead time and capacity planning/management were discussed in the previous chapter. There exists very strong interdependencies among them which are also the basis for the tradeoffs. The relationships will be discussed with respect to a single workstation initially and then depicted for a more complete production system.[4] It can be depicted very clearly in the form of a diagram (figure 1). The input to any workstation is in parts per time period or even in some standard units such as labor or machine hours per period. The is a queue in front of the workstation and the units are number of jobs (converted to consistent units). The output of the workstation depends on the capacity and is in the same units as the input. The following relationships, though simple are valid and very important:

Average input = I_{av}

Average output = C; equal to utilized capacity

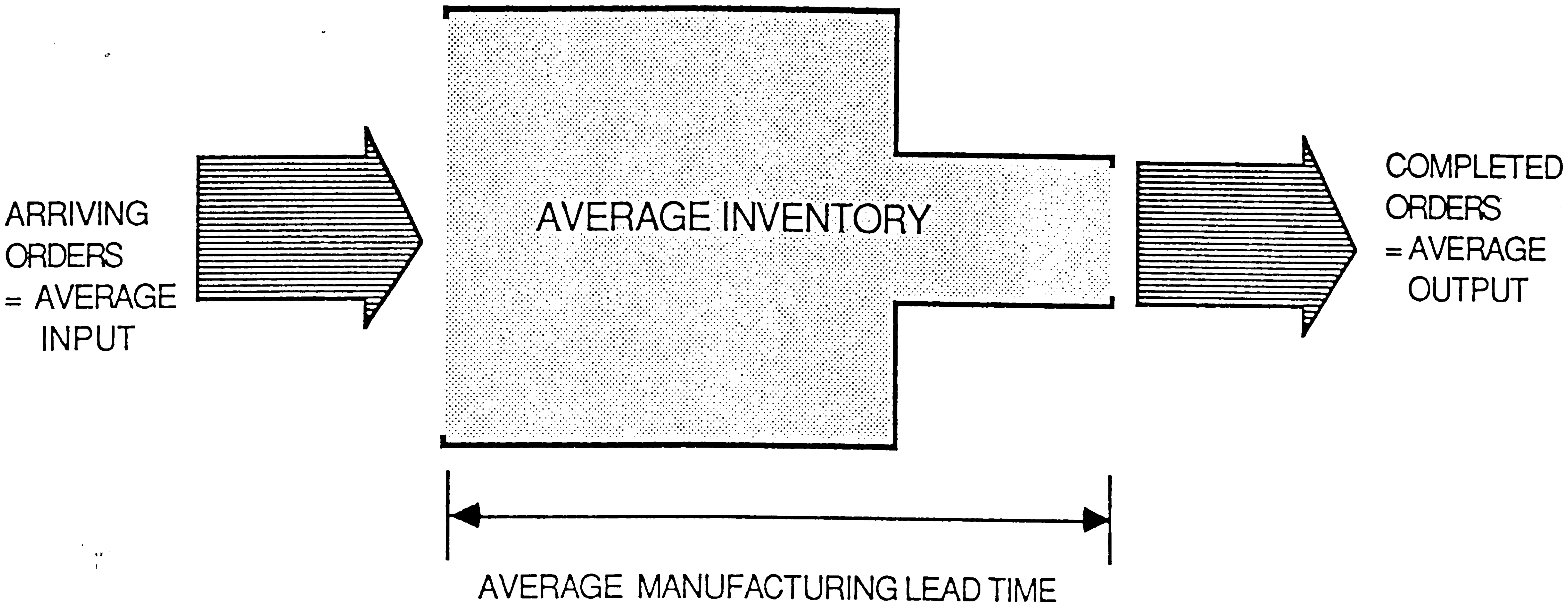

Figure 1: Input/Output relationship for a workstation

Average work-in-process = W_{av};

Thus we have:

Average lead time = average work-in-process/average output

This above very explicitly models the dependence of lead time on the work-in-process size. This also implies that the average lead time is equal to the time required to turn the inventory around once.

The above mentioned relationships are valid at an average level; in reality the transient behavior is very stochastic in nature. This makes the modeling of large real life production systems very difficult and is apparent by the modeling attempts that are available in literature. The input to a workstation is not smooth; the reason is not only the stochastic behavior of the previous workstation's output but also due to the variation in setup times, batch sizes and processing times for different products. The

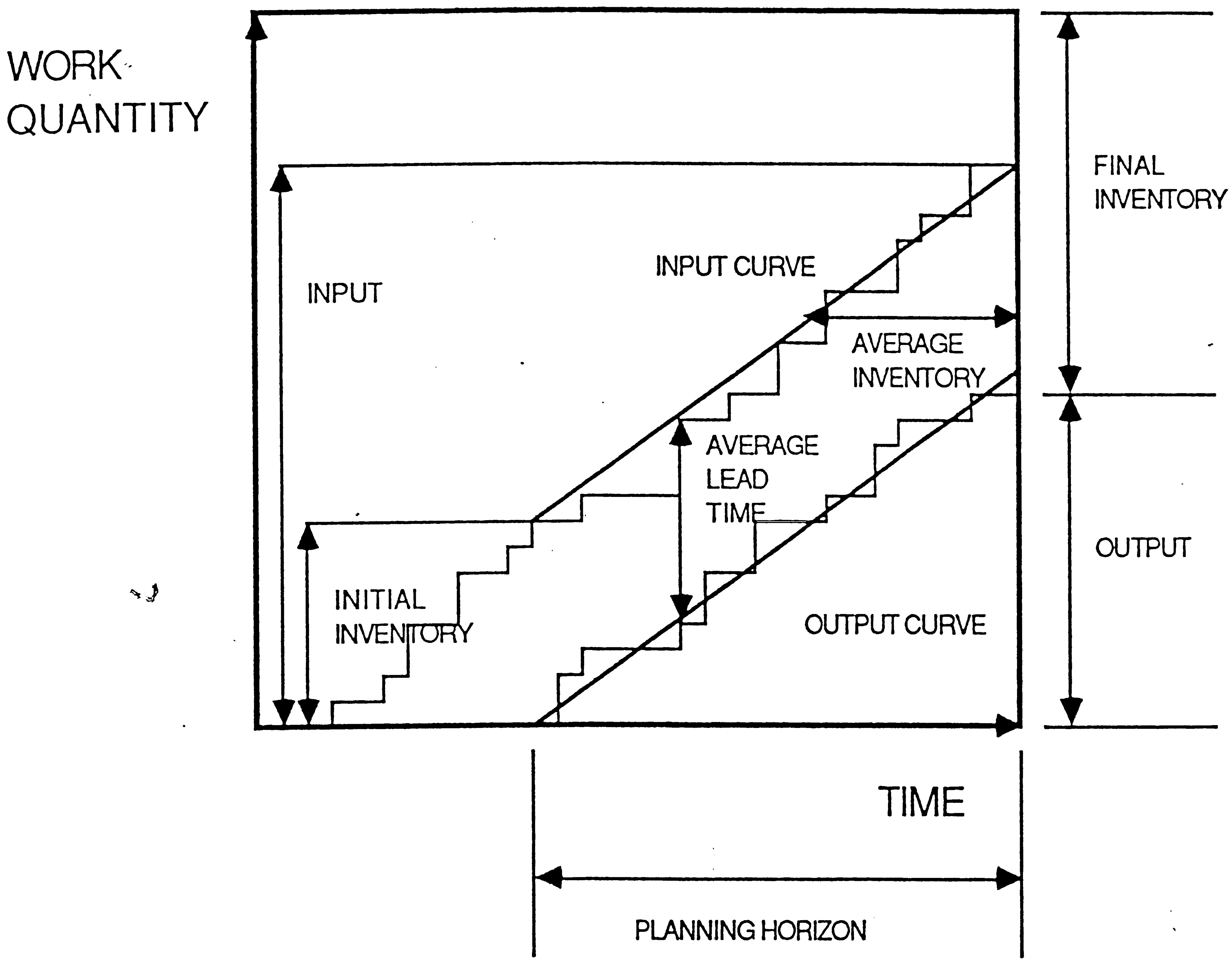

Figure 2: Time varying nature of input/output

.pa

input is computed by using the order arrival dates and using the batch size to represent a step change in the input curve. The time varying nature of work input and output to a workstation is depicted in figure 2.

The graph is drawn between the work value against chronological time. The output is obtained by adding up the processed work order quantities according to completion dates. Thus any vertical axis represents a point in time when an order is completed and the horizontal line is the time interval between the completion of orders. The planning horizon, a time period of time T has also been shown on the graph. At any point in time, the vertical distance between the input and output is equal to the instantaneous work-in-process for the particular workstation. Similarly, the horizontal distance represents the instantaneous lead time i.e. the time that the just completed job would have stayed at the work station. Slopes for inputs and outputs averaged over the period T can be drawn and hence the average lead time and work-in-process inferred from the graph. Also, the cumulative input, ending work-in-process, initial work-in-process, and cumulative output also indicated on the graph. The slope of the averaged input and output lines is equal to ratio of average work-in-process to average lead time.

An effective manufacturing planning and control system enables adequate control of manufacturing lead times and work-in-process while satisfying the needs of the customers with the right parts at the right time and with small lead times. In this situation, the state of the work-in-process/lead time/capacity graph will be as shown

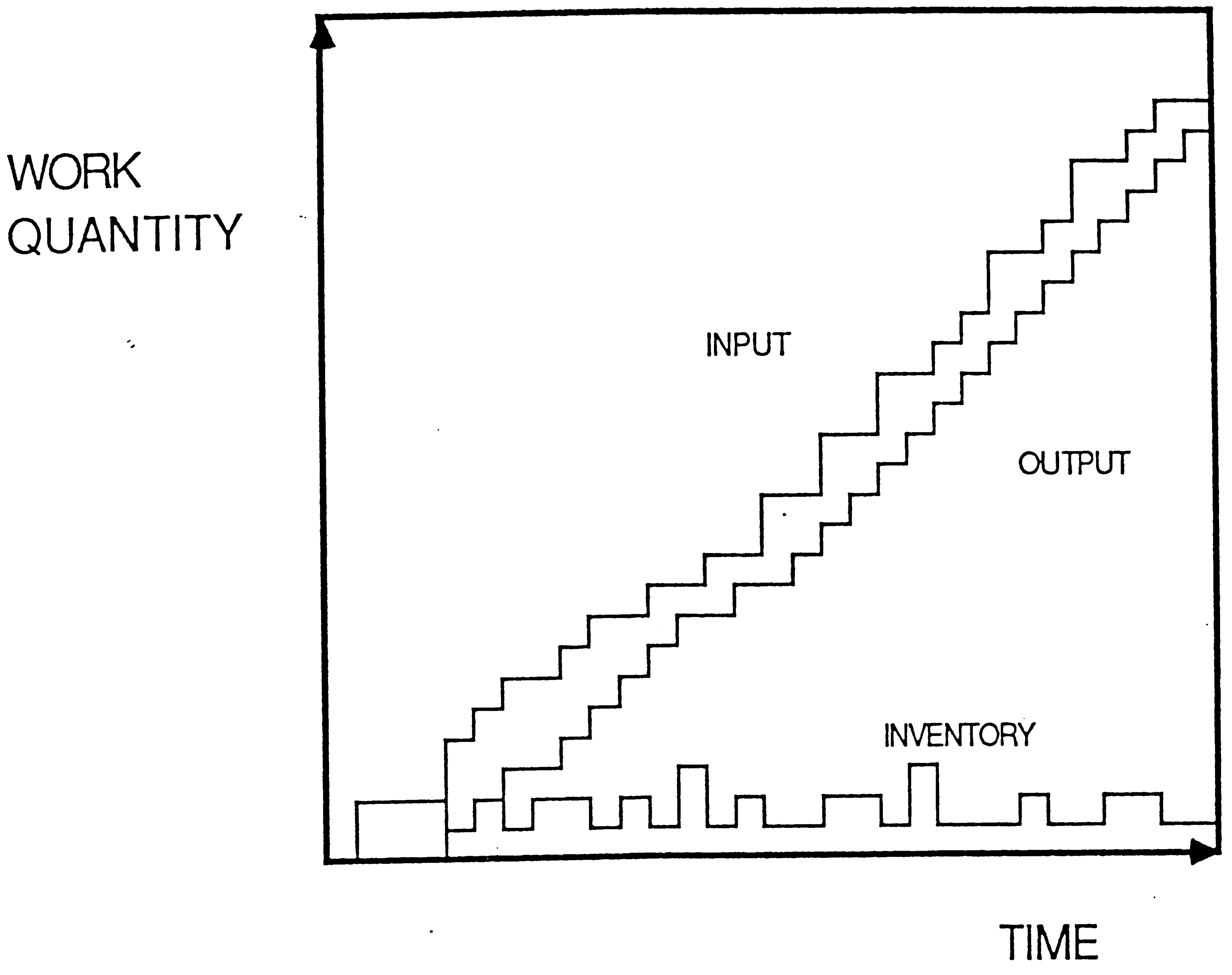

Figure 3: Input/output and inventory levels for well managed workstation

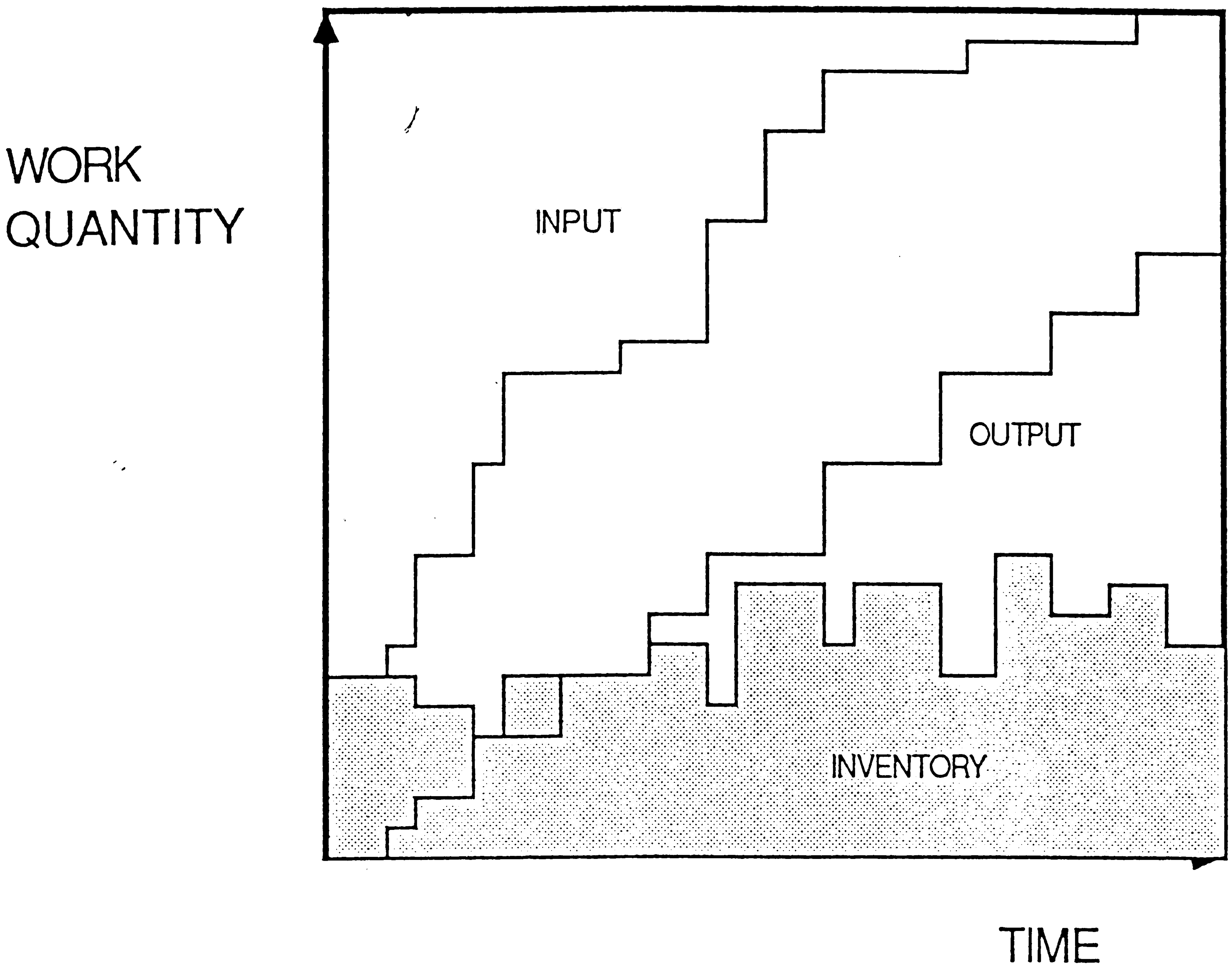

Figure 4: Input/output and inventory levels for not so well managed workstation

in figure 3. The input and output lines should be as close to each other as possible, thereby demonstrating proper monitoring of the input to the workstation and queue management. In this case the horizontal difference between the two lines i.e. is the work-in-process level is very small and under control. The finite level of work-in-process that is present serves to "protect" the workstation from uncertainties in its input. The figure 3 shows the case of a regulated and well controlled workstation whereas figure 4 shows the case of an uncontrolled workstation. In this case the input varies significantly with time and results in fluctuating work-in-process levels.

The above mentioned graphical method of considering tradeoffs between work-in-process/lead time/capacity can very effectively be used to observe, plan and manage the work input to a workstation. The "protection" can be designed into the work-in-process (i.e. the buffer) or in the designed lead time. Alternatively, the imposed lead time by the customer can be a driving factor and capacity computed based on it. Figure 5 is redrawn to be used as a planning tool displaying the relationships between planned values of lead time, work-in-process, output capacity and the input. The phrase "planned work-in-process" should not be misunderstood as unnecessarily designing of queues into a system, but just as an option which can be exercised to "protect" the system. The value of this "planned" work-in-

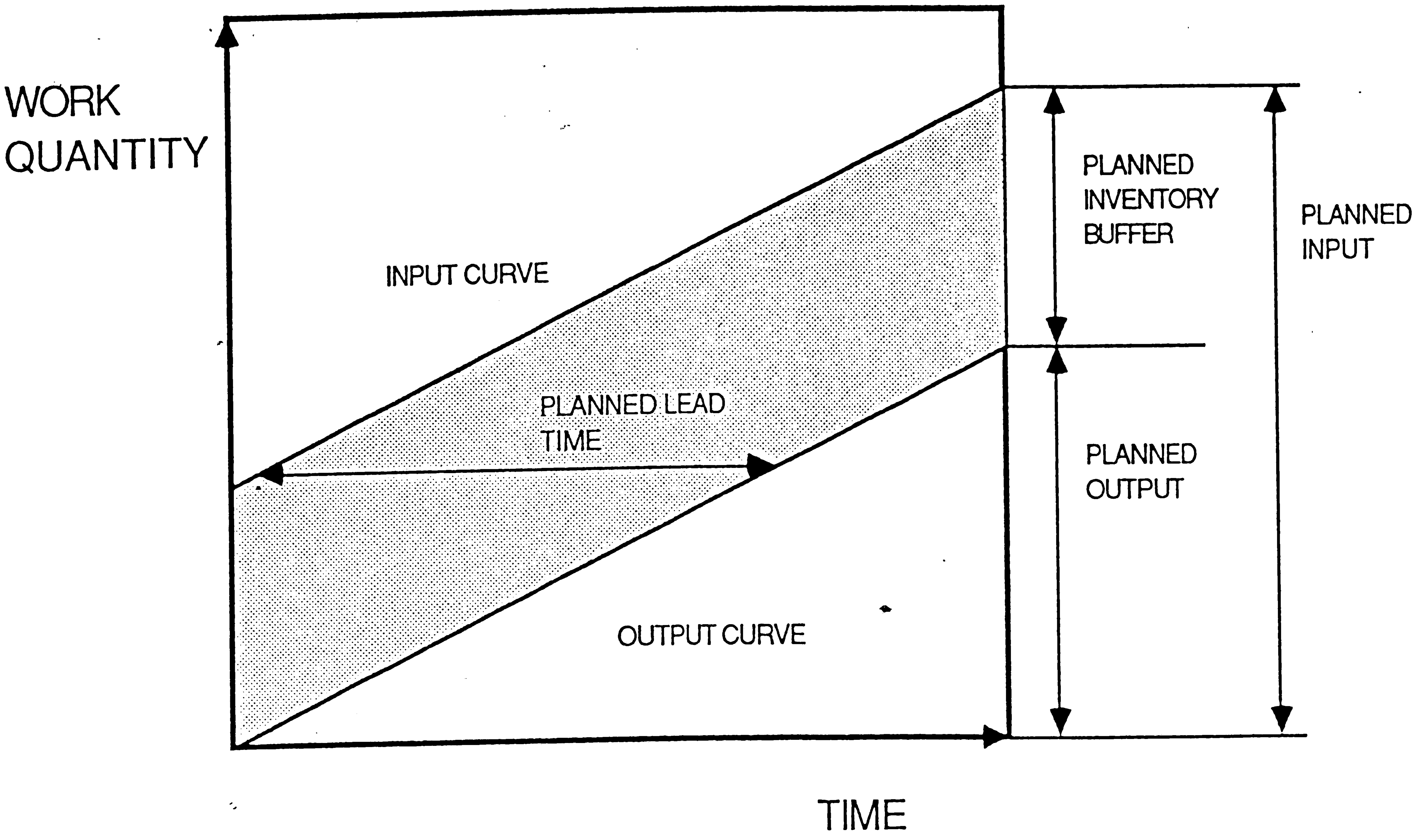

Figure 5: Relationships between planned values of lead time, input, output and inventory

process should be designed at the overall production system level and be a part of the strategy to build "protection" into the system. It will be discussed further in a later chapter. The tradeoffs and the relationships that have been discussed graphically will also be elaborated later using a quantitative model at a workstation level.

3.2 Tradeoffs between work-in-process, manufacturing lead time and capacity for a workstation - mathematical basis

The interdependencies and tradeoffs between work-in-process, manufacturing lead time and capacity will now be depicted using a mathematical rationale based on the capacity planning approach developed by Karni.[46] The approach in question was developed particularly to be used in an MRP framework and is a tool for performing the "capacity requirements planning" as proposed by the author and has its roots in the input/output control methodology.

The approach comes into use after the MRP routine has provided the material requirements plan based on the master production schedule, bill of materials, and "assumed" or "imposed" lead times. The approach uses the MRP time buckets and planning horizon, and imposed lead times to compute the workstation capacity required during the period to facilitate the achievement of the material requirement plan and hence the customer due dates. This capacity requirements planning methodology is a superior method as it focuses management attention on planned lead times and gives the ability to adjust them by computing adequate capacity or by adjusting the output level. Thus it can be planned to run a workstation at less than the full capacity and a rationale for designing "protective" capacity can be made on that basis. Another advantage is the explicit representation of work-in-process and it also provides a means to balance work-in-process with the lead time. The workstation can be operated in two modes: capacitated or uncapacitated. In the capacitated mode, the planned output does not fall below the workstation capacity at any time and in the uncapacitated mode, the underloading is allowed; this means that the planned output can go below the workstation capacity sometimes i.e. there is some extra capacity in the workstation.

The mathematical formulation which is the basis for the above mentioned approach can be effectively used to depict the relationship between work-in-process, lead time and capacity. The complete model which is a capacity requirements planning tool will not be repeated and the reader is encouraged to refer to Karni[46] to study the complete formulation.

3.2.1 Definition of variables

The following variables need to be defined to develop the relationships:

t = time period, with the planning horizon extending from $t=1$ to $t=T$

I_t = planned input to the workstation in period t

Q_t = planned queue at the workstation at the end of period t

W_t = planned work-in-process at the workstation during period t

$= Q_{t-1} + I_t$

L_t = planned lead time of the work at the workstation during period t

$= W_t/C$

U_t = planned underload at the workstation in period t, relative to the nominal workstation capacity (C); the planned output in this case will be $C - U_t$

F_t = fractional underload at the workstation in period t

R_t = cumulative input to the station through time periods 1 to t

$= Q_0 + \Sigma I_t$

Using the above time dependent values, we get the following average values:

C = the constant or nominal capacity of the workstation

Q_{av} = average planned queue

$= \Sigma Q_t/T$

W_{av} = average work-in-process

$= \Sigma W_t/T$

L_{av} = average planned lead time

$= \Sigma L_t/T = \Sigma W_t/CT$

U_{av} = average planned underload

$= \Sigma U_t/T$

F_{av} = average planned fractional underload

$= \Sigma F_t/T = \Sigma U_t/CT$

I_{time_av} = time weighted mean planned input

$= \Sigma(T+1-t)I_{It}/[T(T+1)/2]$

U_{time_av} = time weighted mean planned underload

$= \Sigma(T+1-t) I_{Ut}/[T(T+1)/2]$

The relations between work-in-process, lead time and capacity can now be developed. There are two distinct sets of relations based on the mode of operation of the workstation: capacitated or uncapacitated.

3.2.2 Relations for capacitated mode

The relations are:

Balances:

$W_t = Q_t + C$

Therefore we have:

$W_{av} = C + Q_{av}$

Capacity and work-in-process:

$C = [Q_0+(T+1) \cdot I_{time\_av}/2]/[L_{av}+(T-1)/2]$

where Q_0 is the initial existing queue length

Capacity and lead time:

$L_{av} = [Q_0+(T+1) \cdot I_{time\_av}/2-(T-1)C/2]/C$

Capacity and limits on work-in-process:

The management may decide to place a maximum limit (W_{max}) on the work-in-process level at any workstation. It may even place a minimum limit (W_{min}) so as to maintain the flow of work e.g. the size of kanbans.

W_{max} corresponds to C_{min} and is given by:

$C_{min} = \max(R_t-W_{max})/(t-1)$

W_{min} corresponds to C_{max} and is given by:

$C_{max} = \min(R_t-W_{min})/(t-1)$

Capacity and limits on lead time:

Similar to the limit on the work-in-process level, a maximum and minimum limit may be placed on the lead time. Thus we have:

L_{max} corresponds to C_{min} and is given by:

$C_{min} = max(R_t/(L_{max}+t-1)$

L_{minx} corresponds to C_{max} and is given by:

$C_{max} = min(R_t/(L_{min}+t-1)$

Exact capacitation:

The capacity can be increased to a point when the queue disappears and the workstation becomes exactly capacitated. Underload will occur if it is increased beyond that level. The value at which this will occur is:

$C_{cap} = min(R_t/t)$

Designing of "protective capacity":

Protective capacity can be inserted into the workstation in a particular period also. This is a very short term method and can be achieved by methods of overtime or an extra shift. But it does take care of the excessive input during a part of the planning horizon and keeps the work-in-process and lead time under control. Hence, this is a good method of deciding when to use overtime or extra shifts.

Let the fractional protective capacity be z. So the additional capacity is zC and let it be added in the time period T_{prot}. Thus the constant capacity C and average lead time will be:

$$C = [Q_0(T+1)I_{time\_av}/2]/[L_{av}+(T-1)/2+z(T-T_{prot})/T]$$

$$L_{av} = [Q_0(T+1)I_{time\_av}/2-(T-1)C/2-z(T-T_{prot})C/T]/C$$

3.2.3 Relations for uncapacitated mode

This is the case when the output of the workstation will be less than the workstation capacity in one or more of the periods. This implies some extra designed "protective" capacity in the workstation. The planned output in period t will be C - U_t. The relations will just be stated without the associated explanations which are same as in the case of capacitated operation. The relations are:

Balances:

$$W_t = Q_t+C-U_t$$

$$L_t = D_t+1-F_t$$

Therefore we have:

$$W_{av} = C+Q_{av}-U_{av}$$

$$L_{av} = 1+D_{av}-F_{av}$$

where D_t = planned delay of work in the planned queue at the end of period t

and D_{av} = average planned delay

$$= \Sigma D_t/T = \Sigma Q_t/CT$$

Capacity and work-in-process:

$$W_{av} = Q_0+(T+1)(I_{time\_av}+U_{time\_av})/2-U_{av}-(T-1)C/2$$

Capacity and lead time:

$$L_{av} = [Q_0+(T+1)(I_{time\_av}+U_{time\_av})/2-U_{av}-(T-1)C/2]/C$$

Capacity and limits on work-in-process:

$C_{min} = \max(R_t+\Sigma U_t-W_{max})/(t-1) \quad t>1$

$C_{max} = \min(R_t+\Sigma U_t-W_{min})/(t-1) \quad t>1$

Capacity and limits on lead time:

$C_{min} = \max(R_t+\Sigma U_t)/(L_{max}+t-1) \quad t>1$

$C_{max} = \min(R_t+\Sigma U_t)/(L_{min}+t-1) \quad t>1$

Capacity and limits on fractional underload:

$$\begin{aligned} F_t &= U_t/C \\ &= t-(R_t+\Sigma U_t)/C \\ &\leq F_{max} \end{aligned}$$

Hence, $C_{max} = \min(R_t+\Sigma U_t)/(t-F_{max})$

Capacity and limits on underload:

$$\begin{aligned} U_t &= tC-(Q_0+\Sigma I_t+\Sigma U_t) \\ &\leq U_{max} \end{aligned}$$

Hence, $C_{max} = \min\ [R_t+\Sigma U_t+U_{max})/t]$

3.2.4 Discussion of the tradeoff relations

The relations given above aptly model the tradeoffs involved between work-in-process, lead time and capacity. They also show how capacity for a workstation can be computed in case of imposed limits on lead times and work-in-process levels. The calculation of a constant value of "protective" capacity is also possible by underloading the station for the complete planning horizon. It is also to obtain "protective" capacity for a very short term i.e. in a particular time period by overtime or an extra shift to take care of some extra load in any time period. In this manner there is no need to build the constant extra capacity over the complete planning horizon and at an extra cost. Obviously, the decision is managerial and should be based on both quantitative and qualitative factors.

At workstation level, the relations together with the graphical depiction of the tradeoffs, provide a very powerful tool to achieve the correct levels of lead time, work-in-process and the required capacity in a period (throughput). It enables the achievement of planned (imposed) lead times as done by the MRP module of the manufacturing planning and control system. Thus protection in the system at the workstation level can be built by (i) adjusting the capacity for a short time period (ii) using "protective" capacity over the total period (iii) or if load exceeds the capacity, then feeding the computed and realistically achievable lead time back into the MRP system. This is necessary because the lead time, as it is clear by now, depends on the capacity of the workstation and the input load, and is not a constant value for a particular workstation or a particular product as assumed by the MRP technique. The relations also give the planner a feel of the system and allow him to weigh the tradeoffs such as protection by longer lead time or by "protective" capacity; or capacitated against uncapacitated workstations; or long lead times and high utilization against short lead times and low utilization of the workstations.

3.3 Capacity planning at the production system level

The previous two sections covered the tradeoffs between work-in-process, lead time and capacity and the issue of capacity planning at the workstation level. The discussion was within the framework of a manufacturing planning and control system where the capacity requirements and loads are an result of the master production schedule. The needed capacity and/or "protective" capacity was then computed based on the imposed lead times and generated requirements. It is obvious that such a technique is more suited at an operational level for day to day or period to period operations. The three distinct capacity planning methods (Johri[40], Solberg[72], and Sadowski[66]) which will now be briefly discussed are intended for long range capacity planning purposes and compute capacity at an aggregate level. Each of them have their advantages and limitations and assumptions and these will now be elaborated.

3.3.1 Capacity planning with finite and known buffer sizes (work-in-process levels)

This linear programming based model by Johri is very useful for capacity estimation and is very suited for continuous batch production. It models the system flows and hence takes care of the delays due to queuing which are a very significant part of any manufacturing lead time. Thus blocking and starving are considered. The maximum buffer sizes and the product mix can be specified and the author makes a point that the sequence of the mix can considerably affect the throughput of the line. This may result in loss of the nominal and also the "protective" capacity of the system which may have been designed in it.[40]

The model makes the following assumptions:

1. The process times are deterministic and average rates of machine breakdowns are considered.
2. Each workstation will produce one type of product at a time but there is no requirement that all workstations produce the same product at the same time.
3. The sequence of product mix is the same at all the workstations.

The nomenclature will now be defined:

s = number of workstations

m_i = number of machines at workstation i

a_i = availability of each machine at workstation i

b_i = max buffer size allowed between i and i+1
where $i = 1,2....s-1$

c = number of product types

n_j = batch size of product j

p_{ij} = processing time for a piece of product j at workstation i

s_{ij} = setup time for product j at workstation i

d_{ij} = duration of time required by station i to process the whole batch of product j

P_{ij} = average processing time for a piece of i at workstation i

$= p_{ij}/(a_i \cdot m_i)$

The initial step is a bottleneck analysis. If only one product type is to produced, then the bottleneck station is the one with longest average processing time. But, since this is rarely the case, the actual limiting or bottleneck station will depend on the product mix. Thus define T_i as the time required at workstation i to process the product mix in a production cycle. Hence, we have:

$T_i = \Sigma(n_j \cdot P_{ij} + st_{ij})$

$T_b = \max\{T_i\}$, then the workstation b is the aggregate bottleneck and T_b is the duration of the production cycle.

It should be understood that workstation b is the overall bottleneck over a long period of time and at an average level. The interactions between stations are always significant even in the case of very small systems thus the effect of blocking and starving have to be considered. Thus, for a given product mix, buffer sizes and sequence, the actual bottleneck workstation will be the one that takes the maximum amount of time (including the blocked/starved time) to process this mix. This means that the actual short term bottleneck is time dependent and may shift with time.

The model has been formulated as a linear program to compute the production capacity of the system. The detailed formulation will not be repeated and can be referred from Johri.[40] The objective function is to minimize T so as to obtain maximum throughput. The set of constraints have been classified by the author as production, input side, output side, and cycle constraints. The solution would give the values of d_{ij}, T and the binding constraints. The above values can be used to compute:

R = average flow rate of products

U_i = average utilization of workstation i

B_i = percentage of time that station i is blocked or starved

3.3.2 Capacity planning using a deterministic and stochastic workflow model

Solberg has developed a deterministic model and also an extended version of it i.e. the stochastic model to determine the production capacity of a system.[72] The detailed modeling and derivations will not be discussed and can be found in the above mentioned reference and a conceptual discussion will be done.

Solberg defines a production system as being composed of workstations and a transport mechanism. A mean transport time is assumed for the entire system and for all movements of units between the stations. It allows for the assigning of frequency of visits by a part to a workstation as to facilitate the modeling of inspections or rework. In the deterministic model (also called the bottleneck model by its author), the bottleneck workstation is defined to be the one with the maximum workload. This workload, as discussed in the previous model is very much dependent on the product mix. The system production rate is defined by this bottleneck and other parameters such as utilizations and mean number of busy servers.

The stochastic model is supposed to be superior to the deterministic one and accounts explicitly for variability in processing times. It also considers the queuing behavior of the workflow and computes true capacity as opposed to the deterministic model which overestimates it. Actually, the capacity determined by the deterministic model is the upper bound on the system capacity and this value would only be achieved with infinite work-in-process in the system. In such a situation, the bottleneck workstation will never be idle due to the variations in the processes. The dependence of throughput (actual utilized capacity) on the work-in-process level is shown in figure

6. It is interesting to compare this behavior of the throughput and lead time with work-in-process level (these are the tradeoffs at the production system level) with those at a workstation level described earlier.

The stochastic model suffers from some serious drawbacks in making some assumptions and thus is not a true representation of the reality of any production system. Firstly, it assumes the transport time as same for all the parts. This may be in serious error as parts just wait in some cases waiting to be transported and then some arbitrary or some informal priority rule is used to decide as to which parts need to be transported first. Also, there can be many other reasons to support arguments against the particular assumption.

Secondly, the model assumes a constant total work-in-process (N) in the system which implies that whenever a unit is done processing, it is immediately replaced by a new unit. This assumption is definitely questionable as the total work-in-process in the system will depend on a number of parameters the most important being the market forecasts and customer demands. Thus, the above assumption implicitly presupposes an order release system which may or may not reflect the reality. He also derives and plots the relationship of lead time as a function of work-in-process (figure 7). It is very interesting to note these relation

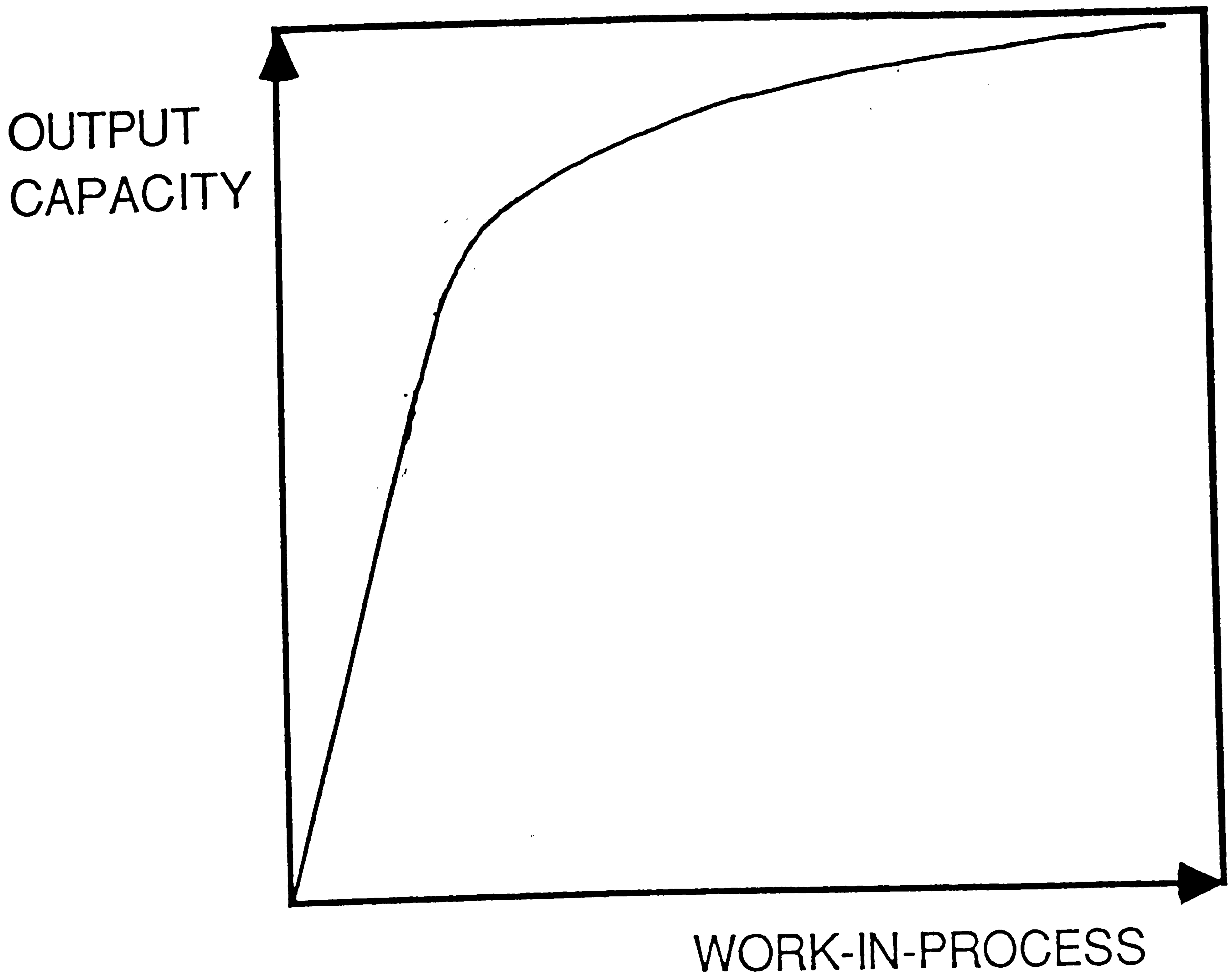

Figure 6: Dependence of throughput on work-in-process at the production system level

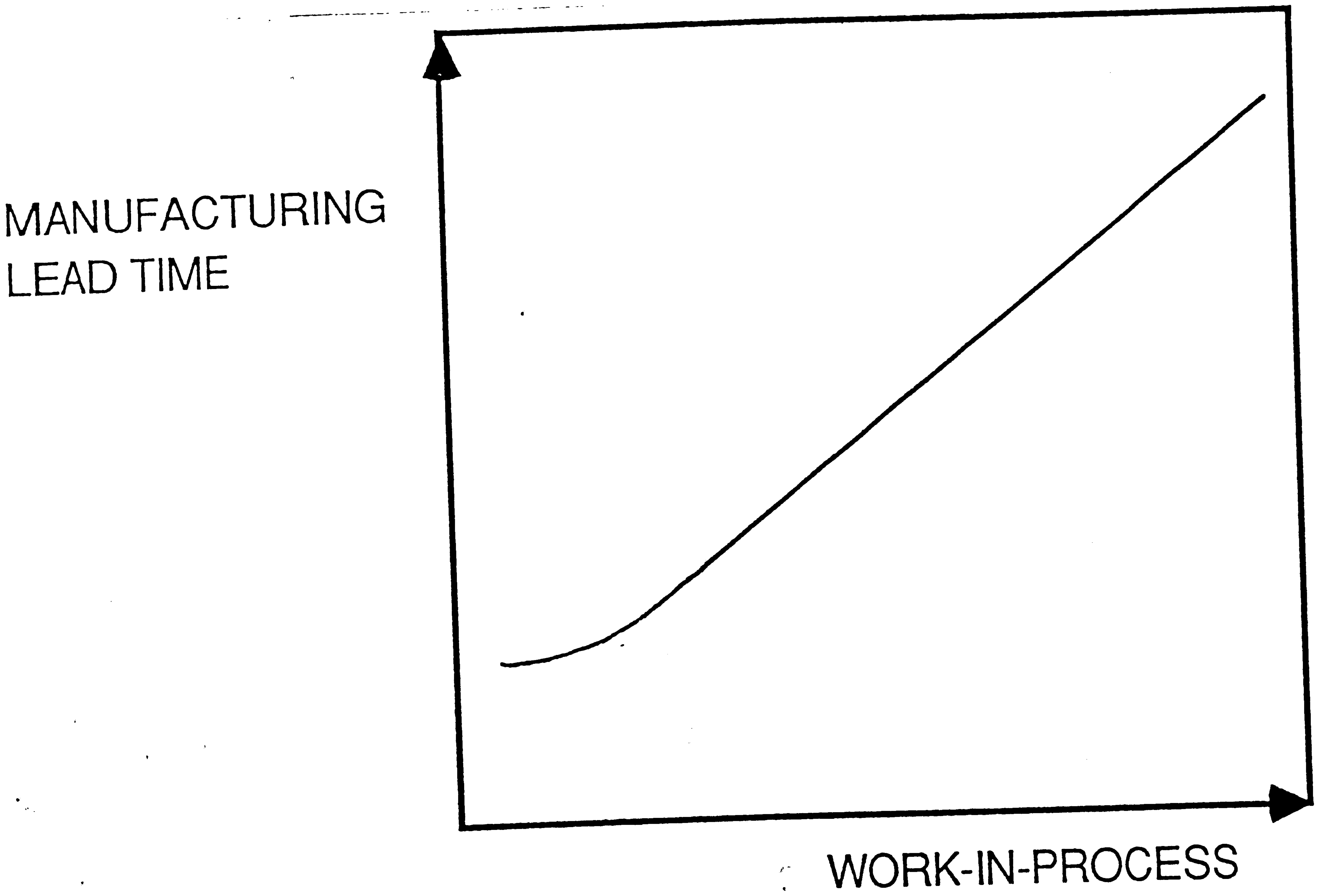

Figure 7: Dependence of lead time on work-in-process at the production system level

ships between work-in-process (N), throughput (capacity) and lead time in the figures 6 and 7. Hence, to follow the logic of this capacity planning method, one will first assume the level of work-in-process in the system to obtain the usable capacity and the lead time or vice versa. Another drawback in the model is that there is no consideration for the location of the work-in-process in the system. The planner has no idea or any control over the size and location of individual buffers so as to gain better control of the production system. The overall work-in-process level (N) also does not help in any way to further this purpose. The model, really, does not has any parameter or method to enable the design of "protective" capacity in the production system.

3.3.3 The constrained machine model

This model for determining the production system capacity developed by Sadowski uses the concept that the system capacity is constrained by and is dependent of a set of "critical" machine resources.[66] It is intended by the author to be used for two purposes: to provide an aggregate level technique to estimate the capability of the system to produce a given product mix in the specified time horizon; and to estimate the capacity of the production system.

The model is intended for manufacturing environments with well defined product lines. This implies the existence of:

1. known number of products
2. known number of machines
3. known process plans for each product with processing times

The flow of parts through the system is not modeled as compared to Johri's or Solberg's models thus eliminating the effect of queuing delays on the results. It loads the system at an aggregate level to determine the critical resources which govern the throughput. The methodology consists of three stages.

The first stage is the general formulation and computes the estimated fraction of time required on each workstation for the given production plan. The machine time for any product has been defined as the sum of processing, failure allowance and setup times. After the above computation of total machine times, the product mix is superimposed and workstation time requirements computed. A cutoff value of this machine time requirement is then used to identify the potential "congested" machines with high utilizations. If the product mix does not vary too much beyond the values used in the previous stage, the computations uptill this stage need not be done again.

The final stage of this model computes the "shop load limit" which is the maximum effective capacity. This helps in adjusting the actual production level to be in agreement with the effective production capacity. The adjustments may be in the actual production levels of one or more products. This stage is an iterative method and is repeated until the desired effective capacity level are achieved.

The model, as stated by the author, is not intended to replace the standard machine loading techniques that are used in conjunction with MRP systems. However, if the master production schedule (mps) seems to be overloading the system, then this methodology can aid in mps modification more realistically manner. It is particularly useful as a strategic planning tool to evaluate long term expansion strategies in order to respond to changing customer demands. It is also useful to evaluate the gain in production system capacity obtained by new machines or better methods and processes. A major drawback of the methodology is that it does not consider product flows and thus ignores phenomena such as conflicts, starving and blocking or problems due to poor scheduling and shop floor control. The factors have a very substantial effect on the effective capacity of a production system. Another limitation is that it allows changing of production volumes of different products so as to change the shop load limit to find a feasible solution. From the market point of view, it is not advisable to do that; a better way to achieve should be to change actual workstation capacities to produce what the market wants; or better still to build enough "protective" capacity so as to dampen the effect of such variations. A good method would thus, compute the required capacities for workstations and the system capacities to achieve the production goals and also determine the "protection" needed to handle the variations in customer demands.

CHAPTER IV

A FRAMEWORK FOR DESIGNING "PROTECTIVE" CAPACITY

The objective of this chapter is to present a conceptual framework for capacity planning & management with emphasis on designing/building "protective" capacity in a production system. This framework is based on current and existing methodologies and algorithms which have been suggested by their authors for planning, estimating and managing capacity. None of these methodologies and algorithms have suggested the use of or the designing of "protective" capacity in a production system. The framework in fact makes use of the relations and expressions developed in these methodologies to suggest ways to build and design for "protective" capacity. The methodologies discussed in the framework are a subset of those that were discussed in the previous chapter. After briefly discussing the framework in a general manner, the description will be based on a series of questions pertaining to different aspects of designing/building and managing "protective" capacity in a production system.

4.1 Description of the framework

A graphical description and mathematical model of the tradeoffs involved between work-in-process, lead time and capacity has been presented, in the previous chapter, along with summarized descriptions of three methodologies for capacity planning. The descriptions also contained the advantages and limitations of these methodologies. Based on the above, the author suggests the use of a combination of the "capacity estimation model" by Johri and the trade-off models, both graphical and quantitative, as a framework for designing "protective" capacity. A flowchart depicting the basic features is shown in figure 8 and it shows the input data requirements, scope/level of planning, and associated factors which influence the planning process or are affected by it are also shown.

The "capacity estimation model" by Johri (as it will be referred to in the future discussion) is intended to serve as a long range planning tool. Given a configuration of equipment, desired buffer sizes and product mix, the capacity can be computed using the linear programming formulation. A very important byproduct of this procedure is the information about the "lost" capacity at each workstation including the cause i.e. blocking or starving. This is very valuable information at the design stage of a production system and will lead to actions to solve these problems at this stage. Thus, at this stage "protective" capacity can be designed at the trouble spots in the system. To achieve that, the actions may be one of the following:

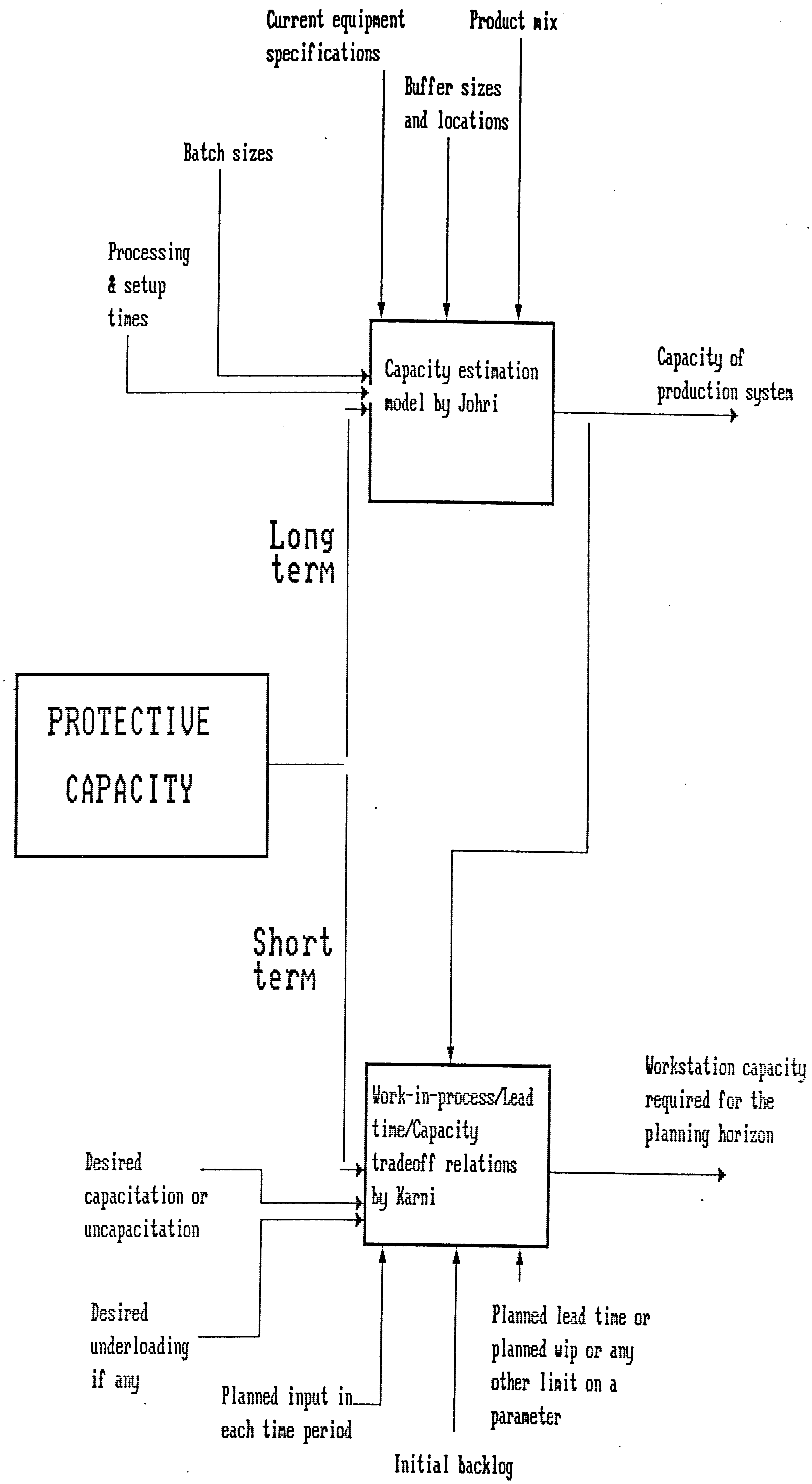

Figure 8: The framework flowchart

* Change the designed work-in-process level by

  (1) modifying the size and location of the buffers;

  (2) changing the batch sizes

* Altering the product mix or the sequence of their production

* Adding extra equipment

It should be noted the above actions may be taken individually or in some combination of each other. Thus, if no more improvement is possible by buffer size modification and short term capacity problems still exist regularly, then extra equipment would be the only method. The tradeoffs, although not very explicit in the formulation, are an inherent part of the model. The effects of increasing the planned work-in-process level will definitely be felt on the capacity which will increase with the resulting increase in lead time and decrease in flexibility. As these changes are executed in the model, the trade-off consequences will definitely be observed in the results to aid in making the right decision.

This "capacity estimation model" should serve as a design/planning tool whose output will be the input to the short term "protective capacity" managing tool i.e. the graphical depictions of work-in-process/lead time/capacity tradeoffs and the quantitative relations by Karni. A well designed production system would have taken care of building enough "protective" capacity to take care of most variations. But, as is always the case, even though the design is done keeping in mind as many possible future variations, some changes that have not been anticipated will occur with the resulting need to protect the system from them. This protection at the short term level is done by taking actions at the workstation level. This is to take care of capacity inconsistencies within a time period in a planning horizon. The trade-off relations will be useful in analyzing different alternatives such overtime in one period, or a constant second shift for the entire planning horizon, or even in demonstrating the unfeasibility of any of the actions. This may mean designing more long term "protective" capacity into the system by going back to the first part of the framework.

The inputs to the framework at this stage are the planned work input in each period of the planning horizon, initial backlog and one more design parameter. This parameter could be one of the following: maximum or imposed lead time by the MRP or some other manufacturing planning and control system; limits on work-in-process or designed work-in-process at the workstation; designed inserted capacity i.e. short term "protective" capacity in a particular period; and the like. Uncapacitated operation of a workstation (i.e. underloading) in a part of the planning horizon is also a possible alternative and actually results in "protective" capacity in the entire planning horizon.

4.2 Questions and issues relevant to "protective" capacity

The various issues related to the planning, design and management of "protective" capacity will now be discussed with reference to the framework described in the previous section.

4.2.1 How much "protective" capacity does one need? How much of it should be long term and how much short term?

The type and amount of capacity depends on the type of business one is into which also affects the amount of variability in the market; thus the need for the type of "protective" capacity. In a process focused production system, the nature of demand is intermittent and the flow is also, as a result, very intermittent. This causes irregular loading of the workstations.[11] In contrast to the above is the product focused production system where highly standardized products result in continuous use of workstations and better product flow.

Thus, it seems that the process focused system will need more long term "protective" capacity due to the nature of the business and the type of equipment that is needed. This has to be taken care of in the long term planning by using the "capacity estimation model". It has been shown in reference [43] that there is great value to maintaining "protective" capacity where there is uncertainty in demand and too little "protective" capacity may result in bottlenecks showing up with ever growing queues. The important factors in this case will be selecting initial equipment, batch sizes, setup time considerations, and the sequence in which the product mix is fed to the system. Due to nature of the business (more product variety with lesser volumes), the benefit of computing and placing accurate buffer sizes may not be very helpful in designing the system. The short term peaks and valleys in the market demand can be accommodated by inserting "protective" capacity using the additional protective capacity factor (z) in the period T_{prot}.

In a <u>product</u> <u>focused</u> <u>system</u> key to long term, the key to designing "protective" capacity is through proper balancing the flow. The disastrous results of any effort to design and build a totally balanced plant have already been stated. It must also be noted that more protection in this type of business is needed, not against rapidly changing market demand, but against variations internal to the organization such as breakdowns, absenteeism and the like. Thus, an effective way to design "protective" capacity is through the suitable way of computing the correct sizes and then placing them at the correct locations. The method of designing short term "protective" capacity would be to design a continuous underload in the workstation for a planning horizon by using the fractional underload desired or an underload limit. This limit may actually be a physical limit on the system due to overtime policies or second shift policies or even a technological limit.

4.2.2 Relationship of "protective" capacity to flexibility and the ability of the system to react to changes

The significance of increased flexibility has been emphasized in an earlier section. It is not only important in the case of customized products market, but also in the other more mass production markets. In such cases the ability to quickly adapt to new and changing markets with products at the least cost and effort guarantees increased market share and survival. Apart from the necessary organizational changes, the equipment and machinery i.e. the production system should also be able to adapt to it. This may mean proper design of "protective" capacity in the system which in turn requires correct prediction of anticipated changes. Technological changes may require complete change of equipment altogether, but, production volume requirement changes will require adequate protection. Karmarkar [43] defines flexibility as the ability to adapt to changing conditions and has classified these changes as exogenous and endogenous:

1. Exogenous changes are changes in demand or output. These require flexibility with respect to

production capacity to take care of seasonal variations, volume changes in the product mix and addition of new products.

2. Endogenous changes are changes in the characteristics or the abilities in the process itself. These include engineering changes and new technologies.

Both endogenous and exogenous changes have to be taken care of by designing long term and short term "protective" capacity. The ability to respond to <u>seasonal variations, new product introduction</u> and also, may be, <u>new technologies</u> should be accounted for in the long term "protective" capacity design using the "capacity estimation model" of the framework. The model allows explicit inputs for defining the number of end products in the mix and also for defining their sequence. A proper design should consider all possible combinations and extensions of this product mix along with the varying sequences to compute the effective production capacity of the system. This will reveal the moving bottlenecks, if there will be any, and adequate buffer protection could then be designed into the system. The effect of these actions can iteratively be checked by recomputing the capacity. This method will be sufficient to take care of varying loads due to seasonal variations and technologically similar new products. To account for flexibility for new technologies, the decision is very subjective and based on an understanding of the technology involved and the market. This decision may involve tradeoffs between general purpose and special purpose machines and will depend on the type of business i.e. processed focused as compared to product focused.

The flexibility in a production system with respect to <u>product mix volume changes and engineering changes</u> have to be taken care of by providing adequate short term "protective" capacity. The product mix volume changes will cause changes in the input loading for the particular planning horizon and this in turn causes problems to surface at the bottleneck resources or even in the shifting of the bottleneck and increasing queues. The effect of these changes can be analyzed using the graphical tradeoffs method and the relations between work-in-process, lead time and capacity. The solution would be to insert "protective" capacity in the heavily loaded time periods or to even the output load by using constant extra capacity throughout the planning horizon. If these variations are persistent and regular, then it would be advisable to design constant underloading at the trouble spots so as to minimize the constant monitoring and changing the capacity. Again, the level of the fractional underloading will depend on management policies and technological limits of the equipment. The engineering changes in any product can affect its routing and/or only one or many workstations. The appropriate action for this would be to insert "protective" capacity in these workstations using the relevant relations or if more than one workstation is affected, the set of actions described for the case of product mix volume changes would be necessary.

4.2.3 How to convert a suitable work-in-process protection to "protective" capacity?

The role of work-in-process is to decouple two stages or two workstations of a production system so as to allow some degree of independence to them. Without this work-in-process, the workstations will have to be perfectly synchronized and balanced to operate effectively. The disastrous results of any effort to design and build a totally balanced plant have already been stated. Thus the omnipresent need and desire to protect workstations by work-in-process or "buffers". This protection prevents capacity loss due to blocking and starving and also due to other uncertainties such as breakdowns, etc. Another important use of work-in-process is that it allows two neighboring workstations to work on different products by taking care of the needed setup time for the changeover. On the contrary the costs of work-in-process are in the carrying costs, and on its influence on manufacturing lead time with the associated harmful effects.

The need to minimize work-in-process is very important; what is really needed is the information about how much and where to place the minimal work-in-process that we want. Wip at some locations, even in small quantities, is very useful and at most other locations is highly counterproductive. But, the most important issue that must be understood in the use of work-in-process protection is that, it should only be used as a protection against short term uncertainties and not against long term capacity problems. Thus, it seems that a minimal work-in-process at strategic points coupled with greater "protective" capacity would be the ideal action to take in order to keep lead times low and preserve system flexibility.

Following from the above discussion, for the protective buffers that have traditionally been used, we need to find the equivalent value of "protective" capacity for both long term and short term purposes. In the case of designing <u>long term</u> "protective" capacity which is done at the design/planning stage of a production system, the "capacity estimation model" is to be used. Given a set of equipment specifications and other inputs, the production system capacity can be computed. Now, this system will probably have buffers at each workstation; definitely in front of the bottleneck workstation if the design is proper. The buffers in front of other stations are not really needed and it may in fact be advisable to convert them to "protective" capacity. The tradeoff relations will be very useful for this purpose; assume the input per period based on the forecasted requirements and product mix which was used for long term planning; then using limits on maximum work-in-process levels, compute the constant workstation capacity or even run the workstation in the uncapacitated mode. The results may insert a lot of so called "idle or extra" capacity into the system but with lot less work-in-process/buffers. This would result in much shorter lead times, the same protection against breakdowns and a much more responsive system even though the system is much more unbalanced now due to these capacity additions. An important prerequisite is that the buffer sizes and locations were computed to obtain the maximum capacity out of the system using some algorithm or heuristic.

For the shorter time horizon, the problem becomes that of keeping queues in front of the workstations in control and of limited size. These queue sizes will be dictated by the overall system design. Thus we have a maximum limit on queue size or work-in-process size (W_{max}) and this can be very easily converted into equivalent capacity by the tradeoff relations. Alternatively, "protective" capacity can be inserted into a particular period; or non bottleneck stations can be made to operate in uncapacitated mode throughout the planning horizon.

4.2.4 Relationship of designing/planning "protective" capacity to forecasting

The activity of capacity planning is tightly coupled to the forecasting of manufacturing requirements. This is evident from the various capacity planning methods available to be used at varying levels of aggregation and planning horizons. The long range plan may extend a year or more into the future and gross aggregated requirements are computed using the aggregated forecasts of groups of end products. The short range forecasting generates more accurate forecasts and is used for preparing the master production schedule. The extent and length of planning horizons will also depend significantly on the type of business. Considering the previously mentioned classification, the process focused organization may offer a lot more customer options and these may be specified very late in the ordering process; on the contrary, the product focused organization may provide very standardized products with little or no customer options leading to much more reliable and earlier frozen forecasts. This issue has very substantial effect on the capacity planning picture which is directly based on the material requirements for the end products.

In the first type of production system, due to the nature of the business, an organization may have a relatively shorter stable planning horizon and variation in forecasts; Obviously, the prediction task is much more difficult. The risk of making the wrong predictions is also higher due to cost of the possible lost sales. Thus in this case the need for "protective" capacity is high. More "protective" capacity should be inserted into the system at the planning stage with adequate provision for short term "protective" capacity management also (overtime, extra shift, subcontracting etc.).

In the second type of production system, the planner is able to get a dependable production plan for a much longer horizon and also there will be lesser variation in the demand. This implies that if there was adequate knowledge about this not too much varying demand, the system does not need too much long term "protective" capacity. If at all there are some variations which are always possible, the short term "protective" capacity managing methods (inserted capacity, underloading etc.) will be sufficient to take care of them. The long range capacity planning exercise i.e. the design of production system should go on simultaneously so as take care of demand increases and prevent use of short term actions. The relevant issue is how and when to provide the capacity increases i.e. identify the size and timing of projected capacity gaps. A good indication is the extent of use of short term methods to protect the system and provide the needed capacity to meet production targets. But it should also be realized that inadequate short term "protective" capacity may result in lost sales and other harmful effects of utilizing a system at its full capacity such as shift premiums, productivity losses. In this way, the capacity can be added in proper time phased increments through planned use of short term and alternate sources of capacity.

4.2.5 Relationship of "protective" capacity to manufacturing lead time

Inflated lead times have traditionally been used as a form of protection against the uncertainties and "arbitrary" customer orders. The importance of shorter lead times and their reduced variability and the effect on the bottom line measurements of any manufacturing organization have been described in a previous section. Thus, considering the tradeoffs, this implies a low ratio of work-in-process to throughput (current utilization of capacity). The long lead time is a result of a combination of large queues in front of workstations, uneven input to the system and mismanagement of the above situation. We very well know the effect that, lead time inflation for protection of the system, has on work-in-process and the resulting customer lead time; it results in the vicious cycle of constantly increasing lead times as management tries to protect themselves by this method. The situation manifests itself in two situations; in the customer lead times and in the purchasing relationship with the vendor.

A small bottleneck in the production system can be the cause of this lead time inflation and thus has to be well protected. Also, the unnecessary buffers at other places will have to be eliminated to decrease the flow time of the product and replaced by "protective" capacity. This has the dual effect of reducing the lead times and also giving the system the capacity to react to changing customer demands much faster. The reduced lead time also has a beneficial effect on the forecast accuracy and its time horizon; the master production schedule will consist of more firm orders and less production plan. The reduced lead time will lead to fewer forecast errors and lesser needed protection by work-in-process or still longer quoted lead times. If at all there are some modifications in the actual customer orders on some occasions, the "protective" capacity can help take care of it.

CHAPTER V

SUMMARY AND CONCLUSIONS

5.1 Summary

The long term strategy of a company must give due consideration to manufacturing issues because they affect the competitive position of the company. Capacity related decisions are an integral part of these issues and hence the need for effective capacity planning - both long term and short term. Based on the market and the manufacturing system and the uncertainties inherent in them, the planning must be done to provide the needed or nominal capacity as well as adequate "protective" capacity. Traditionally, organizational slacks have existed in all forms, particularly as excessive inventory and long quoted lead times with the intended purpose being to protect the system; but they have caused more harm than good. Thus, a case has been made for a better form of protection through "protective" capacity.

Any manufacturing plant cannot be completely capacity balanced due to the variations in the processing times and will have one or a very small set of bottleneck resources which must be protected so as not to lose any system production capacity. Also, traditional cost accounting systems have emphasized high workstation capacity utilizations which lead to deceptive protection by excessive work-in-process and long & highly variable lead time values. In fact, management decreed increases in planned lead times to clear the backlog have also led to a vicious cycle of constantly increasing lead times negating all protection effects that were originally intended. This is specially important in light of today's business environment where flexibility and reduced order lead times are very important. Actually, work-in-process has been used as a protection much more than long lead times and the reasons are many and varied; psychological, decoupling two workstations, keeping workers busy, EOQ batch sizes, and the like. Due to strong dependency of lead time on work-in-process, the effects of such large work-in-process levels can be very harmful; hence the need to properly control it.

There exist strong interdependencies between work-in-process, lead time and capacity at the workstation and at the production system level. The tradeoffs at the workstation level have been depicted through a graphical approach and through relations developed by Karni for his "capacity requirements planning" model. The graphical approach is very useful to keep track of and to observe the workstation behavior (input load and output capacity) for a planning horizon. The relations can be used to compute the needed workstation capacity given the input, initial backlog and any other limiting design parameter. Short term protective capacity design can be done by underloading, inserting extra capacity in only one period, deciding between capacitated or uncapacitated operation etc. using the relations. At the production system level, three capacity planning/ estimation models are discussed briefly. The tradeoffs are valid at an average level allowing "protective" capacity to be designed in the production system at the system design/ planning stage.

A framework based on the tradeoffs at the workstation level and the "capacity estimation model" by Johri has been suggested for designing and planning "protective" capacity. The tradeoff relations at the workstation level are used for the planning "protective" capacity for short term purposes i.e. for one or more time periods within a planning horizon. The "capacity estimation model" is used for long term "protective" capacity design at the planning stage. The model, during the design phase also provides information about the aggregate bottlenecks which depend on the product mix and other factors enabling adequate protection to be designed to preserve system's production capacity.

The use of framework has been discussed with reference to a set of questions and issues relevant to a production system. The type and amount of "protective" capacity that is needed depends on the kind of business that one is into (product focused or process focused) and some technological limits that may exist. Protective capacity is also sensitive to product mix and volume changes; thus flexibility and response of the system can be modified by effective use of "protective capacity". The framework can be used to convert a work-in-process protection (which may be harmful in many other ways) to a capacity protection. The amount of "protective" capacity in a system affects the forecasting accuracy and time horizon; the extent of this effect, again, depends on the nature of market and business.

5.2 Conclusions

The importance of capacity related decisions and both long and short range capacity planning is undisputable to be able to achieve a sustainable competitive advantage. To take care of the uncertainties and variations inherent in the market, the manufacturing system and its environment, "protection" must be built into the system. Instead of building this protection through excessive slacks in all parts of the organization, particularly through excessive inventory or long lead times, "protective" capacity should be the preferred method. Apart from providing similar protection, it does not has the associated drawbacks of high work-in-process inventory or long lead times.

There exist very strong interdependencies between work-in-process, lead time and capacity. These are the basis for the tradeoffs between the above factors, both at the workstation and system level. The framework that is based on the tradeoffs, will enable the management to properly observe & monitor the workstation and system behavior and also effectively design/plan and manage "protective" capacity. Adequate design of long term "protective" capacity and its suitable management would thus ensure satisfactory performance from the manufacturing organization.

BIBLIOGRAPHY

[1]. APICS. Capacity Planning and Control. APICS. 1975.

[2]. Andreas, Lloyd, "40 Days to the due date: by Willy Makeit: The Role of Capacity Management in MRP-II" Capacity Management Reprints - Revised Edition, p. 1.

[3]. Bean, James C., and Smith, Robert, L., "Optimal Capacity Expansion over an Infinite Horizon" Management Science, Vol. 31, No. 12, December 1985, p. 1523.

[4]. Bechte, Wolfgang, "Controlling Manufacturing Lead Time and Work-in-process Inventory by means of Load-oriented Order Release" APICS 1982 Conference Proceedings, p. 67.

[5]. Behling, Richard L., "Supply Chain Management with Capacity Constraints" APICS 1982 Conference Proceedings, p. 67.

[6]. Bellofatto, William R. "Lead time vs. The Production Control System" Production & Inventory Management, 2nd Quarter, 1974, p. 14.

[7]. Belt, Bill, "Integrating Capacity Planning and Capacity Control" Production & Inventory Management, 1st Quarter, 1976, p. 6.

[8]. Belt, Bill, "Input/Output Planning Illustrated" Production & Inventory Management, 2nd Quarter, 1978, p. 11.

[9]. Belt, Bill, "The New ABC's of Lead-time Management" Production & Inventory Management, 2nd Quarter, 1974, p. 80.

[10]. Bender, Paul S., Northrup, William D., and Shapiro, Jeremy F., "Practical Modeling for Resource Management" Harvard Business Review, March-April 1981, p. 163.

[11]. Buffa, Elwood S. Meeting the Competitive Challenge: Manufacturing Strategy for U.S. Companies. Dow Jones Irwin, 1984.

[12]. Carter, Phillip, and Ho, Chawan-jyh, "Vendor Capacity Planning: An Approach to Vendor Scheduling" Production & Inventory Management, 4th Quarter, 1984, p. 14.

[13]. Cheng, T. C. E., "A Simulation study of MRP capacity planning with uncertain operation times" Int. J. Prod. Res., 1987, Vol. 25, No. 2., p. 245.

[14]. Chong, Philip S., "System Combines Manufacturing Resource Planning and Decision Support Approach" Industrial Engineering, October 1985, p. 72.

[15]. Clark, James, T., "Capacity Management - Part Two" APICS 1980 Conference Proceedings, p. 335.

[16]. Cross, Kelvin F. Manufacturing Planning: Key to Improving Industrial Productivity. New York. Marcel Dekker, Inc. 1986.

[17]. Davis, K. Roscoe, and Taylor, Bernard W., "Want to Reduce WIP Inventory? Try Linear Line Balancing" Production & Inventory Management, 2nd Quarter, 1974, p. 56.

[18]. Fox, Robert E., "MRP, Kanban, and OPT - Whats best?" APICS 1982 Conference Proceedings, p. 482.

[19]. Fox, Robert E., "OPT vs. MRP: Thoughtware vs. Software" APICS 1983 Conference Proceedings, p. 556.

[20]. Friedenfields, John. Capacity Expansion: Analysis of Simple Models with Applications. New York. Elsevier North Holland, Inc., 1981.

[21]. Fry, Timothy D., and Blackstone, John H. Jr., "Planning for Idle Time: A Rationale for Underutilization of Capacity" Int. J. Prod. Res., 1988, Vol. 26, No. 12, p. 1853.

[22]. Goldratt, Eliyahu, "Computerized Shop Floor Scheduling" Int. J. Prod. Res., 1988, Vol. 26, No. 3, p. 433.

[23]. Goldratt, Eliyahu, "100% Data Accuracy - Need or Myth?" APICS 1983 Conference Proceedings, p. 64.

[24]. Goldratt, Eliyahu, "Cost Accounting: The Number one Enemy of Productivity" APICS 1983 Conference Proceedings, p. 433.

[25]. Goldratt, Eliyahu, "Optimized Production Time Table: A Revolutionary Program for the Industry" APICS 1980 Conference Proceedings, p. 172.

[26]. Goldratt, Eliyahu, The Theory of Constraints Journal. Volume 1. Avraham Y. Goldratt Institute, 1987.

[27]. Goldratt, Eliyahu, "The Unbalanced Plant" APICS 1981 Conference Proceedings, p. 195.

[28]. Goldratt, Eliyahu. The Goal. North River Press, 1986.

[29]. Goldratt, Eliyahu. The Race. North River Press, 1986.

[30]. Grunwald, H., Striekwold, P. E. T., and Weeda, P. J., "A Framework for Quantitative Comparison of Production Control Concepts" Int. J. Prod. Res., 1989, Vol. 27, No. 2, p. 281.

[31]. Gulati, Vince K., "Hidden Culprit in the Inventory Heap" APICS 1983 Conference Proceedings, p. 279.

[32]. Gray, James C., and Ehlers, Merle L., "The Business Resources Planning System" AT&T Technical Journal, Sept.-Oct. 1987, Vol. 66, Issue 5, p. 49.

[33]. Gutzman, Kurt M., and Wysk, Richard A., "Capacity Control Policies for MRP Systems" Int. J. Prod. Res., 1986, Vol. 24, No. 2, p. 359.

[34]. Heard, Ed., and Plossl, George, "Lead Times Revisited" Production & Inventory Management, 3rd Quarter, 1984, p. 32.

[35]. Hiller, Randall S., and Shapiro, Jeremy S., "Optimal Capacity Expansion when there are Learning Effects" Management Science, Vol. 32, No. 9, Sept. 1986, p. 1153.

[36]. Hinomoto, Hirohide, "Capacity Expansion with facilities under Technological Improvement" Management Science, Vol. 11, No. 5, March 1965, p. 581.

[37]. Hosni, Yasser A., and Alsebaise, Ali, "Capacity Planning in Job-Shop Environment" Computers Ind. Engng., Vol. 13, Nos. 1-4, 1987, p. 96.

[38]. Hoyt, Jack, "Dynamic Lead times that Fit Today's Dynamic Planning (Q.U.O.A.T. Lead Times)" Production & Inventory Management, 1st Quarter, 1978, p. 63.

[39]. Jacobs, F. Robert, "The OPT Scheduling System: A Review of a New Production System" Production & Inventory Management, 3rd Quarter, 1983, p. 47.

[40]. Johri, Pravin K., "A Linear Programming Approach to Capacity Estimation of Automated Production Lines with Finite Buffers" Int. J. Prod. Res., 1987, Vol. 25, No. 6, p. 851.

[41]. Johri, Pravin K., "Iterative Algorithms for estimating the Capacity of Automated Production Lines with Finite Buffers" Int. J. Prod. Res., 1989, Vol. 27, No. 4, p. 687.

[42]. Kanet, John, J., "Towards Understanding Lead Times

in MRP Systems" Production & Inventory Management, 3rd Quarter, 1982, p. 1.

[43]. Karmarkar, Uday, and Kekre, Sunder, "Manufacturing Configuration, Capacity and Mix Decisions considering Operational Costs" Journal of Manufacturing Systems, Vol. 6, No. 4, p. 315.

[44]. Karmarkar, Uday S., Kekre, Sham, and Kekre, Sunder, "Capacity Analysis of a Manufacturing Cell" Journal of Manufacturing Systems, Vol. 6, No. 3, p. 165.

[45]. Karni, Reuven, "Capacity Requirements Planning - Optimal Workstation Capacities" Int. J. Prod. Res., 1981, Vol. 19, No. 5, p. 595.

[46]. Karni, Reuven, "Capacity Requirements Planning - a Systematization" Int. J. Prod. Res., 1982, Vol. 20, No. 6, p. 715.

[47]. Koenigsberg, E., and Mamer, J., "The Analysis of Production Systems" Int. J. Prod. Res., 1982, Vol. 20, No. 1, p. 1.

[48]. Koenigsberg, Ernest, "Production Lines and Internal Storage - A Review".

[49]. Kusiak, Andrew, "The Production Equipment Requirements Problem" Int. J. Prod. Res., 1987, Vol. 25, No. 3, p. 319.

[50]. Lee, Sang-Bum, and Luss, Hanan, "Multifacility-Type Capacity Expansion Planning: Algorithms and Complexities" Operations Research, Vol. 35, No. 2, March-April 1987, p. 249.

[51]. Lin, Chem-Win, and Moodie, Colin L., "Hierarchical Production Planning for a Modern Steel Manufacturing System" Int. J. Prod. Res., 1989, Vol. 27, No. 4, p. 613.

[52]. Lundigran, Robert, "What is this thing called OPT" Production & Inventory Management, 2nd Quarter, 1986, p. 2.

[53]. Lunz, Alfred G., "The Missing Factors - The Real keys to effective capacity requirements planning and control" Production & Inventory Management, 2nd Quarter, 1981, p. 1.

[54]. Luss, Hanan, "Operations Research and Capacity Expansion Problems: A Survey" Operations Research, Vol. 30, No. 5, Sept.-Oct. 1982, p. 907.

[55]. Maturo, Michael P., "The Language of OPT" APICS

1983 Conference Proceedings, p. 567.

[56]. McAreavey, D., Hoey, J., and Leonard, R., "Designing the Closed Loop Elements of a Material Requirements Planning System in a Low, Make-to-order Company (with Case Study)" Int. J. Prod. Res., 1988, Vol. 26, No. 7, p. 1141.

[57]. Meal, Harlan C., "Putting Production Decisions Where They Belong" Harvard Business Review, March-April 1984, p.102.

[58]. Meal, Harlan C., Wachter, Manfred H., and Whybark, D. Clay, "Material Requirement Planning in Hierarchical Production Planning Systems" Int. J. Prod. Res., 1987, Vol. 25, No. 7, p.947.

[59]. Meleton, Marcus P., "OPT - Fantasy or Breakthrough" Production & Inventory Management, 2nd Quarter, 1961, p. 13.

[60]. Parnaby, J., "Concept of a Manufacturing System" Int. J. Prod. Res., 1979, Vol. 17, No. 2, p. 123.

[61]. Plenert, Gerhard, and Best, Thomas, D., "MRP, JIT, and OPT: What's Best?" Production & Inventory Management, 2nd Quarter, 1986, p. 22.

[62]. Plossl, Keith, "Production in the Factory of the Future" Int. J. Prod. Res., 1988, Vol. 26, No. 3, p. 501.

[63]. Plossl, Keith, "Throughput Time Control" Int. J. Prod. Res., 1988, Vol. 26, No. 3, p. 493.

[64]. Plossl, Keith. Manufacturing Control: The Last Frontier for Profits. Reston, Virginia. Reston Publishing Company, Inc. 1973.

[65]. Rocklin, Sol M., Kasper, Arik, and Varvaloucas, George C., "Capacity Expansion/Contraction of a Facility with Demand Augmentation Dynamics" Operations Research, Vol. 32, No. 1, Jan.-Feb. 1984, p. 133.

[66]. Sadowski, R. P., Waller, P. M., and McNeely, R. A., "A Constrained Machine Model for Calculating Effective Production Capacity" Comput. & Ind. Engng., Vol. 5, No. 3, p. 173.

[67]. Sari, F. John, "How to Master Schedule if your Problem is Capacity, not Materials" APICS 1987 Conference Proceedings, p. 582.

[68]. Sarkar, Bhaba R., "Some Comparative and Design Aspects of Series Production Systems" IIE Transactions,

September 1984, p. 229.

[69]. Sarkar, Bhaba, R., and Harris, Roy D., "The Effect of Imbalance in a Just-in-time Production System: A Simulation Study" Int. J. Prod. Res., 1988, Vol. 26, No. 1, p. 1.

[70]. Schmitt, Thomas G., Klastorin, and Shtub, Avraham, "Production Classification System: Concepts, Models and Strategies" Int. J. Prod. Res., 1985, Vol. 23, No. 3, p. 563.

[71]. Shantikumar, J. G., "On the Production Capacity of Automatic Transfer Lines with Unlimited Buffer Space" AIIE Transactions, September 1980, p. 273.

[72]. Solberg, James J., "Capacity Planning with a Stochastic Workflow Model" AIIE Transactions, June 1981, Volume 13, No. 2, p. 116.

[73]. South, John B., "Continuous Excess Capacity versus Intermittent Extra Capacity to Control Average Queue Size in a Random Environment" Production & Inventory Management, 1st Quarter, 1985, p. 103.

[74]. South, John B., and Stewart, Robert J., "A 50% increase in Lot Size and Planned Manufacturing Lead Time with no Increase in Work-in-process Inventory" Production & Inventory Management, 2nd Quarter, 1986, p. 18.

[75]. Stewart, Simon P. J., "Using OPT to Schedule a Machine Shop" APICS 1982 Conference Proceedings, p. 92.

[76]. Sung, C. S., and Chang, S. H., "A Capacity Planning Model for a Multi-Product Facility" Eng. Opt., 1987, Vol. 10, p. 263.

[77]. Swann, Don, "Using MRP for Optimized Schedules (emulating OPT)" Production & Inventory Management, 2nd Quarter, 1986, p. 30.

[78]. Thurwachter, William A., "Capacity Driven Planning" APICS 1982 Conference Proceedings, p. 384.

[79]. Udaybhanu, V., and Morton, T. E., "Planning Horizons for Capacity Expansion" European Journal of Operational Research, 34, 1988, p. 297.

[80]. Vollmann, Thomas E., "Capacity Planning: the Missing Link" Production & Inventory Management, 1st Quarter, 1973, p. 61.

[81]. Vollmann, Thomas E., "OPT as an Enhancement to MRP-

II" Production & Inventory Management, 2nd Quarter, 1986, p. 38.

[82]. Vollmann, Thomas E., Berry, William L., and Whybark, D. Clay. Manufacturing Planning and Control Systems. Homewood, Illinois. Dow Jones-Irwin. 1984

[83]. Weeks, James K., "WIP Inventory Management, Productivity and Costs" Production & Inventory Management, 1st Quarter, 1978, p. 30.

[84]. Wild, A. M. A., "Control of work-in-progress" Production & Inventory Management, 2nd Quarter, 1977, p. 30.

[85]. Wilhelm, W. E., Saboo, S., and Johnson, R. A., "An Approach for designing Capacity and Managing Material Flow in Small Lot Assembly Lines" Journal of Manufacturing Systems, Vol. 5, No. 3, p. 147.

[86]. Wilhelm, Wilbert E., "A Model to Evaluate Job Shop Productive Capability" AIIE Transactions, Vol. 8, No. 2, p. 176.

[87]. Wolfmeyer, Kevin, "Lead Time Control in the Job Shop" Production & Inventory Management, 1st Quarter, 1980, p. 30.

VITA

Ashish Masih, son of Jonathan and Ida Masih, was born in Allahabad, India on October 3, 1965. He graduated from St. Joseph's High School, Allahabad in 1983. In 1987 he received a Bachelor of Technology in Mechanical Engineering from Indian Institute of Technology, New Delhi. He then joined the graduate program in Manufacturing Systems Engineering program at Lehigh University and will be graduating with a Master of Science degree. During his stay at Lehigh, he was also a Research Assistant with the Manufacturing Systems Engineering program. Upon graduation, Ashish will be joining Peat Marwick Main & Co. as a consultant.