CONTROLLING W.I.P. AND LEADTIMES IN JOB SHOPS

by

Uday S. Karmarkar

William E. Simon Graduate School of Business Administration
University of Rochester

Controlling W.I.P. and Leadtimes in Job Shops

by

Uday S. Karmarkar

William E. Simon Graduate School of Business Administration
University of Rochester
Rochester, New York 14627

ABSTRACT

In many manufacturing problems, the total lead time taken to manufacture a product is an important consideration. Long lead times impose costs due to higher work-in-process inventory, increased uncertainty about requirements, larger safety stocks and poorer performance to due dates. Traditional lot sizing models ignore lead time related costs, although there are systematic relationships between lot sizes and lead times. This paper examines these relationships and their implications for lot sizing and work-in-process inventories, for batch manufacturing shops with queues.

CONTROLLING W.I.P. AND LEADTIMES IN JOB SHOPS

Uday S. Karmarkar
University of Rochester

INTRODUCTION

A significant proportion of all value added in manufacturing in the U.S. is in repetitive small batch manufacturing which largely occurs in the complex multi-machine multi-product facilities termed closed job shops. A typical example of such a facility is a parts manufacturing shop that produces parts for assembly, to stock or against orders from regular customers. The prevalence of the job shop configuration in manufacturing makes it imperative to understand the behavior of such shops and to develop effective methods of managing them.

Job shops, by their nature, vary a great deal in terms of the machines in the shop, the mix of products manufactured, the degree of automation employed and the pattern of work flow within the shop. However, their outstanding common characteristic is that they exhibit significant queueing behavior. It is not unusual for jobs to spend over 95% of their time in a shop waiting in queues for processing or for transportation. The resulting work-in-process (W.I.P.) levels alone constitute a significant cost of operation. For example, a shop processing $10 million worth of raw materials annually, with a physical "turn" of 4, has a W.I.P. level of $2.5 million. At a 20% carrying cost, the resulting annual capital charges are $500,000 which are a significant proportion of the contribution margin that can be obtained on the $10 million flow. What is more, these are costs directly attributable to materials management in the shop. Controlling queues, W.I.P. and manufacturing leadtime or shop time is thus central to the effective management of job shops.

JUST-IN-TIME FROM JOB SHOPS?

The queues that appear in job shops lead to delays which comprise the major portion of total manufacturing leadtime. Yet, there is little knowledge about why these queues occur and how they are to be controlled. Current practice typically consists of estimating leadtimes for individual items based on historical data. The problem with this is that leadtimes are dynamic and depend on many changing factors including product mix, volume and batch sizes. As a result, leadtime offsets, as used in an MRP system, must be inflated to account for the uncertainty about leadtimes. Compounding the problem even further is the tendency on the part of assembly departments, to "game" the system by inflating leadtimes so as to ensure part availability. A somewhat better view of queues is provided by "input-output" analysis, which identifies points of queue buildup in aggregate terms, but cannot suggest remedies.

Long and variable leadtimes have many costs other than W.I.P.. First, it is difficult to see how such a job shop can be operated in a just-in-time (JIT) manner. A key ingredient for "pull" or JIT operation is reliable or constant leadtimes. Variability of leadtimes is a particular problem when part arrivals must be coordinated for assembly. Furthermore, there are other inventory related consequences. Allowances for leadtime variability as "safety time" tend to cause early arrivals of parts into inventory in addition to inventory in queues. High total manufacturing leadtimes also mean that higher levels of finished goods safety stocks are required to protect against the leadtime. Long leadtimes cause a loss of information: production must be initiated against future demands which are more uncertain; long delays between production and use create difficulties in detecting and correcting quality problems; reacting to short term problems or demands is impossible.

In order to control leadtime, it is essential to understand queueing behavior in job shops. A conceptual view of the queueing phenomenon allows us to design solutions to the many problems described above.

QUEUES IN JOB SHOPS

The ideal production facility is a single product, continuous flow process like a chemical plant or a dedicated automated assembly line. These are examples of a balanced process while job shops at the other end of the spectrum are unbalanced processes. If we look at a particular work center in a job shop, streams of discrete batches arrive from many different locations. The overlapping of these streams creates an arrival stream that is highly uneven. Furthermore, the work required to process each batch also varies due to the heterogeneity of items and batch sizes. Unlike a station in perfectly balanced flow line, such a work center has no natural "cycle time" and furthermore, there is little coordination between one work center's pattern of work and another. A computer scientist might describe such a facility as "asynchronous". This mismatch between arrivals and processing patterns is the fundamental reason for queues in job shops.

It is extremely important to recognize that these queues occur not only at processing centers but also at material handling and movement points. In other words, a material handling system, whether it be an automated train or a group of fork lifts, can be thought of as a work center with a queue of jobs waiting to be "processed". A failure to understand this leads to anomalies such as careful prioritization and scheduling at work centers while the movement of batches is left to chance. A symptom of this myopia is that setup and processing times at work centers are often measured in minutes on route sheets while a succeeding move operation will sometimes be described as taking 48 hours. Of course, what makes dispatching in move queues difficult is that the queue for a material handling resource is dispersed throughout the shop and cannot be "seen" at one time.

We have all waited in many kinds of queues, in banks and stores, and such situations give us useful analogies for thinking about shop behavior. A particularly apt analogy to a job shop is a traffic network, where intersections correspond to work centers or other resources. Batches of different items follow a route through this network and compete for possession of intersections. Of course, traffic intersections are relatively simple places with only two to four competing streams whereas a machine center in a job shop may process 30 different product streams. Batches of work are released into the network like convoys and these production batches may be split into smaller batches under the rules by which intersections are operated.

What intuitive insights can we gain from such analogies? First, the degree of loading of a resource - its "traffic intensity" or utilization - is obviously a major determinant of queues. However, other factors are equally important. Clearly batch sizes -- the timing pattern -- at an intersection will affect the waiting time for all other streams. Very small batches are workable only at relatively idle intersections: the analogy is with 4-way stops in lightly travelled areas. As the traffic load increases, larger batch sizes are appropriate. However, very large batch sizes for one item -- think of a military convoy or a motorcade -- cause unacceptably large waiting times for other streams. Two other analogies: Expedited lots are like emergency or police vehicles; they cause severe disruptions in the normal pattern of flow and result in worse queues. Overlapped operations are analogous to synchronized traffic lights; it is necessary for batches to move quickly from machine to machine to make synchronization work.

There are many ways to control queueing behavior and the resulting W.I.P. and leadtimes. At the design level they include capacity choice, equipment choice, material handling system design, shop layout, item grouping and cellular configuration. At the operational level, the major alternatives are batch sizing and detailed scheduling including job release and sequencing at work centers. Recent research shows that batch sizing effects are very important which seems apparent given the traffic network analogy. In retrospect, it is curious that the relationship between batch sizes and queues has received such little attention in either queueing or scheduling research.

LOT-SIZES AND QUEUEING BEHAVIOR

Lot-sizes or batch sizes have three major effects on queues. First, there is a "scale" effect: as lotsizes are scaled upwards, the queueing delay rises proportionally. The number of batches waiting in a queue remains the same, but each batch is larger. This suggests that batch sizes should be scaled downwards except that at some point, a second "intensity" effect is encountered. Smaller batch sizes lead to increased setups which consume productive time. As a result, the ratio of the processing to be done to the capacity available begins to rise.

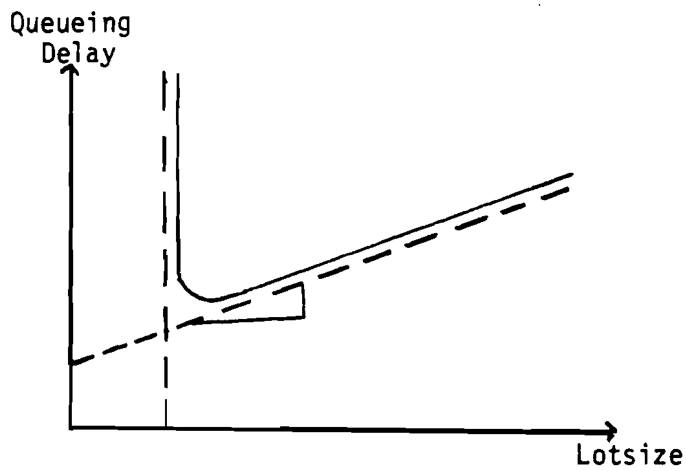

Figure 1: Relationship Between Lotsizes and Queueing Delays

In queueing situations, this leads to the phenomenon of saturation where large queues build up even when there is nominally enough capacity to process incoming work. These two effects can be conveniently illustrated as in Figure 1. The rise in queueing delay with small batch sizes is always very steep below the minimum. The increase in queueing delay with increasing size is more gradual, but is worse for heavily loaded resources. In between there is a point were queueing delays are minimized.

The third effect of batching on queueing behavior is more subtle. Proper lot-sizing can make batches arriving at a work center more uniform in terms of their processing demands. Uniformity in processing times tends to reduce queues and also leads to a more uniform stream of departures from the work center. This effect is difficult to visualize; in a sense it is a way of creating a more balanced process. This last effect makes it especially clear that queueing is a multi-product, joint phenomenon. Lot-size choices for one item affect all others that compete with it for resources. Conversely, changing a single items batch size cannot improve its queueing delays, since it must wait for other items.

The qualitative ideas described here have been thoroughly explored using approximate queueing models (Karmarkar, 1983a; Karmarkar, Kekre and Kekre, 1983; and Zipkin, 1983), exact numerical analysis (Sumita and Kijima, 1984) deterministic sequencing models (Karmarkar, 1983b; Dobson, Karmarkar and Rummel, 1984) and simulation (Freeman, 1982; Karmarkar, 1983a). There is thus substantial evidence for the validity of the basic ideas and in fact they can be easily confirmed by a simple simulation. In order to model a realistic job shop, these

Lot Size Scaling	Average Leadtime (days)		W.I.P.($)	
	Q-LOTS	Simulation	Q-LOTS	Simulation
0.60	49.56	17.05	391,210	168,140
0.80	9.32	7.90	75,290	65,810
1.00	8.57	7.26	68,840	60,400
1.25	9.06	8.24	72,750	68,860
1.50	9.91	9.16	79,680	76,320
1.75	10.89	10.30	87,650	85,440
2.00	11.95	11.47	96,170	95,910

Table I: Comparison of predicted leadtimes with W.I.P. with GPSS simulation of cell.

methods have been extended to a multi-item multi-center model (Q-LOTS) described by Karmarkar, Kekre and Kekre (1983). The model has been successfully tested against a detailed simulation of an actual manufacturing cell with 13 items and 10 work centers. The results are described in a paper by Karmarkar, Kekre, Kekre and Freeman (1983). Table I shows a comparison of predicted and simulated average leadtimes and average W.I.P.($) as the lot sizes of all items in the cell, are scaled up and down in the same proportion.

IMPLICATIONS FOR MASTER SCHEDULING

In addition to predicting queue times and leadtimes for a given set of lot size choices, Q-LOTS also automatically searches for the best batching policy for a given objective function. For the manufacturing cell mentioned above, the model was used to obtain the batch sizes that would minimize the average demand-weighted leadtime for the cell. This is a very convenient objective function since minimizing leadtime is essentially equivalent to minimizing many of the costs described earlier. Yet, no cost data is required. However, if necessary another objective such as minimizing W.I.P. plus finished part inventory, could have been used. Terms accounting for setup costs due to material losses or added labor costs can also be added.

Batch sizes determined by the model, should be used to provide min-max limits on lot sizes in an MRP system. Furthermore, the leadtime estimates from the model can be used to provide released dates for the lots. Note that this is basically an "order launch" approach with no detailed scheduling at the shop floor level. It may be that the poor reputation of order launch methods is somewhat unwarranted, since they work quite well once lot sizes and release dates are selected so as to give the best pattern of work flow through the shop. Such methods also have the advantage of having very low implementation costs, since the policy can be introduced at the master scheduling level, and requires no interference or change in shop floor control systems, responsibilities or organization. In the cell described earlier, the simulation verified that W.I.P. inventories and leadtimes could be cut by 60-70% over current levels by these methods. In a separate case study on a cell in a different firm, savings were projected at 40% of current levels. Detailed scheduling can improve performance beyond this point, but the savings are quite modest once the lot sizes and release dates are properly selected (Kekre, 1984).

PRODUCTION PLANNING

The basic Q-LOTS model assumes a certain product or item mix in the shop. Over a period of time, seasonal variations or trends in demand will cause this mix to change -- often in a statistically predictable way. Current methods of aggregate production planning attempt to match nominal capacity with demand fluctuations while minimizing the accumulation of seasonal finished goods inventories. This conventional approach does not account for the fluctuations in W.I.P. that are caused by changing product and part mix. The fluctuations can be quite severe, and ironically, production smoothing methods focussing only on finished goods inventories can make them worse.

The lot-sizing model can be imbedded in a seasonal planning model which now accounts for W.I.P. as well as finished inventories. Broadly speaking, the effect is to smooth production levels as well as production mix over the planning horizon. Kekre (1984) has tested this approach on data from a small manufacturing facility and the results indicated 10-15% savings in total annual inventory holding costs.

The integrated planning model is quite complex. A simpler approach is to use Q-LOTS to evaluate the effect of mix and volume changes as shop behavior. This is an effective tool for communicating information on W.I.P. effects to production planning. In our experience, the adverse impact of aggregate production changes on shop behavior, queues and W.I.P. is often not well understood by production planners. The basic idea is simply that keeping production volumes as well as mix as level as possible, will allow a stable lot sizing policy that minimizes W.I.P., leadtimes, and lead time variability

CAPACITY AND DESIGN ISSUES

Once the characteristics of shop behavior are captured, the model can be used to analyze the effect of capacity and design decisions on the shop. For example, there is considerable attention being given to setup time reduction. However, it is often difficult for a shop manager to determine exactly which items and which machines should be the targets of such improvements. The models described above can quantify the effect of a given improvement by predicting its effect on W.I.P. and leadtimes. It is straightforward to produce a prioritized list of setups that have the greatest potential for impact on shop performance. Interestingly, the Q-LOTS model predicts that there is a linear relationship between setup time and inventory costs and leadtimes. This is at variance with the conventional EOQ model.

The model can also be used to perform a what-if analysis of changes in shift and overtime policy, the addition of machines to work centers, the addition of items to a cell, and cell configuration. For example, Table II shows the results of such an analysis on a small parts manufacturing cell producing 27 parts on 9 major work centers. When used in this way the model automatically adjusts lot size policy to allow for the change in capacity. For the example shown in Table II, the addition of two machines is probably the best option from a cost benefit point of view. Although a second shift results in lower lead time and W.I.P., it is a much more costly alternative.

	LEAD TIME DAYS	W.I.P. $
Current lotsizes & overtime	68.56	139,480
Best lotsizes & overtime	38.08	75,720
Current lotsizes & 2 shifts	8.47	16,355
Best lotsizes & 2 shifts	4.80	9,555
One machine added at work Center B, no over time	44.23	86,970
Two machines added: one at work center B, other at D.	16.39	32,435

Table II: Example of capacity alternatives analysis.

SUMMARY

This paper has described a new conceptual approach to the analysis of job shops with queues. The approach relies on modelling the dynamic behavior of material flow through a shop. It reveals the connection between lot-sizing policies and the W.I.P. and leadtime problems that are endemic to job shops. In addition to providing an intuitive view of shop behavior, the Q-LOTS model quantifies the relationships involved. Tests of the methods have shown that dramatic deductions in W.I.P. and leadtimes are possible by choice of lot sizing policy alone. Furthermore, these savings can be substantially augmented by suitable planning and design changes. The Q-LOTS model provides a quantiative way of analyzing the costs and benefits of such changes.

REFERENCES

Dobson, G., Karmarkar, U. S. and Rummel, J., "Single Machine Sequencing with Lot Sizing", Graduate School of Management Working Paper No. QM8419, University of Rochester.

Freeman, S., (1982), personal communication.

Karmarkar, U. S., (1983a), "Lot Sizes, Lead Times and In-Process Inventories", Graduate School of Management Working Paper No. QM8312, University of Rochester.

Karmarkar, U. S., (1983b), "Lot Sizes and Sequencing Delays", Graduate School of Management Working Paper No. QM8314, University of Rochester.

Karmarkar, U. S., Kekre, S. and Kekre, S., (1983), "Multi-Item Lotsizing and Lead Times", Graduate School of Management Working Paper No. QM8325, University of Rochester.

Karmarkar, U. S., Kekre, S. and Kekre, S., (1984), "Lot Sizing in Multi-item, Multi-machine Job Shops", Graduate School of Management Working Paper No. QM8402, University of Rochester.

Karmarkar, U. S., Kekre, S., Kekre, S., and Freeman, S., (1983), "Lotsizing and Lead Time Performance in a Manufacturing Cell", Graduate School of Management Working Paper No. QM8328, University of Rochester. To appear in Interfaces.

Kekre, S., (1984), "Management of Job Shops", unpublished Ph.D. Thesis, Graduate School of Management, University of Rochester.

Kekre, S., (1984), "Some Issues in Job Shop Design", unpublished Ph.D. Thesis, Graduate School of Management, University of Rochester.

Sumita, U. and Kijima, M., (1984), "On Optimal Bulk Size and Single-Server Bulk-Arrival Queueing Systems with Set-up Times -- Numerical Exploration via the Laguerre Transform", Graduate School of Management Working Paper No. QM8412, University of Rochester.

Zipkin, P. H., (1983), "Models for Design and Control of Stochastic, Multi-Item Batch Production Systems", Columbia Business School Research Working Paper No. 496A, Columbia University, New York.

BIOGRAPHICAL SKETCH

Uday S. Karmarkar is Associate Professor of Operations Management and Operations Research and Director of the Center for Manufacturing and Operations Management at the Graduate School of Management of the University of Rochester. His research interests lie in manufacturing management as well as service issues such as product support. His most recent research efforts in manufacturing have been directed toward scheduling, lotsizing and capacity design problems in small batch manufacturing systems. In other research he is studying the role of strategies such as loaners, service contracts, field service systems and modular design in the market for product support. His interests also include manufacturing cost and performance measurement, organization and incentives.

Professor Karmarkar's work has been published in Management Science, Operations Research, Naval Research Logistics Quarterly, IIE Transactions, Econometrica, Interfaces, Harvard Business Review, and Organizational Behavior and Human Performance. He is an Associate Editor of Operations Research and of the Naval Research Logistics Quarterly.

Professor Karmarkar is a member of ORSA, TIMS, OMA and IIE. He is a co-founder of ORSA's Special Interest Group in Manufacturing. He is a member of the Manufacturing Management Council of the Society for Manufacturing Engineers and an honorary member of the Rochester APICS Chapter.

Professor Karmarkar has been involved in consulting, research projects and management seminars with companies such as Deere and Co., Eastman Kodak, Motorola and Xerox.