

**OPTIMAL PART DELIVERY DATES IN  
SMALL LOT STOCHASTIC ASSEMBLY SYSTEMS**

by

Rajiv K. Srivastava

Dissertation submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in  
Industrial Engineering and Operations Research

APPROVED:

\_\_\_\_\_  
S. C. Sarin, Ph.D, Chairman

\_\_\_\_\_  
H. D. Sherali, Ph.D.

\_\_\_\_\_  
J. J. Greene, Ph.D

\_\_\_\_\_  
R. S. Russell, Ph.D

\_\_\_\_\_  
M. S. Jones, Ph.D

July, 1989

Blacksburg, Virginia

**OPTIMAL PART DELIVERY DATES IN  
SMALL LOT STOCHASTIC ASSEMBLY SYSTEMS**

by

Rajiv K. Srivastava

Subhash C. Sarin, Ph.D, Chairman

Industrial Engineering and Operations Research

(ABSTRACT)

An important issue in the design and operation of assembly systems is the coordination of part deliveries and processing operations. These decisions can have a significant impact on inventory cost and customer service. The problem is especially complex when actual delivery and processing times are stochastic in nature, as is the case in small lot manufacturing.

In this research a new methodology is developed for determining optimal part delivery dates in stochastic small lot assembly systems. This methodology is based on the descriptive model that comprises of taking the maximum of several random variables. The part arrival and processing times are assumed to follow various known probability distributions. The model includes consideration of limited buffers between stations. The overall objective is to minimize the expected total of part and subassembly inventory cost, makespan cost and tardiness cost.

An approach based on the optimization of individual stations in isolation is used to obtain the part delivery dates at each station. Comparison of the approach with the nonlinear programming based approach to the problem indicates that it generates almost as good solutions in a fraction of the computation time. This approach is then used to study system behavior under various operating conditions. Results indicate

that the lognormal and gamma distributions result in higher total costs than the normal distribution. However, the normal distribution can be used to determine part delivery dates even if the actual distribution is lognormal or gamma, with relatively small errors compared to the solutions obtained using the correct distribution. Variability is the most important factor in the design of the system, and affects the determination of due dates, buffer capacity requirements, choice of distribution, and estimates of system performance. The role of buffer capacities, however, is not very critical in the design of small lot unbalanced lines.

## **Acknowledgements**

I would like to take this opportunity to express my appreciation to the many persons whose contributions have made this dissertation a reality. I am particularly grateful to Dr. Subhash C. Sarin, my chairman, for his inspiring guidance throughout my education at Virginia Tech, and especially during the course of this research.

I am also thankful to the other members of my committee, Drs. Hanif D. Sherali, Timothy J. Greene, Roberta S. Russell, and Marilyn S. Jones, for their encouragement and valuable suggestions. In addition, I wish to thank the faculty and staff of the Department of Industrial Engineering and Operations Research for their support during my stay at Virginia Tech.

It is not possible to adequately express my gratitude to my wife, \_\_\_\_\_ for patiently enduring a break in her career in order to be with me during this important step in my life. Finally, I am most grateful to my daughters, \_\_\_\_\_ and \_\_\_\_\_ for being a constant source of inspiration at the most trying times.

# Table of Contents

<b>1.0 Introduction</b>	<b>1</b>
1.1 Introduction to Assembly Systems	1
1.2 Production Planning in Small Lot Assembly Systems	2
1.3 Determining Lead Times and Due Dates	5
1.4 Descriptive vs. Prescriptive Methods	6
1.5 Nature and Limitations of Previous Research	8
1.6 Description of the Problem and Objectives	12
1.7 Organization of the Dissertation	16
<b>2.0 Literature Review</b>	<b>18</b>
2.1 Introduction	18
2.2 Descriptive Models of Stochastic Assembly Systems with Part Input	19
2.3 Simulation Based Models for Planned Lead Times in Assembly Systems	25
2.4 Prescriptive Models for Setting Part Due Dates in Assembly Systems	28
2.5 Multi-stage Flow Lines with Limited Storage	35
2.6 Related Research	39
2.7 Summary	45
<b>3.0 Cost Model Development</b>	<b>47</b>
3.1 Introduction	47
3.2 Notation	48
3.3 A Fundamental Relationship	50

3.4	Distributions Considered	54
3.5	Moments of the Maximum of Two Random Variables	58
3.5.1	Normal Distribution	59
3.5.2	Lognormal Distribution	60
3.5.3	Gamma Distribution	63
3.6	Test for the Distribution of the Maximum	66
3.7	Moments and Distribution of the Sum of Two Random Variables	71
3.8	Expectation of the Maximum of a Random Variable and a Constant	76
3.9	Summary	77
<b>4.0</b>	<b>Solution of Single Job or Lot-for-lot Problem</b>	<b>79</b>
4.1	Introduction	79
4.2	Single Station Optimization	80
4.2.1	Single Station Solution for Normal Distribution	82
4.2.2	Single Station Solution for Non-Normal Distributions	84
4.3	Multi-station Assembly Problem	88
4.3.1	Initial Solution Approaches	89
4.3.2	Comparison of Solutions	97
4.4	Incorporation of Random Processing Times	102
4.5	Development of Approximate Solution Method for Normal Distribution	105
4.5.1	Sources of Error in Independent Solution Method	107
4.5.2	Correction for Variance	110
4.5.3	Correction for Cost Proportion	114
4.5.4	Correction for Random Processing Times	118
4.5.5	Development of Hybrid Solution Technique	119
4.6	Implementation of the Approximation Method and Results	123
4.7	Solution of Multi-station Problem for Non-normal Distributions	127
4.7.1	Optimal and Independent Solutions	127

4.7.2	Approximate Solution Technique for Non-normal Distributions	132
4.8	Summary	134
<b>5.0</b>	<b>Solution of Multiple Job Problem</b>	<b>135</b>
5.1	Introduction	135
5.2	Precedence Network Representation of Multiple Job Assembly	136
5.2.1	Assembly Systems with Unlimited Buffers	136
5.2.2	Assembly Systems with Limited Buffers	138
5.2.3	Objective Function for Multiple Job Assembly	141
5.2.4	Evaluation of the Objective Function	142
5.3	Determination of Optimal Solution for Multiple Job Problem	146
5.3.1	Solution Approach	146
5.3.2	Analysis of Results	148
5.4	Development of Heuristic Procedure for Multiple Job Problem	157
5.4.1	Bottleneck Station and Critical Path	157
5.4.2	Incorporating Buffers in Deterministic Networks	159
5.4.3	Incorporating Random Processing Times	160
5.4.4	Heuristic Procedure for the Multiple Job Problem	164
5.5	Implementation of Heuristic and Comparison of Results	169
5.6	Generalization of Optimal and Heuristic Methods for Non-normal Distributions	175
5.7	Summary	178
<b>6.0</b>	<b>Analysis of Results and System Behavior</b>	<b>180</b>
6.1	Introduction	180
6.2	Sensitivity to System Parameters	182
6.2.1	Effect of Changes in Costs	182
6.2.2	Effect of Changes in Variability	186
6.3	Effect of Randomness on Cost Components	191

6.4	Effect of Buffer Capacities	195
6.5	Effect of Distributions	198
6.5.1	Relationship Between Buffer Capacities and Distributions	200
6.5.2	Choice of Distribution	203
6.6	Summary	207
<b>7.0</b>	<b>Conclusion</b>	<b>208</b>
7.1	Summary and Achievements of Research	208
7.2	Major Findings of This Research	210
7.3	Avenues for Future Research	212
<b>References</b>		<b>215</b>
<b>Appendix A. Moments and Correlations of the Maximum of Two Lognormals</b>		<b>224</b>
<b>Appendix B. Moments of the Maximum of Two Gamma Random Variables</b>		<b>228</b>
B.1	Erlang Distribution	229
B.2	General Gamma Distribution	231
<b>Appendix C. Correction Equations for Different Distributions</b>		<b>234</b>
<b>Vita</b>		<b>237</b>



## List of Illustrations

Figure 1. The Assembly System	13
Figure 2. Relationship Between Various Times for Operation (i j)	51
Figure 3. Determination of Error Between Approximating and Exact Distributions	68
Figure 4. Single Job Assembly System	81
Figure 5. Behavior of Mean Assembly Start Time vs. Mean Part Delivery Date	85
Figure 6. Relationships in Single Job Two Station Problem	92
Figure 7. Objective Function Contours for a 2 Station Problem	95
Figure 8. Effect of Change in Variance for Different DS Values	111
Figure 9. Effect of Change in Variance	112
Figure 10. Effect of Change in Cost Proportion for Different CP Values	115
Figure 11. Effect of Change in Cost Proportion	116
Figure 12. Precedence Network Representation of a Multiple Job Assembly System with Unlimited Buffer Capacities	137
Figure 13. Precedence Network Representation of a Multiple Job Assembly System with Limited Buffer Capacities	140
Figure 14. Precedence Levels in a Multiple Job Assembly System	144
Figure 15. Optimal Intervals Between Successive Jobs	151
Figure 16. Optimal Part Due Dates Along Critical Path	152
Figure 17. Optimal Part Due Dates at Station 4	153
Figure 18. Flowchart for Implementation of Heuristic	165
Figure 19. Effect of Changes in Variances on Total Cost in Multiple Job Problem	190

## List of Tables

Table 1.	Comparison of Deviations of Actual and Approximating Distributions for the Maximum	70
Table 2.	Comparison of Deviations of Actual and Approximating Distributions for the Sum	75
Table 3.	Comparison of Single Station Optimal Values for Different Distributions	87
Table 4.	Comparison of NLP, ISM and DP Methods for 3 Station Problems	98
Table 5.	Comparison of NLP, ISM and DP Methods for 10 Station Problems	100
Table 6.	Effect of Processing Time Variability on Costs and Decisions for 3 Station Problems	103
Table 7.	Comparison of Results from Approximate Solution Method	124
Table 8.	Comparison of Results from Approximation	126
Table 9.	Comparison of Normal, Lognormal and Gamma Distributions for 3 Station Problems (using NLP method)	129
Table 10.	Performance of ISM Method for Lognormal and Gamma Distributions for 10 Station Problems	130
Table 11.	Performance of HYM Method for Lognormal and Gamma Distributions for Test Problems	133
Table 12.	Comparison of Optimal Solutions for Sample 5 Station, 5 Job Problems	149
Table 13.	Comparison of Optimal Solutions for Sample 5 Station, 5 Job Problems	150
Table 14.	Frequency of Critical Activities in a Random Processing Network	162
Table 15.	Heuristic Solutions for Sample 5 Station, 5 Job Problems	170
Table 16.	Results for Multiple Job Problems for Normal Distribution	173
Table 17.	Comparison of Techniques for Multiple Job Problems for Normal Distribution	174
Table 18.	Results for Multiple Job Problems for Different Distributions	176
Table 19.	Comparison of Techniques for Multiple Job Problems for Different Distributions	177

Table 20. Effect of Changes in Selected Parameters on Total Costs at Solutions	184
Table 21. Effect of Changes in Variances on Total Costs at Solutions . . . . .	189
Table 22. Effect of Randomness on System Cost Components at Solutions . . . .	193
Table 23. Effect of Buffers on System Cost Components at Solutions . . . . .	197
Table 24. Effect of Distributions on System Cost Components at Solutions . . . .	199
Table 25. Effect of Distributions and Buffer Capacities on Total Costs at Solutions	201
Table 26. Effect of Distributions and Buffer Capacities on Total Costs at Solutions	204
Table 27. Errors Due to Choice of Distribution . . . . .	206

# **1.0 Introduction**

## **1.1 Introduction to Assembly Systems**

Assembly systems are used to transform parts and subassemblies into finished products. They are composed of a series of operations performed in accordance with the product structure. These tasks are arranged in workstations made up of machines, equipment and workers. Depending on the nature, volume, and complexity of the product, assembly systems can range from single station configurations to long assembly lines dedicated to a particular product.

Assembly is the final stage of value addition in production, and parts used have generally undergone other manufacturing processes beforehand. Value added due to labor and material is at its highest at this stage. Assembly may also require specialized equipment, which involves significant capital investment. It is therefore an important production stage, and has a major impact on the overall cost of finished products (Blumenfeld, 1987).

Design of assembly systems includes the planning of material flow to the system, job flow through the facility, and finished product flow to the customer. The parts required for the assembly are delivered at the necessary station to be assembled on to the mainframe. These parts may be produced in-house or may be procured directly

from vendors. Each of these parts may in turn be the output of a branch line. An important issue in designing the material flow is the coordination of deliveries of the parts to stations and determining the timing of various operations along the line.

## **1.2 Production Planning in Small Lot Assembly Systems**

Traditionally, design of assembly systems has focussed on high-volume, standardized production to stock. Such systems are characterized by fixed automation and relatively invariant production schedules over a period of time. Jobs to be assembled flow through the system at a constant rate, and are processed at each station for a fixed time. Since processing times are fixed and constant across workstations, deliveries to stations and transfers between stations are synchronized. In such situations, the main objective is to streamline the flow of the assembly line and minimize the avoidable idleness. This is achieved by allocating the operations to workstations, subject to precedence restrictions, in order to balance the line.

Competitive forces and increasing customer requirements have increased the number of product variations, while reducing the run sizes of individual products. Thus manufacturers of such products do not have the luxury of dedicating facilities and making long runs. Mixed model assembly line balancing techniques have been developed to manage the flow of work along the line in case there is multiplicity of products, but that still assumes steady production to stock. In many fields, such as heavy equipment and weapon systems made to specific requirements, job production and small lots have always been the mode of operation. One of the major priorities

in the design and operation of manufacturing systems is the development of technologies and management techniques for such small lot systems.

At the level of part production, a number of technological developments over the last few years have improved the capabilities of coping with small lot production. These include Cellular Manufacturing, Flexible Manufacturing Systems, and Robotics. Appropriate tools and techniques for managing and coordinating these systems as part of the whole system are constantly being developed. However, the focus of these systems is on discrete parts manufacturing, predominantly in the metal-working environment. Almost all part manufacturing systems eventually feed assembly systems where the final product is put together. Assembly systems have seen relatively less attention, both in the development of new technologies and management techniques.

Assembly of small lots of large products is a complex, slow moving process. Low volumes and inventory costs coupled with high variations between stations and jobs make balancing of the line inappropriate. Moreover, because of the variability inherent in any new job, there is considerable random variation as well in delivery and processing times. This causes queues to develop and enforces waiting times for parts, subassemblies, and stations. This results in inventory costs and underutilization of facilities. Scheduling such systems must include coordination of the manufacturing tasks and part deliveries to ensure that customer requirements are met in the most cost effective manner. Analysis based on deterministic estimates would not provide realistic predictions of system performance. A model to study such a system must incorporate the stochastic nature of the variables to be able to measure the effects of uncertainty and evaluate the design of the line.

Material Requirements Planning Systems (MRP) is one of the methods that have been developed to manage production of assembled items. These items may be produced on a job basis, or as a batch of jobs processed through the manufacturing facility. The MRP approach relies on exploding the Master Production Schedule of end item requirements, using the Bill of Materials (BOM), to generate the time phased part requirements at different levels of the product structure. These requirements are offset by the planned lead times to determine due dates and the appropriate times for initiation of procurement and manufacturing action.

One of the most important inputs to the MRP system is the estimated lead time for procurement and manufacturing of different parts. Planned lead times and due dates play an important role in MRP systems (Orlicky, 1975). Their structure and level can impact every aspect of the performance of the system (Kanet, 1982). These due date estimates are used as a monitoring tool as well for coordinating the different sub-systems. Poor estimates of lead time allowances can negate all the advantages expected from the system. In multi-stage systems, this problem can get compounded even further, resulting in poor customer response, excessive inventories, and frequent rescheduling (Kanet and Christy, 1984). MRP systems do not inherently possess the optimization capability to be able to select the best values of planned lead times.

### **1.3 Determining Lead Times and Due Dates**

The task of determining reasonable planned lead times and due dates is not easy due to the dynamic and variable nature of the procurement and manufacturing process. Variability in part delivery and processing can arise due to a number of reasons. The most significant causes of delays and variability are queueing times at work stations, transportation delays, and setups required due to the non-repetitive nature of the jobs.

Coping with uncertainty in such systems can take a number of forms. First, efforts need to be made to reduce variability in the system by adding on appropriate control systems and managing the lead times through human intervention. Cost-benefit analysis therefore needs to be done on the reduction of variability compared to the errors arising out of it. Second, efforts can be made to improve the reliability of lead time estimates. This may be done by adjusting the Master Production Schedule to reflect the actual capacity, or by tracking lead time errors and updating the estimates constantly.

Finally, the remaining uncertainty has to be dealt with. One way of providing for it is to construct lead times that result in an appropriate probability of delivery, by adding on an allowance that is some multiple of the standard error. This method is also referred to as the safety buffer rule (Wilhelm, 1982). Such allowances are a more effective way of dealing with uncertainty than buffer stock (Whybark and Williams, 1976). In job-lot manufacturing, safety stocks may not be an option in any case due to risks of obsolescence.



A more complete method of providing for uncertainty is to determine the values of planned lead times that yield the best overall expected costs of inventory and tardiness. This method of dealing with uncertainty is the focus of this research. This method may be used to determine the cost of variability and provide a means of coping with it. The end result of the research will be the mean due dates for each part and operation, representing the net effect of lead times and allowances for uncertainty. The research is not confined to MRP systems, but encompasses any production planning system for assembly network type structures.

## **1.4 Descriptive vs. Prescriptive Methods**

Due to the complexity of small lot manufacturing systems, it is useful to have models to aid in their design and operation. Suri (1985) classifies modelling approaches in two main categories. Descriptive or evaluative models evaluate a given set of decisions based on certain performance criteria, while prescriptive or generative models are used to find decisions. Some features of these types of models are discussed below.

Descriptive or evaluative models are tools to help the decision maker develop an understanding of the behavior of the system. They help provide insight rather than decisions, and are useful for sharpening intuition. They also enable the designer to experiment with candidate decisions, although it may take a long time to arrive at a good decision. In problems with a large candidate solution space, it is difficult to be certain of finding even a reasonably good solution. Queueing networks and simu-

lation are among the commonly used evaluative methods for analysis of stochastic systems.

The complexity of many systems prohibits the development of prescriptive models. However, it may be possible to develop analytical models of system performance, which may be used to generate and understand system behavior and thus to guide the design. Simulation models are frequently used for analysis of complex systems, especially when special purpose packages and languages can be applied to ease the tedium of model building. However, in the case of simulation, it is necessary to choose appropriate factor levels and conduct experiments to determine the best combination. The validation of simulation models is another issue that must be addressed by the designer.

Response Surface Methodology is one of the methods that is used in conjunction with simulation to determine preferred factor combinations in stochastic problems. However, there is no guarantee of finding the optimal solution, and a large number of simulation runs may be required (Law and Kelton, 1982). Perturbation Analysis (Ho, 1984, Ho et al., 1985) is a relatively new approach that uses the results of a simulation run for a set of decision parameters to predict system behavior if these decisions are changed. However, this technique can currently be used only for small changes in decisions (Suri, 1985).

Prescriptive or generative models are used to find good decisions that achieve some pre-specified criterion, such as minimizing cost or maximizing performance. If such a model can be constructed, it can provide solutions to the problem faster and with greater likelihood than descriptive models. However, such a model may be extremely difficult to develop. It may also be difficult to modify decisions, or to examine

the sensitivity of the solution, since such a model functions as a 'black box'. Frequently, when optimal decisions are difficult or expensive to determine, it may be necessary to resort to heuristic methods which try to achieve a balance between computational effort and the quality of the solution.

Of course, it is possible to combine features of the two types of models. The feedback from an evaluative model can be used to modify the decisions and even the logic of a prescriptive model. Likewise, a prescriptive model can be used in conjunction with an evaluative model to generate insights into optimal policy behavior under different conditions. Also, one can build user options and evaluative modules into the prescriptive models. In this research, the focus will be on the development of a prescriptive model for determining 'optimal' due dates, although attempts will be made to incorporate evaluative features and generation of insights into system behavior.

## **1.5 Nature and Limitations of Previous Research**

One of the methods reported to provide for uncertainty in delivery and processing times is to add on a safety time to the mean lead time for each stage. This safety time is some multiple of the standard deviation of the lead time for that stage (Wilhelm and Ahmadi-Marandi, 1982, Walker and Wysk, 1983, Melnyk and Piper, 1985). The multiplier is a management decision variable that provides for an appropriate probability of the job completion at a stage within the total time provided. The effect of this kind of allowance is that safety times are built in at every stage. For long lines, these ac-

cumulated allowances can be substantial. This piecemeal approach may result in higher inventory levels than if a global solution method was used that minimized the overall cost, rather than provide a high probability of job completion at each stage.

Moreover, such safety allowances are a function of the relative magnitudes of costs and variabilities at the neighboring stages. It is unrealistic to expect the same pattern of costs to prevail all through the network. Therefore a fixed, across the board, safety factor for all stages may cause some errors as well.

The majority of published research in this area focuses on evaluative models. The analytical descriptive models of Wilhelm and his co-workers are useful in terms of the understanding they give of the small lot manufacturing process. None of them have attempted to optimize system performance. They do, however, provide the structure to be used for prescriptive studies as well. The simulation-based models in MRP related literature (St.John, 1985, Walker and Wysk, 1983, Grasso and Taylor, 1984, Melnyk and Piper, 1985) are primarily investigative in nature, although attempts have been made to use them to determine the decisions required. However, the quality of the decisions is not guaranteed, since only a restricted range of alternatives can be experimented with.

The prescriptive models developed by Yano (1987a,b) are restricted to simple structures such as serial and two-level networks. The serial line model can be extended to a multi-stage serial line but without part input. The two level network studied is the largest merging tree type problem that can be handled by the approach, and is confined to lot-for-lot procurement and assembly. Although detailed analysis has been done on convexity, the models are extremely difficult to extend to larger or more general assembly networks.

The models of Sarin and Das (1987) and Das and Sarin (1988) are for assembly systems with one part input at each stage, and use the normal distribution for delivery times, while processing times are assumed to be deterministic. In the multiple job problem they assume unlimited buffers between assembly stations. The Dynamic Programming methodology developed is dependent upon the normal distribution to derive closed form solutions for the optimal stage return, and to develop approximations for the transferred return. Extension to the random processing time case will require further approximation. In small lot assembly, the use of the normal distribution could cause some errors due to the significant probability of a negative tail for high coefficients of variation, and the fact that real life distributions tend to be skewed to the right. If non-normal distributions are considered, closed form solutions even for single stage optimization may not be possible to obtain, so the methodology may be difficult to extend. In view of the difficulty of extending the DP methodology to the multiple job problem, only heuristic solution can be attempted.

The nature of previous research and some of its limitations are summarized below.

1. The safety buffer rule adds on a multiplier of the standard deviation to the mean lead time to provide for an appropriate probability of completion of each stage. It does not distinguish between changes in costs and variances at different stages and cannot be optimal in all situations.
2. The analytical models of Wilhelm and his co-workers are evaluative in nature. While they may be used to generate an understanding of the process, they have not attempted to optimize system performance.

3. The simulation models in MRP literature are also evaluative in nature. They have been used in investigative studies to guide decision making, but such decisions cannot be claimed to be optimal in view of the limited choices that can be experimented with.
4. The models of Yano (1987a, b) are for simple situations such as serial and two-level systems with lot-for-lot manufacturing. They are extremely difficult to extend to multi-level systems with multiple jobs and part input at each stage.
5. The models of Sarin and Das (1987, 1988) are developed for the scenario of normally distributed part delivery times, deterministic processing times, and unlimited buffers between stations. They are difficult to generalize to non-normal distributions and random processing times. Also, limited buffers need to be considered in the multiple job problem.

The aim of this research is to redress some of these limitations in developing a prescriptive model for the determination of optimal mean part delivery dates. Distributions other than normal will be considered, and descriptive models for the other distributions will be developed if necessary. It will be attempted to develop a model and solution methodology that is general enough to incorporate different distributions. Random delivery times as well as processing times will be incorporated. In the multiple job problem, attempts will be made to achieve optimal solutions and to incorporate limited buffers between stations. A detailed description of the problem considered is given in the next section.

## **1.6 Description of the Problem and Objectives**

The assembly system being modelled operates according to the scenario described below. In the description, the term assembly job refers to a particular job produced by the system, while the term subassembly refers to the incomplete product as it progresses through the assembly stations. A schematic diagram of the system is shown in Figure 1. The statement of the problem considered follows the description of the operating conditions of the system.

1.  $J$  end products of the same type are to be produced.
2. After being launched into production, each assembly job progresses sequentially through a series of  $N$  workstations.
3. At each workstation a different part is assembled on to the subassembly. Upon completion of processing at a station, the resulting subassembly can be transferred with no delay to the next station.
4. Parts are supplied by external vendors or may be supplied in-house by branch lines. The parts are delivered separately for each job and each station, and their delivery times are independent random variables with known probability distributions.
5. Each station is a single server queue with queueing space for parts and subassemblies. Each assembly operation time is an independent random variable with a known probability distribution.
6. If the station is busy, the subassembly waits in the buffer ahead of the station. The buffer may be of limited capacity, in which case the subassembly may be blocked at the previous station if the buffer is full.

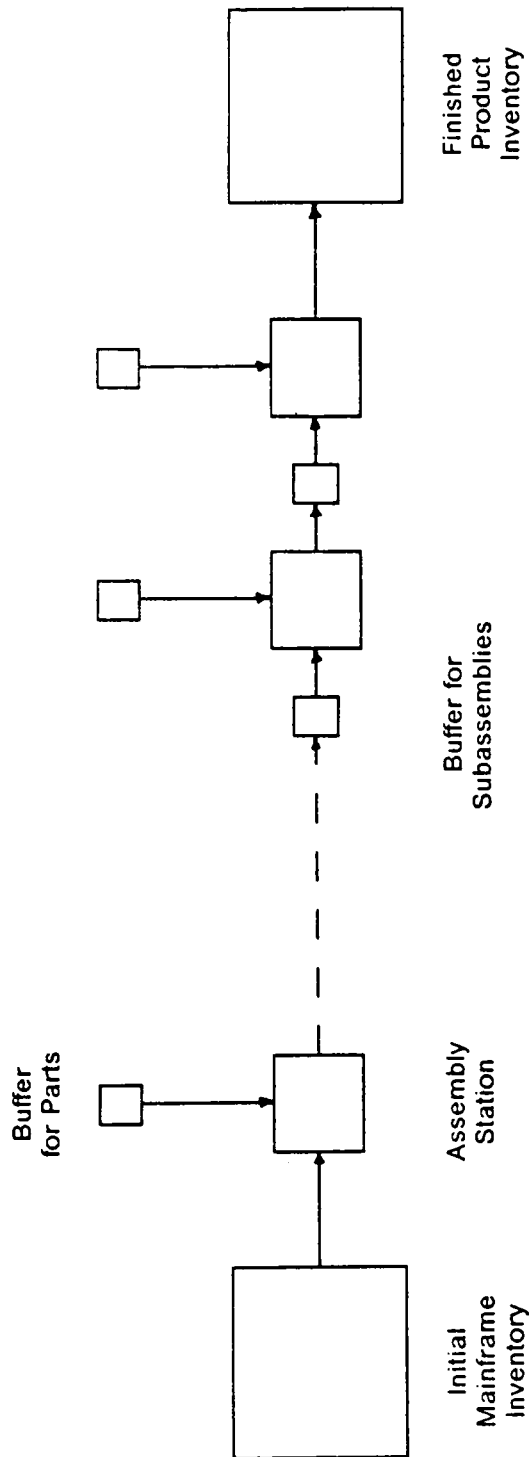


Figure 1. The Assembly System



7. Each part and subassembly stage have associated inventory costs. Due to randomness, parts and subassemblies may be early or late.
8. The completed assemblies are transferred to finished product storage, and are delivered to the customer as a batch. The finished products incur inventory cost while waiting for the batch to be completed.
9. There is a customer specified due date by which the batch must be completed. The batch incurs inventory cost if completed ahead of the due date, and tardiness penalty if completed beyond the due date.

If the delivery of parts and assembly operation times can be assumed to be completely reliable or deterministic, the required part delivery dates can be trivially obtained. However, in the case of small lot manufacturing these times cannot be assumed to be deterministic, and hence they cause random queueing delays and idle times for parts, subassemblies and finished products. The problem considered in this dissertation explicitly considers the trade-offs between these factors, and is stated as follows:

‘Given the stochastic nature of the arrival of parts at each station of an assembly line, the problem is to determine the mean planned delivery dates for the parts required at each station for each assembly job, in order to minimize the expected total of part and subassembly inventory cost, makespan cost, finished inventory cost and tardiness cost.’

The general scheme of the model can be adapted to fit a variety of scenarios such as flow shops with multiple parts, robotic workcells, component accumulation in kits, etc. Also, the structure can incorporate multiple part input at a station. It is assumed

that there are no lot-sizing issues involved and that capacities for part production are sufficient to meet the due dates determined by the model.

The objectives of this research are as follows:

1. Development of a solution methodology for determining the optimal part delivery dates when the actual delivery and processing times follow a known distribution. This will be done separately for the lot-for-lot or single job case and for the multiple job case.
2. Incorporation of different system configurations and additional features such as limited buffers in the descriptive and prescriptive methodologies.
3. Computer implementation of the solution methodology, and assessment of its performance compared to previously reported results and practice.
4. Investigation of the sensitivity of system behavior and optimal policies to changes in distributions, configurations and system parameters, in order to generate insights into system behavior that may be used to guide system design.

This dissertation falls in an important area facing researchers in Industrial Engineering and Operations Research - that of finding effective ways of managing operations in small lot manufacturing. It seeks to determine an optimal balance between work-in-process inventory, customer service, and utilization of facilities. It provides a means of effectively coordinating the procurement and manufacturing of assembled products. It will develop a methodology for the general problem that has not been developed so far. It additionally has extrinsic value due to its applicability to PERT networks, to optimize the costs based on the completion times of individual activities in addition to the aggregate project completion times.

This research represents an attempt to integrate different aspects of manufacturing systems analysis and design. It considers the product structure and processing requirements, and their effect on manufacturing and assembly requirements. It considers explicitly the uncertainties of small lot manufacturing and provides a means of dealing with the uncertainty. Finally, it attempts to combine descriptive modelling with optimization techniques to facilitate better managerial decision making.

## **1.7 Organization of the Dissertation**

The first chapter of this dissertation introduces the nature and relevance of small lot assembly manufacturing. The issues pertinent to the problem of determining lead times allowed for procurement and manufacturing are examined. An overview of the nature of previous research and its limitations is presented. The system being modelled is described and the problem being addressed is defined. The objectives of the research and its relevance are outlined.

Chapter 2 presents a review of the literature relevant to the problem area. This review is organized according to the nature of the studies and the approach followed. The major categories are: Descriptive Models of stochastic assembly systems, Simulation Based Models for investigation and determination of lead times, and Prescriptive Models for the determination of the due dates of parts to achieve some specified criteria. Related literature in the areas of Multi-stage Flow Shops with Limited Storage, PERT networks, and Assembly Shop Scheduling is also reviewed.

In Chapter 3 the basic model for determining the measures of system performance and deriving total costs is developed. The use of different distributions for part delivery and processing times is also examined. The relationships regarding the moments of a distribution that is the maximum of random variables pertaining to the normal and lognormal distributions are described, while similar relations for the gamma distribution are derived.

Chapter 4 describes the methodology that is used for the solution of the single job or lot-for-lot problem. The nature of the objective function is examined, and the motivation behind the solution methodology is explained. The implementation of the solution methodology is described along with its comparison with previously reported research and other methods.

Chapter 5 describes the generalization of the solution methodology to the case of procurement and assembly of multiple jobs of the same type through the same assembly facility. Extensions of the problem to incorporate other features are also described.

Chapter 6 describes the results from the application of the solution methodology to a variety of problem scenarios. The insights gained from the experimentation as to the behavior of the system under different operating conditions are presented.

Chapter 7 summarizes the results and achievements of this dissertation. The insights gained into the problem during the course of the research are summarized. Avenues for future research in this area are also identified.

## **2.0 Literature Review**

### **2.1 Introduction**

This chapter presents a review of the literature pertinent to the assembly systems design problem of setting part delivery dates. Although a number of papers have been published that develop models to describe the characteristics of an assembly network system, most of the papers pertaining to the setting of part delivery dates in such systems are simulation based. Only a few papers address the problem in the form of analytical prescriptive models. The literature review is organized in accordance with: descriptive models of stochastic assembly systems with part input, simulation based models and prescriptive models for setting part delivery dates in assembly systems. Related literature in the areas of multi-stage flow lines with limited buffers, PERT networks, and assembly shop scheduling is also reviewed.

## **2.2 Descriptive Models of Stochastic Assembly Systems with Part Input**

A relationship that is basic to describe the assembly of parts to a subassembly involves the maximum of two random variables, pertaining specifically to the arrival of the subassembly and the part to the station. It is therefore essential to compute the moments of the maximum of two random variables in order to develop a descriptive model of the assembly of parts to a subassembly. Some work presented in the literature in this regard is described first.

Clarke (1961) derived relations for the exact moments of the maximum of two normally distributed random variables with non-identical moments. He presented a recursive method for approximating the moments of the maximum of several normally distributed random variables by repetitive application of the equations, invoking the assumption that the maximum of two variables is approximately normally distributed. Although the distribution of the maximum is not normal, rather it is positively skewed as shown by Tippett (1925), Clarke showed that the error is small, but accumulates as the number of variables increases. These relations are discussed further in the next chapter.

Considering the positive skewness of the distribution of the maximum, Wilhelm (1986c) presented an application of the lognormal distribution as a model of ready times. The lognormal distribution is a transformation of the normal distribution, and incorporates finite lower support and positive skewness. The moments of the maximum of two lognormal variables were derived analytically for the special case of

common location parameter. Procedures for determining the appropriate correlations were also developed. An approximation procedure was described to handle the case of unequal location parameters. The model was tested for the component accumulation problem and for modelling a flexible manufacturing system. It was found that the lognormal model gives fairly accurate approximations for the maximum, especially for large coefficients of variation.

As another approximation, Wilhelm (1986d) has presented consideration of the  $S_v$  distribution, which is also based on a transformation of the normal distribution. Analytical relations for the moments of the maximum of  $S_v$  random variables for the two parameter and restricted four parameter distributions were derived. Relations for calculating the resulting correlations were also developed. An example application was also described of the component accumulation process in small lot assembly. The results showed that the resulting distribution of the maximum can be reasonably approximated by another  $S_v$  distribution with the computed parameters.

Rao (1975a) considered the problem of determining the mean production rate of a two-stage tandem production line with no intermediate storage, which is again a problem of determining the mean of the maximum of two random variables. Analytical expressions were derived for normal and Erlang distributions of processing times. Some special cases were worked out for different combinations of distributions and parameters. For the Erlang distribution, which is a special case of the gamma distribution, this requires evaluation of a finite Hypergeometric series. For the general gamma case, the Hypergeometric series is infinite but converges to a limit and can therefore be computed to the desired accuracy. However, the relations are limited to the first moment, and need to be extended to higher moments to enable use in the

problem at hand. Also, it must be confirmed if the maximum of two gamma variables can be approximated as another gamma variable.

In light of the fact that the distribution of the maximum of two random variables is not of the same form as the constituent variables, Sculli and Wong (1985) described a study to explore the errors involved in approximating the maximum and the sum of two independent Beta random variables as Beta variables for the analysis of PERT networks. Since closed form relations are not available, a numerical approach was used for evaluation of the moments, which can be matched to derive the parameters of the resulting distribution. The Kolmogorov-Smirnov test was used to test the accuracy of the approximation, by treating the values obtained from the approximating distribution as random samples from an unknown distribution. It was seen that the sum of two Beta variables can be approximated by another Beta variable with small and almost constant error. The actual distribution of the maximum, however, is skewed to the right, but the approximation is reasonable if one of the distributions dominates or when the variances do not differ considerably. The approach presented in this paper will be used later in the study of similar approximation errors for the distributions being considered in this research.

Sculli (1983) used Clarke's method to approximate the moments of PERT networks with independent, normally distributed activity durations. Additionally, the different paths were assumed to be independent, and the distribution of the maximum of two normal random variables was approximated by another normal with the computed parameters. The approximation was numerically shown to yield better estimates than the PERT approach based on the critical path alone.



Considering now the development of the descriptive models, Wilhelm and Ahmadi-Marandi (1982) first developed a model to describe the operating characteristics of small lot assembly systems, using Clarke's relations to determine the moments of operation start times. The assembly system that they considered consists of a set of stations through which each mainframe progresses, a different part being added on to the mainframe at each station and the resulting subassembly transferred to the next station with no delay. Unlimited queueing space was assumed to be available for parts and subassemblies. The model was based on the fundamental relation that the start time for a job at a station is the maximum of the finish time of the preceding job at that station, the preceding operation for that job, and the arrival of the requisite part. Relations were developed for earliness and tardiness of jobs. The concept of metering stations was introduced to identify the relative bottleneck operations that control the flow of work along the line. Different launching intervals for the mainframe were examined in the deterministic case with respect to different performance criteria. In the stochastic case, delivery times were assumed to be normally distributed and a management decision factor, in number of standard deviations, was used to provide safety buffer time beyond the mean lead time. A seven-step procedure was described to enable recursive computation of the correlation coefficients required to implement Clarke's method.

Wilhelm et al. (1986) extended the model of Wilhelm and Ahmadi-Marandi (1982) to incorporate more features and enable application to more general situations. They included the cases of random processing times, correlated component ready times, parallel machines at a station, and consequent redefinition of the metering station concept. The approach developed in the above papers was generalized by Saboo and Wilhelm (1986) to the study of transient performance in multi-level assembly

networks, where the input of parts to an assembly station may itself be the output of a branch line. Performance times were again assumed to be normally distributed, and Clarke's method was used for determining the moments of finish times. Computational procedures were developed for estimating the correlations between operation finish times. Some properties of these correlations were identified to develop bounds for expected start times.

The descriptive methodology used for assembly systems can also be adapted to model the operating characteristics in other manufacturing scenarios such as flowshops and robotized work cells. Wilhelm (1986a) described an adaptation of the descriptive model to approximate transient performance of the flowshop where operation start and finish times are related by the multivariate normal distribution. Fundamental relationships among the correlations of various start and finish times were derived to yield bounds and estimates of operation start times. Wilhelm (1986b) presented an approach for analysis of transient material flow in robotized manufacturing cells, where various operation times are normally distributed random variables, and all parts are available for processing at time zero. The analysis included two machines and a robot for part handling from and within the cell. Although the availability of a machine for processing is determined by the minimum of the current processing times, the analysis was based on selecting the machine with a greater probability of earlier completion. Procedures were developed for computing the necessary correlations. Modifications to adapt the approach to incorporate other features, such as more than two machines, alternate cell configurations, layout, robot design etc., were also discussed.

Wilhelm and Wang (1986) presented a model for managing the component accumulation process in small lot assembly systems. Accumulation time was defined by the maximum of all part ready times, and relations were derived as before for estimating its moments using Clarke's method. Analytical relations were derived for the various measures of performance of the accumulation process such as component waiting times, kit earliness and kit tardiness. The model was used for sensitivity analysis of system performance to different parameters. A safety factor, in terms of number of standard deviations of the lead time, was used to determine the lead time allowance for the best trade-off between earliness and lateness costs.

Harrison (1973) described a queueing theoretic model of an assembly operation, consisting of a multi-input process of assembly parts and a single server performing the assembly operation. Part arrival times and processing times were general independent random variables. This scenario is similar to the model of Wilhelm and Wang (1986) in that all accumulation is assumed to take place before assembly starts. Limit theorems were derived for appropriately normalized versions of the waiting time process under various load conditions. The approach is restricted to one station assembly, and extension to the multi-station case would be an elaborate exercise in queueing network theory.

The models that have been discussed in this section are descriptive in nature. They enable analytical description of the process of small lot assembly, and allow the determination of the measures of system performance for a given set of decisions. As such, they do not directly yield decisions and fall into the general category of evaluative models (Suri, 1985). However, the models of Wilhelm and his co-workers do provide the framework necessary to develop prescriptive models.

## **2.3 Simulation Based Models for Planned Lead Times in Assembly Systems**

The determination of mean delivery dates for parts at necessary stations along the assembly line is equivalent to the problem of determining planned lead times for procurement and assembly. The planned lead time is the total of the mean lead time and an appropriate safety time, and is one of the most important inputs to MRP systems. It determines the appropriate times for initiation of necessary procurement or processing, and affects the overall inventory cost and delivery performance. The models in the literature for study of planned lead times in stochastic MRP systems are generally simulation based, and investigative rather than prescriptive in their focus. These models are discussed in this section.

The effect of inflated planned lead times on overall costs in MRP systems was studied by St. John (1985) using a simulation model. The study assumed normally distributed demand, constant purchased part lead times, and that processing times follow the triangular distribution. The planned lead time values tested ranged from zero slack to 95% slack time. The resulting response was analyzed by regression analysis. The costs were seen to increase rapidly with increase in allowed slack time, with exponential increases beyond the 80-90% level.

Grasso and Taylor (1984) used a simulation model to investigate the effect of variability in the timings of deliveries on the performance of MRP systems. The experiments were conducted using a five level product structure with different levels of lead time variability, buffering alternatives, lot size rules, and holding and penalty

costs. System performance was found to be significantly affected by the amount of safety lead time. The amount of variability in lead times also affected the choice of lot sizing rule. The lead times therefore need to be managed effectively to reduce variability and to prevent increases in cost. The lot-for-lot rule was found to be the preferred lot sizing method, especially when lead times were highly variable. Changes in holding and penalty costs also affected the choice of buffering alternative, so efforts must be made to assign realistic values to the costs.

Melnyk and Piper (1985) studied the interaction between lead time decisions and lot-sizing decisions, and obtained measures for the lead time error effects of five lot-sizing rules. The study was conducted using a SLAM simulation model to study the effect of lot-sizing and lead times on setup costs, late deliveries and component lead time errors. The results showed that lead time allowances influenced lot-sizing effectiveness, and increase in safety multipliers improved end item service levels. Likewise, lot-sizing influenced lead time allowance effectiveness, and the lot-for-lot rule produced the tightest error distributions and the best overall performance.

Walker and Wysk (1983) presented a study to identify the least cost purchased-part planned lead time strategy for items in an MRP system. A simulation model was used to explode and time phase requirements, generate capacity, simulate the shop production of the schedule, and record performance. The lead time for all parts was taken to follow the lognormal distribution, with a mean of 3 weeks and a standard deviation of 1 week to coincide with the MRP time-bucket unit. Flat, multi-level and mixed product structure types were tested. Other conditions included no safety stocks, lot-for-lot procurement priority setting by due date, and an 80% shop utilization rate. The performance criterion was the sum of average inventory costs and

lateness penalty costs. The model was run for planned lead time values in increments of one standard deviation, starting from the mean lead time, and measuring the total cost for each level of lead time.

For the flat product structure, a lead time allowance of 1 standard deviation beyond the mean yielded the least cost, whereas for multi-level structures the best strategy was to use the mean lead time itself. For mixed structures, the best strategy was to use the mean lead time for low-level parts, and add on a 1 standard deviation allowance for level one parts. Thus, the amount of allowance depends on the level of the part in the assembly. This model is the closest in objective to the problem being studied in this research. However, only discrete, across the board levels are experimented with so optimality is not assured.

The planned lead time including the safety factor provided also has a direct relation to the number of Kanbans required in implementation of the JIT technique as shown by Philipoom et al. (1987). A simulation model was constructed for a prototype shop to determine the planned lead time and the number of Kanbans necessary to prevent backorders. Investigation of the factors showed that the number of Kanbans required increased as the coefficients of variation, the utilization levels in the shop and the positive autocorrelation between jobs increased.

In a related study, Rees et al. (1987) described a simulation based procedure to dynamically adjust the number of Kanbans in a JIT production system in an unstable environment. The idea was to estimate the density function of the lead time using simulation and combine it with the forecasted demand to produce the probability mass function for the number of Kanbans. Given the holding and shortage costs, the optimal number of Kanbans can then be determined. This value can then be imple-

mented and the process repeated until convergence is obtained. The methodology was shown to adjust very quickly to incorrect initial specification of the number of Kanbans.

Simulation models of the type discussed in this section also belong to the class of evaluative models, in that they are used to obtain measures of system performance for a given set of decisions. They are useful if the complexity of the system being modelled does not permit analytical description. Also, in the context of the assembly problem, they do not require any approximations or assumptions about the distributions of the sum or the maximum of random variables that are necessary for analytical models. However, it is difficult to be certain about the quality of the solution obtained using simulation models, since only a limited number of alternatives and factor levels can be experimented with.

## **2.4 Prescriptive Models for Setting Part Due Dates in**

### **Assembly Systems**

So far, the review has covered the areas of analytical descriptive models and simulation models, both of which are primarily evaluative in nature. In this section, the prescriptive models in the literature pertaining to the setting of due dates in assembly systems are reviewed. Some of these models are for simple structures such as serial lines or where all part accumulation takes place at the first station prior to assembly. Very few models incorporate multi-level assembly with part input at different stations along the line, as is necessary in the problem being considered.

Jensen and Khan (1972) addressed the issue of determining the optimal start-up and shut-down schedule of a multi-stage production system to minimize the total of inventory and start-up costs in a deterministic system. Accumulation of raw materials was assumed to take place prior to the first stage. Dynamic Programming was used to determine the cycle times and production rates at adjacent stages.

Bigham and Mogg (1981) enriched the above model to include the expected costs of earliness and lateness, due to stochastic deliveries at each stage, in the cost function. Production rates were assumed fixed, and finished goods could be back-ordered at a fixed penalty. Materials were input at each stage but the entire requirement was assumed to arrive simultaneously, making it equivalent to the Jensen and Khan model in this respect. Also, production was permitted to start only at the scheduled time even if all materials arrived earlier - a system of the forbidden early departure type (Kanet and Christy, 1984). With deterministic processing times, this causes no risks, but in the case of random processing times, this increases the probability of backorders. Other assumptions were that backorders exist for a negligible fraction of the time, probability of excessive lateness of an order is zero, and the cycle time at stage N is at least as great as at stages 1 to N. The objective function was reported to be convex, and an iterative method was used for solution, since direct, simultaneous solution of the equations for the first order conditions is not possible. In view of the restrictions, this model is limited in its applicability.

Lee and Rosenblatt (1985) examined the effect of in process inventory on the design of assembly lines. An analytical framework was developed for incorporating consideration of value addition along the line, and the resulting inventory holding costs, in the solution of line balancing problems. A branch and bound procedure was



developed for grouping the tasks into stations, in accordance with precedence relationships, to minimize the overall cost. The procedure is designed for deterministic assembly lines producing in volumes large enough to justify balancing, and as such addresses a different scenario from the problem at hand.

Yano (1987a) presented an analytical treatment of the problem of determining optimal planned lead times, including safety times, in a serial production system where a fixed lot size is processed at each stage, and the actual lead time follows a known distribution. The delivery of parts can be interpreted as being the first stage in the system. Each order has a specified due date with tardiness penalty, and batches finishing a stage early are held back till the scheduled start time of the next stage. This is also a forbidden early departure system that enforces idle time if a stage finishes early. The objective was to minimize the total of expected holding and tardiness costs.

The objective function for the two stage system was shown to be convex for certain cost combinations, and quasi-convex for most others. The optimal solution for the last stage is a simple 'newsboy' problem, as shown by Weeks(1981) in his model for single stage systems with inventory and tardiness costs. The algorithm described first determines the optimal value for the last stage, and then the optimal value for the first stage by one-dimensional search. If the first stage value does not fall within the acceptable range of probability, the first stage lead time is set to zero, and the final stage lead time is found by collapsing the two stages together. The approach was also extended for the general multi-stage serial case. The method can be used to investigate the sensitivity to changes or errors in estimation of the mean, variance and distribution at each stage, without having to resort to simulation.

It is claimed that the forbidden early departure policy is more effective than immediate dispatching, except when inventory costs are almost constant across stages and/or lead times at initial stages have low variances. However, it is also suggested that there are advantages in reducing the number of stages by merging them, which is equivalent to immediate dispatching at each constituent stage. A forbidden early departure system results in adding safety time at each stage, thus inflating the total allowed time compared to the immediate dispatching case, where the allowances are merged.

Yano (1987b) described an analytical approach to the problem of determining optimal planned lead times in a two level assembly system. The model was for a product assembled from two components, with assembly being carried out after both parts arrive. Part delivery times and the assembly time are random variables. As in the series system, forbidden early departure is assumed for the assembly operation. The final product has a specified due date with tardiness penalty. Lot sizing is assumed to be done on a lot-for-lot basis. The objective, as before, was to minimize the total of expected holding and tardiness costs.

It was found that the objective function was not convex in all variables. However, it was shown to be jointly convex in the planned lead times for the parts for a fixed value of assembly lead time, while it was convex in the assembly lead time for fixed part lead times. The solution method suggested was to find the optimal values of the part lead times for each candidate value of assembly lead time, and determine the solution iteratively. Some numerical insights were also presented on the effects of changes in lead time parameters, using the Poisson and Negative Binomial distributions.

It is stated that the objective function for the n-component case is likely to be even less well-behaved, and even the objective function and first order conditions are not easy to derive. This approach thus has severe limitations as to the size of the problem and its ability to incorporate multi-level assembly. The difficulty in analysis arises primarily out of the enforced idleness at each stage. Immediate dispatching, on the other hand, permits extension of the formulation to the multi-stage case with relatively less difficulty.

Sarin and Das (1987) developed an approach to determine optimal part delivery dates in single job assembly with one part input at each stage. Delivery times were taken to be normally distributed random variables, while processing times were deterministic. The model was based on the descriptive model of Wilhelm and Ahmadi-Marandi (1982), where the start time of an assembly operation is the maximum of the finish times of the previous operation of that job, the previous job at that stage, and the arrival of the part required at that stage. This corresponds to the immediate dispatching case, where there are no enforced delays. The moments of the start time were determined using Clarke's formula for the maximum of two random normal variables, the resulting distribution being approximated as another normal. The objective was to minimize the total expected cost of part and subassembly waiting over all the stages.

A Dynamic Programming model using backward recursion was developed with each assembly station corresponding to a stage, the decision variable for each stage being the mean due date for the part required at that stage. The objective function for each individual stage was determined by modelling the waiting time as the difference between two normal random variables, which is also normal, and computing

the partial expectations. Optimization of the stage return with respect to the stage variable yields a closed form result for the optimal policy and the optimal stage return in terms of the variances and other parameters. For multi-stage optimization, the return transferred from previous stages must also be included in addition to the individual stage return. The derivative of this transferred return proved to be of an extremely tedious form, and a regression model was employed to approximate it in a form similar to the stage return. The model can then be applied recursively to each stage to determine the optimal policy for each stage. The generalization of the model to N stages, and modifications necessary to incorporate customer due date and cost of setting part delivery dates were also described.

The use of the normal distribution enables optimizing the individual stage return in closed form. The regression model is developed using random values from specific ranges for the means, variances, and costs. It has to be re-run if the actual ranges of values anticipated are significantly different from the ones used to develop the model. The total return function even for the two station problem is not guaranteed to be convex or unimodal. However, the approximation is convex under certain conditions which were generally found to hold during numerical testing.

Das and Sarin (1988) considered the problem of setting mean due dates for parts required at different assembly stations when a set of assemblies is produced sequentially at the stations, and delivered to the customer as a batch. Part delivery times were taken to be normally distributed, while processing times were deterministic. The availability of the station from the previous assembly must now be included in determining the distribution and moments of the start time. Also, the procedure for determining correlations as outlined by Wilhelm and Ahmadi-Marandi

(1982) must be used to determine moments of start and finish times, since successive assemblies are not independent. All assemblies were assumed to require identical processing, but the parts required for at a particular station for different assemblies were taken to be distinct. There were assumed to be unlimited buffers between stages. The objective function was the sum of all subassembly and part waiting costs, the facility cost determined by the makespan, and finished product holding and tardiness costs.

The DP approach used for the single job problem was difficult to extend to the multiple job problem because of difficulties with the dimensionality of the problem. Also, correlations need to be determined at each step for application of Clarke's method. Consequently, a heuristic method was employed for piecemeal optimization using the DP procedure developed earlier. The application of the heuristic was based upon determining the critical path through the network of stages. This critical path was identified as the one passing through the bottleneck station, termed the first order metering station. The DP procedure was first applied to this path, and the part due dates determined accordingly. Any path before the bottleneck that was closely critical was optimized using the DP procedure, creating a slack to ensure that it did not interfere with the critical path.

The remaining stations before the bottleneck were optimized one job at a time using the DP procedure. To ensure that the chain being optimized feeds the critical path, the assembly start time of the appropriate node on the critical path was treated as the due date for the chain. Stages downstream of the critical path were scheduled as late as possible using similar single job optimization, starting with the station with the largest operation time downstream of the bottleneck, and using previously de-

terminated assembly start times as prescribed due dates for the chain being optimized. The procedure is applied until all stations and jobs have been scheduled. The expected system cost was determined using simulation, and the results showed the heuristic method to be superior to the one standard deviation safety buffer rule, although the difference declines as the number of near-critical stations increases.

The research in the area of multi-level assembly systems with part input at different stations is therefore limited. The models of Yano (1987a,b), though analytically well developed, are for simple structures and are difficult to generalize to the problem on hand. The models of Sarin and Das (1987, 1988) are restricted to normally distributed part delivery time, deterministic processing times and unlimited buffers between stations. All aspects of the problem being considered have not been researched yet in prescriptive models.

## **2.5 Multi-stage Flow Lines with Limited Storage**

In the latter part of this research, the issues of scheduling of assembly lines with multiple jobs will be addressed. The imbalances and variability in small lot stochastic assembly systems cause queues to form. Since it is unrealistic to assume unrestricted buffer space between stations, the relationships between start and finish times of operations can get affected due to the possibility of blocking and starving of stations. The literature in the area of production lines with limited buffers was reviewed to understand the effects of limited buffers, and examine methods of incorporating limited buffers in the analysis, and providing for them in system design.

Koenigsberg (1959) identified the number of stages, location of buffers and buffer capacities as important parameters in flowline design. Hillier and Boling (1967) developed procedures for exact Markov chain analysis of small lines and approximate procedures for small systems. They demonstrated the existence of the 'bowl' phenomenon, where the plot of the optimal allocation of processing time versus line position is bowl shaped. Buzacott (1967) developed models of transfer lines with buffer stocks applicable to several configurations and described how they might be used to specify buffer stock levels.

Rao (1975b) used the results in Rao (1975a) to develop a recursive methodology for determining the effect of intermediate storage on the mean production rate of a two-stage production line. The model was restricted to exponential service at the first stage, with the second stage service being normal or gamma distributed. The effects of unbalancing of the line were also examined, it was found that balanced division was no longer optimal if the variabilities were non-identical.

Altiook and Stidham (1983) used a queueing theoretic model for the allocation of buffer capacity in flow lines where stages are subject to breakdown. Altiook (1985) described a Markov chain approach to obtain performance measures of flow lines with limited buffers with phase-type service and exponentially distributed times to breakdown.

Wilhelm (1986a) described a procedure to incorporate limited buffers between stations in the basic descriptive model developed in Wilhelm and Ahmadi-Marandi (1982). Since the start times of jobs can now also get affected by previous jobs at later stations, the model includes additional restrictions to capture the dependency of activity start times on previous jobs at later stations. The representation retains the

same Activity-on-Arrow structure, but has additional precedence relationships incorporated in the evaluation of the maximum operator. A nested computational procedure was described for determining the various correlations. The method can be used to evaluate the performance of an assembly system with specified buffer capacities.

An alternative network modelling approach was developed by Muth (1984) to study the effects of blocking and starving in unbuffered stochastic serial production lines. Part input at stations was not considered. The line was represented as an Activity-on-Arrow network, with nodes representing start times. Some special cases for exponential and Erlang service time distributions were worked out for lines up to 4 stations long. However, the issue of dependence of waiting times was not resolved. Non-zero buffer capacities can be incorporated by adding on a dummy activity with zero processing time for each buffer position. Part input can be included by adding additional nodes to separate subassembly arrival at a station and the start of assembly. With these modifications this representation becomes equivalent to the Activity-on-Arrow representation in Wilhelm (1986a). However, the problem size is considerably larger due to the number of dummy nodes required, and becomes unwieldy when buffer capacities are large or unrestricted.

Yamashina and Okamura (1983) used a simulation model to study the role of buffer stocks in a multi-stage transfer line incorporating station breakdown and station repair. They studied the effects of the number of production stages and variation in production rates on the output of the line. It was found that uniform allocation was not optimal even for balanced lines, and that the need for buffer stocks increases as



the number of stages increases. Based on the studies of line performance, some guidelines were developed regarding the allocation of buffer stocks.

In an exhaustive simulation study, Conway et al. (1988) investigated the role of work-in-process inventory in serial production lines. They studied the effect of buffer stock and its allocation on the production rates of balanced lines. They also examined the allocation of buffer stocks in lines with unequal variabilities, unbalanced lines and unreliable stations. It was found that buffers provide less increase in capacity in unbalanced lines, and that the preferable allocation is displaced towards the bottleneck stations.

Kraemer and Love (1970) developed a queueing theory based model for optimizing buffer storage levels in sequential production flowlines. Smith and Daskalaki (1988) combined queueing network analysis with classical optimization in a study of optimal allocation of buffer spaces in automated assembly lines with exponentially distributed service times. The assembly lines were modelled as finite open queueing networks, and the performance of the system was optimized using Powell's unconstrained optimization procedure. The methodology was used to evaluate line topologies, system throughputs and optimal buffer sizes for the series, merge and splitting configurations.

In this research, the concern is to determine optimal part delivery dates when there are limited buffers between stations. Although the optimal placement of buffers is not the objective, the results and cost values can be used to guide the allocation of buffers. The models discussed in this section provide some insights as to the role of buffers in the design of flow lines, and will be useful in the development of heuristic procedures. The model of Wilhelm (1986a) for handling buffers appears to be the most

convenient to use in the context of the problem being considered, and is also compatible with the basic descriptive model discussed earlier.

## **2.6 Related Research**

In this section, some related literature is reviewed in the areas of activity networks, resource constrained project scheduling, and assembly shop scheduling. While studies in these areas address somewhat different scenarios, the problems considered share some characteristics with the problem at hand. The literature has been examined to see if any of the approaches can be directly or indirectly applied to this research.

The structure of the descriptive models developed by Wilhelm and Ahmadi-Marandi (1982) is essentially that of Activity-on-Node networks, and part delivery times can also be viewed as activities in an enlarged network. It was therefore investigated if the methods in activity network literature could be used for the development of descriptive and prescriptive models. Elmaghraby (1977) provides a comprehensive review of the basic methods and models used in analysis of such networks.

Studies of Probabilistic Activity Networks (PAN's), also known as PERT networks, are directed at determining the mean project completion time or the probability distribution function of the project completion time. Dodin (1985) classifies these studies in four main categories: (i) Analytical studies to approximate or bound the mean

project completion time, (ii) Analytical procedures to calculate or approximate the distribution function, (iii) Monte Carlo sampling procedures to approximate the distribution function, and (iv) analytical procedures to bound the distribution function.

The PAN methods have been developed for general network structures. As such, they do not exploit the stage-wise structure of the assembly network problem. Moreover, for the development of the prescriptive model we require moments and distribution functions, not only for the overall project completion times, but also for the start and finish times of individual activities. The PAN methods are too cumbersome to apply for this purpose. Methods based on evaluation of the maximum and the sum of random variables, on the other hand, enable determination of these measures by approximating the resultant distributions of each finish time as another distribution of the same form.

The literature in the area of resource constrained activity network scheduling was also examined to see if any of the techniques are relevant to the problem at hand. Reviews of the methods used in this area are provided in Davis (1973) and Elmaghraby (1977). Russell (1983) reviews the application of these techniques in the manufacturing environment. Procedures for optimal scheduling in resource constrained networks are computationally intensive, and have not proved feasible for large, complex networks (Davis, 1973). Consequently, several heuristic procedures have been developed to enable obtaining good solutions to such problems. Some of these heuristic solution approaches are compared in Davis and Patterson (1975). Such methods use a variety of decision rules based on factors such as resource demand, activity duration, criticality of activities, and total slack.

These traditional approaches to scheduling in resource constrained project networks have concentrated on deterministic networks with non-storable resources such as manpower and machines. The concern in the problem at hand, however, are stochastic networks with storable resources in the form of materials, assembly parts and subassemblies.

Aquilano and Smith (1980) developed CPM-MRP, a set of algorithms for scheduling such networks subject to material lead times and inventories. Smith and Aquilano (1984) extended the previous algorithm to include constraints of non-storable resources as well. Smith and Smith (1987) presented a mixed integer programming formulation for determining a schedule of project activities and materials orders that optimizes the total of activity, ordering, holding and penalty costs. The algorithms described above are, however, restricted only to networks with deterministic activity durations. No previous research was found that combined scheduling of stochastic networks with ordering of storable resources.

Assembly shop scheduling is concerned with development of scheduling and dispatching methods in job shops where individual jobs may be made up of multiple components, and the operations required follow a converging tree structure. In such shops, scheduling requires consideration of queueing times at work centers, as well as staging delays arising out of items having to wait for the remaining parts before assembly can commence. The assembly network structure of this problem, and the objective of reducing queueing and assembly delays, is similar to the problem being considered in this research. The research in this area was investigated to examine if some of the ideas and methods could be utilized.

Trilling (1966) developed a framework for simulating job shops where orders use varying combinations of parts and routings. This necessitates the use of network structures, and a binary coding scheme is developed for representation of the network. Maxwell and Mehra (1968) developed and tested simple and composite dispatching rules in an assembly environment in which jobs have a tree structure of operations. Factors considered in the investigation included operation slack, processing time, operation urgency and precedence constraints. It was found that composite rules that combined operation and job status performed better than single factor rules, and that higher allowances were necessary in setting due dates for assembled products to provide for staging delays.

Sculli (1980) investigated priority dispatching rules for a job shop where jobs are made up of parts, requiring several operations on machine centers, and eventual assembly. The study was directed towards rules that attempt to coordinate the completion times of the constituent parts. This mainly involved rules utilizing job status information, such as operation float based on the critical path in the assembly network, number of parts completed, and number of operations remaining in each part. The rules incorporating job status information consistently gave the best results over all criteria and significantly improved performance.

Goodwin and Goodwin (1982) described a simulation study of the relative impact of different scheduling policies on the performance of a multistage assembly system. The study examined rules for job release date determination, sequencing priorities, and regeneration of priorities. Combining decision rules to form policies was found to have a strong effect on performance. Goodwin and Weeks (1986) further evaluated scheduling policies comprised of dispatching and regeneration rules in a multi-level

assembly system. In addition to the traditionally used expected value preference ordering, second-degree stochastic dominance ordering rules were used to identify the most efficient scheduling policies.

Huang (1984) conducted a simulation based investigation of priority decision rules in a hybrid assembly/job shop, manufacturing both single-component and multiple-component parts. A combined priority rule was tested that considered the different types of jobs in making dispatching decisions at machine centers, giving priority to jobs requiring assembly. The rule consisted of implementing the Assembly Jobs First-SPT rule for machine centers processing both types of jobs, and SPT at other machine centers. The rule was found to reduce staging time, and also resulted in the least variation in flow time when there were more assembly jobs.

Adam et al. (1987) explored the means by which job structural complexity can be incorporated explicitly into the design of priority assignment procedures to reduce lead times in multi-level assembly job shops. Pacing rules, such as the Relative Number of Remaining Operations and Relative Remaining Processing Time, in conjunction with the Total Work Remaining rule, were found to yield significant improvement in staging delay compared to previous rules in literature.

Russell and Taylor (1985) evaluated scheduling policies for the production of assembled products in an assembly job shop, with labor as an additional constraining resource. Two different product structure sets were considered, and the scheduling policies examined included due date assignment, labor assignment and item sequencing rules. A SLAM II model was used to determine if these factors individually or in combination significantly affect the root mean square of job tardiness. All the factors were found to influence job tardiness, while product structure was found to

affect due date assignment, labor assignment, and sequencing decisions. Labor assignment rule decisions further affect the sequencing decision, while the due date assignment rule does not significantly affect the other two decisions. Tall structured jobs were found to be more sensitive to the choice of scheduling policies. Policies which performed well regardless of product structure or other scheduling policies included the branch slack combination sequencing rule, the longest path due date assignment rule, and the longest queue labor assignment rule.

Fry et al. (1986) examined the effect of product structure and sequencing rules on the performance of an assembly job shop. Considerable variability in performance was found between the different product structures, with the taller structures being more tardy than the flat ones. The performance of the due date oriented rules was found to improve as BOM's became taller, and the earliest due date rule was found to perform consistently well for all types of structures and performance measures. A rule combining SPT with product structure information, and giving a higher priority to operations higher up in the BOM, performed well on all criteria. In a related paper, Fry et al. (1986) examined rules for the assignment of due dates or job release dates in a multistage job shop. Rules considering product structure and shop condition additively performed better than rules considering only one of the factors, or considering them multiplicatively.

Fry et al. (1988) evaluated the performance of selected dispatching rules in a multistage job shop where machine capacities are unbalanced, to determine whether the performance is consistent for jobs routed through the bottleneck and jobs that bypass the bottleneck. It was found that performance varied significantly, depending on whether the job required the bottleneck station, especially in tardiness and as-

sembly delay measures. The due date based rules perform well for most tardiness measures for both types of jobs. The best rule for flowtime for both types of jobs was found to be the Number of Uncompleted Branches rule, using SPT to break ties. The Branch Slack Rule also performed well for all tardiness measures for jobs routed through the bottleneck, and for most measures for jobs bypassing the bottleneck.

The studies in assembly shop scheduling, while sharing some features with the problem being considered in this research, address a different scenario. They are not directly concerned with developing optimal solutions, and are evaluative rather than prescriptive in their focus. All of them use simulation to evaluate the results of the scheduling policies considered. However some of the ideas used, such as bottleneck station, critical path, operation slack, and number of remaining operations may be useful in the development of approximate/heuristic solutions to the problem at hand.

## **2.7 Summary**

The literature pertinent to the problem area has been reviewed in this chapter from different viewpoints. The main features of the different models used to study stochastic assembly systems and related areas have been identified. The thrust of research in the area of stochastic assembly systems with multiple part input has been in the area of evaluative modelling. The models of Wilhelm and his co-workers develop detailed methods for analytical description of the small lot assembly process. The related research in studies of planned lead times in MRP systems has been simulation based. As such, these models have not been used for determining opti-



mal decisions, though attempts have been made to obtain decisions from among a restricted range of alternatives. The method that is generally experimented with in the literature, to determine part delivery dates or planned lead times, is to allow safety time to be some multiple of the standard deviation of the part arrival process.

Relatively few publications have addressed the issue of determining optimal due dates and planned lead times in stochastic assembly systems. Yano's models are limited to serial and two level structures and assume lot-for-lot procurement and assembly. They are extremely difficult to extend to larger, multi-level networks with differentiation between jobs. The models of Sarin and Das are for the restricted scenario of normally distributed part arrivals, deterministic processing and finite buffers. In the remainder of this research it will be attempted to enrich the scenario by considering features not included in prescriptive models so far, such as non-normal distributions, random processing times and limited buffers between stations.

## **3.0 Cost Model Development**

### **3.1 Introduction**

This chapter describes the development of the general model for computing the total expected system cost of the assembly system with random part arrival and processing times. The model is based on the descriptive model developed by Wilhelm and Ahmadi-Marandi (1982). The notation and the basic relationships used in developing the descriptive model are stated. Three distributions, namely normal, lognormal and gamma are used to represent part delivery and processing time distributions. The statistical relationships and background necessary for application of the general framework to these distributions are presented and derived where necessary. In this chapter, it is assumed that buffer capacities between stations are unlimited. The changes necessary in the model to incorporate limited buffers between stations will be dealt with in Chapter 5.

## 3.2 Notation

The various terms and notation used in the model development are as follows:

- $D_{ij}$  = Delivery time for part required for assembly job j at station i
- $P_{ij}$  = Processing time for assembly job j at station i
- $S_{ij}$  = Operation start time for assembly job j at station i
- $F_{ij}$  = Operation finish time for assembly job j at station i
- $F_{0j}$  = Arrival time of initial assembly job j to the system

The corresponding means and standard deviations of the times defined above are represented by prefixing the variable name by  $\mu$  and  $\sigma$  respectively. The expected arrival time of the first assembly job to the system is  $\mu F_{01}$ .

- $f(x)$  = Density function of random variable X.
- $L$  = Lower support of the random variable.
- $U$  = Upper support of the random variable.
- $Pr(X \leq t) = \int_L^t f(x) dx$  = Cumulative distribution function of random variable X.
- $E(X) = \int_L^U xf(x) dx$  = Expected value of random variable X.
- $E_i(X) = \int_L^t xf(x) dx$  = Partial expectation of random variable X over the range from the lower support to t.

- $T$  = Customer specified due date for the job or batch
- $CE_{i,j}$  = Holding cost per unit time of part required for assembly  $j$  at station  $i$
- $CS_{i,j}$  = Subassembly holding cost per unit time for assembly  $j$  at station  $i$
- $CI_{i,j}$  = Idleness cost per unit time of station  $i$  due to assembly  $j$
- $CF$  = Finished goods holding cost per unit time
- $CT$  = Tardiness cost per unit time for finished assembly
- $CM$  = Makespan cost per unit time

The subscripts on the variables are as follows

- $i = 1, 2, \dots, N$  is the index for assembly stations
- $j = 1, 2, \dots, J$  is the index for assembly jobs

It is assumed that the subassembly holding cost changes once processing starts, since value is added to it due to the assembly of parts to it. It is expected that the subassembly holding cost at a station is at least as much as the sum of the holding costs of the previous subassembly and the part added to it at the previous station. Since all the jobs are delivered as a batch, the makespan is defined as the time between the start of the first job of the batch and the completion of its last job. This can also be referred to as the flowtime of the batch. The makespan cost comprises of the equipment, labor and overhead costs associated with the use of the facility for the production of the batch. It is expected that the tardiness cost for delayed delivery is

greater than the holding cost of the finished products, otherwise the jobs would be scheduled as late as possible.

For simplicity, it is assumed that the network consists of merge points only of the type shown in Figure 2. These merge points occur at each assembly station. All serial activities between the merge points are assumed to have been replaced by their convolutions, with corresponding durations and costs. This assumption is not crucial to the development of the cost model and can be relaxed if necessary.

### **3.3 A Fundamental Relationship**

The basic building block of the model is a merge point in the network that determines the start time of assembly at that station. For a line with unlimited buffer space between stations, the model is based on a fundamental relationship between the start time of the assembly operation at a station and the following three random variables

1. The finish time of the previous job at that station.
2. The finish time of the job at the previous station.
3. The delivery time of the part required at the station.

Figure 2 shows the relationship between these times for a typical operation. Since assembly can start only when all three of the above operations are completed, the start time of the assembly operation at a station is the maximum of these three random variables. Thus,

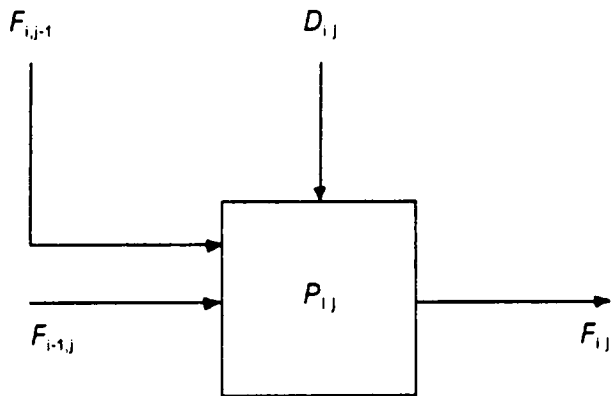


Figure 2. Relationship Between Various Times for Operation (i, j)

$$S_{i,j} = \max\{F_{i,j-1}, F_{i-1,j}, D_{i,j}\} \quad (3.1)$$

$$F_{i,j} = S_{i,j} + P_{i,j} \quad (3.2)$$

The expected start and finish times have to be determined by taking expectations of the corresponding defining equations. These derivations are discussed in further detail in the next section. These equations when recursively applied enable determination of the quantities necessary to measure the behavior of the system.

The measures of system performance required are as follows

1. Expected part waiting time for assembly j at station i

$$E(WP_{i,j}) = E(S_{i,j} - D_{i,j}) = E(\max\{F_{i,j-1}, F_{i-1,j}, D_{i,j}\}) - \mu D_{i,j} \quad (3.3)$$

2. Expected subassembly waiting time for assembly j at station i

$$E(WS_{i,j}) = E(S_{i,j} - F_{i-1,j}) = E(\max\{F_{i,j-1}, F_{i-1,j}, D_{i,j}\}) - \mu F_{i-1,j} \quad (3.4)$$

3. Expected idle time at station i for assembly j

$$E(IS_{i,j}) = E(S_{i,j} - F_{i,j-1}) = E(\max\{F_{i,j-1}, F_{i-1,j}, D_{i,j}\}) - \mu F_{i,j-1} \quad (3.5)$$

4. Expected processing time for assembly j at station i

$$E(PS_{i,j}) = \mu P_{i,j} \quad (3.6)$$

5. Expected makespan for the batch

$$E(FT) = \mu F_{Nj} - \mu F_{01} \quad (3.7)$$

6. Expected earliness of finished assembly j with respect to due date T

$$E(WF_j) = E(\max\{T, F_{Nj}\}) - \mu F_{Nj} \quad (3.8)$$

7. Expected tardiness of finished assembly batch with respect to due date T

$$E(TF) = E(\max\{T, F_{Nj}\}) - T \quad (3.9)$$

The total inventory holding and tardiness costs can therefore be determined by applying these equations for each station and each job, weighting them by the appropriate costs. In order to determine the total expected cost for a given set of values, it is therefore necessary to derive expressions for the quantities in the above defining equations for the distribution being considered. To that end, the following need to be determined.

1. The moments and distribution function of the maximum of two or more random variables.
2. The moments and distribution function of the sum of two random variables.
3. The expectation of the maximum of a random variable and a constant.

Since the equations are going to be applied recursively, in order to maintain the same form of distribution throughout the network, the following properties are also required.



1. If  $X_1$  and  $X_2$  follow a particular distribution, possibly with different parameters, then  $X = \max(X_1, X_2)$  also follows the same form of distribution with moments determined by the formulae used.
2. If  $X_1$  and  $X_2$  follow a particular distribution, possibly with different parameters, then  $X = X_1 + X_2$  also follows the same form of distribution with moments determined by the formulae used.

These properties, if satisfied, enable determination of all the other quantities of interest, given the moments. They also enable recursive application of the fundamental relationship through the network to determine the various measures of performance. On the other hand, if the properties are not satisfied, at each stage there would be mixtures of different types of possibly unknown distributions, resulting in forms difficult to manipulate or even compute. It is therefore desirable to resort to approximations and use a common distribution form for the sake of tractability.

### **3.4 Distributions Considered**

The choice of distributions to be used in developing the cost model has been made considering three main criteria. First, the distribution must be a reasonable representation of the behavior of the random processes, in this case part arrival and processing times. Second, it must retain the properties described above through the course of application of the descriptive model to the network. Finally, it must be analytically tractable for ease of use of the model.

There is no documented literature available on the distribution of purchased part delivery times. The distribution of low-repetitive task times generally exhibits positive skewness (Knott et al., 1987). It seems reasonable to expect part delivery times in small lot manufacturing to also have finite lower support and be skewed to the right, as is the case in most models in applied science (Bury, 1975). It is also expected that in small lot work, variability will be high due to unfamiliarity with the manufacturing methods. Knott and Sury (1987) reported a study of task performance time distributions in unpaced work, as is likely to be the case in this model. The performance times reported also had the same characteristics as stated above. The coefficients of variation ranged between 0.23 and 0.57 in their study.

In this research the objective is to consider randomness in processing times and part delivery times. For tractability it is desirable for probability distributions representing processing times and part arrival times to be of the same form. It appears that in small lot manufacturing, as compared to repetitive operations, processing times are likely to have high coefficients of variation and follow skewed distributions. While it may not be possible to capture all the desired characteristics in one distribution, at least different coefficients of variation and skewness must be incorporated.

The normal distribution has been used in the development of the descriptive models of Wilhelm, as well as in the prescriptive models of Sarin and Das (1987). Wilhelm (1987) clarified the nature of the issues involved in assuming normality. When the times for delivery or processing are sums of a large number of random variables, normality can be invoked by the Central Limit Theorem. Thus for batch sizes between 10 and 50, and for part delivery times which may be made up of a series of operations for the preparation of parts, normality may be a good approxi-

mation. Seidmann et al. (1987) further discussed the normality of batch production times in flexible manufacturing cells. According to the authors, batch times will be approximately normally distributed in the cases of large batches, small batches with many reworks, and small batches with normally distributed processing times. By contrast, batch times will not be normally distributed in small batches with limited rework attempts and small batches with minimal rework attempts and manufacturing times that are not normally distributed.

The normal distribution has the advantage of being capable of mathematical manipulation. However, it is symmetrical in shape and this limits its applicability if the real life behavior is skewed. In the context of the the assembly situation, the maximum operator will be repeatedly applied. The maximum of normally distributed random variables is not normal but is skewed to the right, as shown by Tippett (1925), although Clarke (1961) and Wilhelm (1982) showed that the errors are small.

The other limitation on the normal distribution is the probability of a negative tail. For coefficients of variation of 0.3 or less, this probability is negligible. However, in the small lot assembly situation, as discussed above, the actual coefficients of variation of processing times are likely to be higher, ranging between 0.23 and 0.57, as reported in Knott and Sury (1987). For this range of values, the probability of a negative tail can be significant (3.5 % at  $CV = 0.55$ ). Consequently, the normal distribution, though useful and analytically tractable, has to be used with awareness of its limitations.

Analytical studies of queueing behavior are generally based on the use of the exponential distribution. However, the applicability of the exponential distribution in a procurement-assembly situation is questionable because it implies a memoryless

process. The shape of the exponential distribution is not representative of the processes involved in this problem. Finally, it has a fixed and very high coefficient of variation of 1, and as such does not fall within the range expected.

As discussed earlier, the distribution of the maximum operator tends to be positively skewed. In the absence of exact statistical models of this distribution, one must select an input distribution that more faithfully depicts real life behavior. The lognormal distribution is an attractive one to use in this situation for a number of reasons. It has a finite lower support, and can therefore avoid the problem of having a negative infinite tail that the normal distribution has. In fact, with the use of the location parameter, any finite support can be incorporated. It can be used to fit a variety of distribution shapes along the  $(\beta_1, \beta_2)$  line (Wilhelm, 1986d). Also, it is inherently positively skewed, enabling application as a model of the maximum as well. Finally, since it is a transformation of the normal distribution, its probability integral can be expressed in terms of the associated normal probability integral. It therefore retains some of the mathematical manipulation capabilities of the normal.

The gamma distribution has been chosen for reasons similar to the lognormal distribution. It also has a finite lower support and can therefore be more realistically employed to represent variables like activity durations that are strictly positive or have some finite threshold value. The Erlang distribution is a special case of the gamma distribution with integer parameter, and represents phase type service systems. It has been used to represent service time distributions, for example in queueing theory. The gamma distribution can fit a variety of shapes ranging from the exponential to the limiting case of a normal distribution. It is also unimodal and inherently skewed to the right, making it an attractive potential representation of the

maximum operator. According to Ensore et al. (1980, 1982), the gamma distribution yields the best fit to low-repetitive task times, though the lognormal is also close.

The next sections are devoted to determination of the moments of start and finish times through the network for the three distributions, and verifying the properties required for their application.

### **3.5 Moments of the Maximum of Two Random Variables**

The maximum of two independent random variables can be characterized as follows. If  $X_1$  and  $X_2$  are the two random variables, then the cumulative probability distribution function of  $X$ , the maximum of the two, is given by

$$Pr(X \leq t) = Pr(X_1 \leq t)Pr(X_2 \leq t) \quad (3.10)$$

The density function is obtained by differentiating the above expression.

$$f_X(t) = f_{X_1}(t)Pr(X_2 \leq t) + f_{X_2}(t)Pr(X_1 \leq t) \quad (3.11)$$

This equation can be used to derive relations for the moments of the maximum of two independent random variables for the distribution being considered. If the random variables are correlated, then the bivariate distribution has to be used for the derivations.

### 3.5.1 Normal Distribution

Clarke (1961) derived the exact moments of the maximum of two normally distributed random variables, using the bivariate normal distribution. If  $\eta$  and  $v$  are normally distributed with means  $\mu_1$  and  $\mu_2$  and standard deviations  $\sigma_1$  and  $\sigma_2$  respectively, then the moments of  $\tau = \text{Max} \{ \eta , v \}$  are given by:

$$M_1 = \mu_1\phi(\alpha) + \mu_2\phi(-\alpha) + a\psi(\alpha) \quad (3.12a)$$

$$M_2 = (\mu_1^2 + \sigma_1^2)\phi(\alpha) + (\mu_2^2 + \sigma_2^2)\phi(-\alpha) + (\mu_1 + \mu_2)a\psi(\alpha) \quad (3.12b)$$

where

$$a^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho$$

$$\alpha = (\mu_1 - \mu_2)/a$$

$\psi(t)$  = standard normal density function evaluated at  $t$ .

$\phi(t)$  = cumulative standard normal distribution function evaluated at  $t$ .

$r(x, y)$  = coefficient of linear correlation between  $x$  and  $y$ .

$\rho = r(\eta, v)$

The maximum of more than two random variables can be found by applying the equations in a nested fashion. The variables can be taken two at a time to determine the moments of the maximum, then applying the equations to the resulting maximum and the next variable, and so on.

$$\beta = \max\{\eta, v, \zeta\} = \max\{\max(\eta, v), \zeta\}$$

$$r\{\max(\eta, v), \zeta\} = [\sigma_1\rho_1\phi(\alpha) + \sigma_2\rho_2\phi(-\alpha)]/(M_2 - M_1^2)^{1/2} \quad (3.13)$$

where  $\rho_1 = r(v, \zeta)$  and  $\rho_2 = r(\eta, \zeta)$ .

However, the application of these formulae is based on the assumption that the maximum of two random variables is also normally distributed. This causes errors since the maximum has a positively skewed distribution as shown by Tippett (1925). Clarke showed that the error involved due to the approximation is small, though it increases with repetitive application. The goodness of fit of the normal distribution to the maximum of two normals will be investigated in section 3.6.

Wilhelm et al. (1982) developed a seven step methodology to recursively compute the correlation coefficients between operation times required at each step in the application to the assembly system problem. The procedure assumes that jobs and stations more than two removed from the current one can be taken to be independent. The procedure is also extended to the network case with randomly distributed processing times (Wilhelm et al. 1986).

### **3.5.2 Lognormal Distribution**

The lognormal distribution is defined in terms of an associated normal. If the logarithm of a random variable is normally distributed, the variable is said to follow the lognormal distribution. If the associated normal for variable  $j$  has parameters  $\mu_{n_j}$  and  $\sigma_{n_j}$  then the  $i^{\text{th}}$  moment of the lognormal variable  $j$  is given by

$$\mu_j(i) = \exp[i\mu_{n_j} + i^2\sigma_{n_j}^2]$$

If two variables follow the two parameter lognormal distribution with parameters  $\mu_i$  and  $\sigma_j$  respectively and coefficient of correlation  $\rho$ , then the associated normals can be defined by

$$\mu_{n_j} = \ln \left[ \frac{\mu_j^2}{(\mu_j^2 + \sigma_j^2)^{1/2}} \right]$$

$$\sigma_{n_j}^2 = \ln \left[ \frac{\mu_j^2 + \sigma_j^2}{\mu_j^2} \right]$$

The correlation coefficient between the associated normals is given by

$$\rho_{12} = \frac{1}{\sigma_{n1}\sigma_{n2}} \ln \{ \rho [(\omega_1 - 1)(\omega_2 - 1)]^{1/2} + 1 \}$$

where  $\omega_k = \exp(\sigma_k^2)$ ,  $k = 1, 2$ .

For the shifted or three parameter distribution,  $\log(X - \theta)$  is normally distributed, where  $\theta$  represents the shift or location parameter. Accordingly,  $(\mu_j - \theta)$  replaces  $\mu_j$  in the above equations. The moments about zero are altered accordingly, while the central moments remain the same.

The  $i^{\text{th}}$  moments of the maximum of two two-parameter lognormals can be determined using the relations derived by Wilhelm (1986c).

$$M_i = M_{i1} + M_{i2} \quad (3.14)$$

$$M_{i1} = \mu_1(i) \phi \{ [\mu_{n1} - \mu_{n2} + i\sigma_{n1}(\sigma_{n1} - \rho_{12})] / a \} \quad (3.15a)$$

$$M_{i2} = \mu_2(i) \phi \{ [\mu_{n2} - \mu_{n1} + i\sigma_{n2}(\sigma_{n2} - \rho_{12})] / a \} \quad (3.15b)$$



$$a^2 = \sigma_{n1}^2 + \sigma_{n2}^2 - 2\sigma_{n1}\sigma_{n2}\rho_{12}$$

where  $\phi(z)$  = cumulative standard normal distribution function evaluated at  $z$ .

$M_{i2}$  is obtained by exchanging the subscripts in the formula for  $M_{i1}$ . The new correlation coefficients are derived using the formulae for the product moments detailed in Appendix A.

Extension to the three parameter case requires that the location parameters for the distributions be equal for tractability. The moments can then be determined by shifting the distribution by the same location parameter. This is not a serious limitation, since according to Johnson and Kotz (1972a), there can be considerable variation in  $\theta$  with little effect on percentiles and little effect on the cumulative distribution function. In the event the parameters are unequal, an approximation procedure is also suggested by Wilhelm (1986c). This consists of setting both location parameters equal to the smaller of the two, recalculating the other parameters based on this location parameter and then applying the equations. The location parameter for the maximum is then set equal to this computed mean minus 4 times the standard deviation to approximate the lower support. Wilhelm (1986) also experimented with different methods of setting  $\theta$ , including  $\theta = 0$  for all variables, and found little difference between the methods.

Application of these relations to the assembly problem requires the same extensions as before. The relations can be applied recursively in a nested fashion using the same relations. This of course requires that the distribution of the maximum also be lognormal. This is a more reasonable expectation than normality in view of the inherently skewed nature of the distribution, and was shown by Wilhelm (1986c) to

be so, especially when the coefficients of variation are large. The goodness of fit of the maximum of two lognormals as another lognormal is investigated in section 3.6.

### **3.5.3 Gamma Distribution**

For the gamma distribution, moments have been derived for the case of independent variables. The relations for the bivariate densities of correlated gamma random variables (Johnson and Kotz, 1972b) are not of a tractable form. The density for the two parameter gamma distribution is given by

$$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)}$$

The  $i^{\text{th}}$  moment of the gamma distribution is given by

$$\mu(i) = \alpha(\alpha + 1)\dots(\alpha + i - 1)\beta^i$$

Consequently, the mean of the distribution is  $\mu = \alpha \beta$ , the variance is  $\sigma^2 = \alpha\beta^2$ , and the coefficient of variation is  $CV = \sqrt{\alpha}$ . In the three parameter distribution,  $x$  is replaced by  $(x - \theta)$ , where  $\theta$  is the location parameter, and the moments are altered accordingly.

Given the mean and standard deviation, the  $\alpha$  and  $\beta$  values can be determined using the relations

$$\alpha = (\mu/\sigma)^2$$

$$\beta = \sigma^2/\mu$$

The moments of the maximum have to be determined using defining equation (3.11) for the density of the maximum. The gamma distribution does not have a closed finite form for its cumulative probability function. However, the Erlang distribution, which is a special case of the gamma distribution with integer shape parameter  $\alpha$ , has a closed form representation, given by

$$Pr(X \leq t) = 1 - e^{-t/\beta} \sum_{j=0}^{\alpha-1} \frac{(t/\beta)^j}{j!}$$

Substituting this expression in the density function and integrating yields, after some algebra shown in Appendix B.1, the following expressions for the  $i^{\text{th}}$  moment of the maximum of two Erlangs.

$$M_i = M_{i,1} + M_{i,2} \quad (3.16)$$

$$M_{i,1} = \beta_1^i \prod_{j=0}^{i-1} (\alpha_1 + j) - \sum_{j=0}^{\alpha_2-1} \frac{(\alpha_1 + j + i)! (\beta_1 + \beta_2)^{-\alpha_1 - j}}{(\alpha_1 - 1)! j! \beta_1^j \beta_2^{-\alpha_1}} \left( \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \right)^i \quad (3.17)$$

The corresponding term for  $M_{i,2}$  is determined by exchanging the subscripts in the formula for  $M_{i,1}$ .

This relation gives exact values for the moments of the maximum of two gamma variables, provided the  $\alpha$  values are integer. However, in the course of application of the maximum operator to an assembly network, one cannot be certain of maintaining that restriction. Even if the individual operation times have integer values of  $\alpha$ , their

sum and maximum need not. Therefore relations are also required for the general gamma case. This requires the use of the following infinite series representation for the cumulative distribution function ( Johnson and Kotz, 1972a).

$$Pr(X \leq t) = e^{-t/\beta} \sum_{j=0}^{\infty} \frac{(t/\beta)^{\alpha+j}}{\Gamma(\alpha+j+1)}$$

Substituting this in the expression for the density and evaluating the integral yields, after some lengthy manipulations shown in Appendix B.2, the following relations for the  $i^{\text{th}}$  moment of the maximum

$$M_i = M_{i_1} + M_{i_2}$$

$$M_{i_1} = \frac{\Gamma(\alpha_1 + \alpha_2 + i)}{\Gamma(\alpha_1)\alpha_2\Gamma(\alpha_2)} \frac{\beta_1^{\alpha_2+i}}{(\beta_1 + \beta_2)^{\alpha_2}} F\left(\alpha_2, -\alpha_1 - i - 1, \alpha_2 + 1, \frac{\beta_1}{\beta_1 + \beta_2}\right) \quad (3.18)$$

where  $F(a, b, c, z)$  is the generalized hypergeometric series given by

$$F(a, b, c, z) = 1 + \frac{ab}{c} z + \frac{a(a+1)b(b+1)}{c(c+1)} \frac{z^2}{2!} \dots\dots$$

and the corresponding term for  $M_{i_2}$  is determined by exchanging the subscripts in the above formula.

For integer values of  $\alpha$  the hypergeometric series in the above formula reduces to a finite terminating series and can be calculated exactly. For non-integer values, the series converges and can be evaluated by summation to the accuracy desired, since the argument  $z$  falls within the circle of convergence  $|z| < 1$  (Slater, 1966).

Computational experience showed the above formula to be ill-conditioned for large values of  $\alpha$  (above 25), even in double precision. The moments of the maximum for such values are determined by the Erlang relations, using interpolation for fractional values. This approximation was found to give accurate results when verified by numerical integration. It must be noted, however, that  $\alpha$  values of greater than 25 correspond to coefficients of variation of below 0.2, which falls outside the range of 0.22 - 0.57 reported by Knott and Sury (1987) and being considered here.

The above derivations can be extended to the three parameter case by shifting the distribution by  $\theta$  and calculating the moments accordingly. As in the case of the lognormal distribution, it is necessary for tractability to have both the location parameters equal. The same approximation procedure as used for the lognormal has to be applied in case the  $\theta$  values are unequal.

### **3.6 Test for the Distribution of the Maximum**

As discussed earlier, the suitability of a distribution depends on the ability to represent the maximum of two random variables, following a particular distribution, by the same distribution with parameters calculated from the moments. In order to test this for each of the three distributions considered, tests were conducted for a representative range of values of parameters. For the sake of uniform comparison, the parameters were chosen such that the first two moments were the same for all three distributions. These were chosen to correspond to means ranging from 2 to 14 units (say weeks), and coefficients of variation ranging from 0.25 to 0.55.

In order to test the goodness of the fit of a distribution of the same form, the moments of the maximum were computed using the relations derived. The error measure used is defined as

$$D_{\max} = \max\{abs[AF(t_i) - EF(t_i)], i = 1, \dots, n\} \quad (3.19)$$

where

$AF(.)$  is the approximating cumulative distribution of the same family, obtained using the exact moments computed by the formulae in the previous section,

$EF(.)$  is the exact cumulative distribution function determined from the constituent distributions by equation (3.10), and

$n$  is the number of intervals into which the range is divided.

The relationship between these terms is shown in Figure 3.

The statistic  $D_{\max}$  is the same as used to apply the Kolmogorov-Smirnov test of goodness of fit of data to a postulated distribution. It measures the greatest absolute difference between the actual distribution function  $Pr(X \leq t)$  and the distribution being fitted at each interval. The maximum difference is compared with the critical values, determined for the desired significance level using the number of intervals, to get an estimate of the accuracy of the approximation (Kraft and Eeden, 1968).

Sculli and Wong (1985) used this statistic to measure the goodness of fit of the Beta distribution as an approximation for the maximum and sum of two Beta random variables in the context of PERT network analysis. Since closed form relations are not available for the Beta distribution, the moments were evaluated by numerical integration. The values of the exact cumulative distribution were treated as samples from an unknown distribution, and the  $D_{\max}$  statistic derived as above was used to apply the

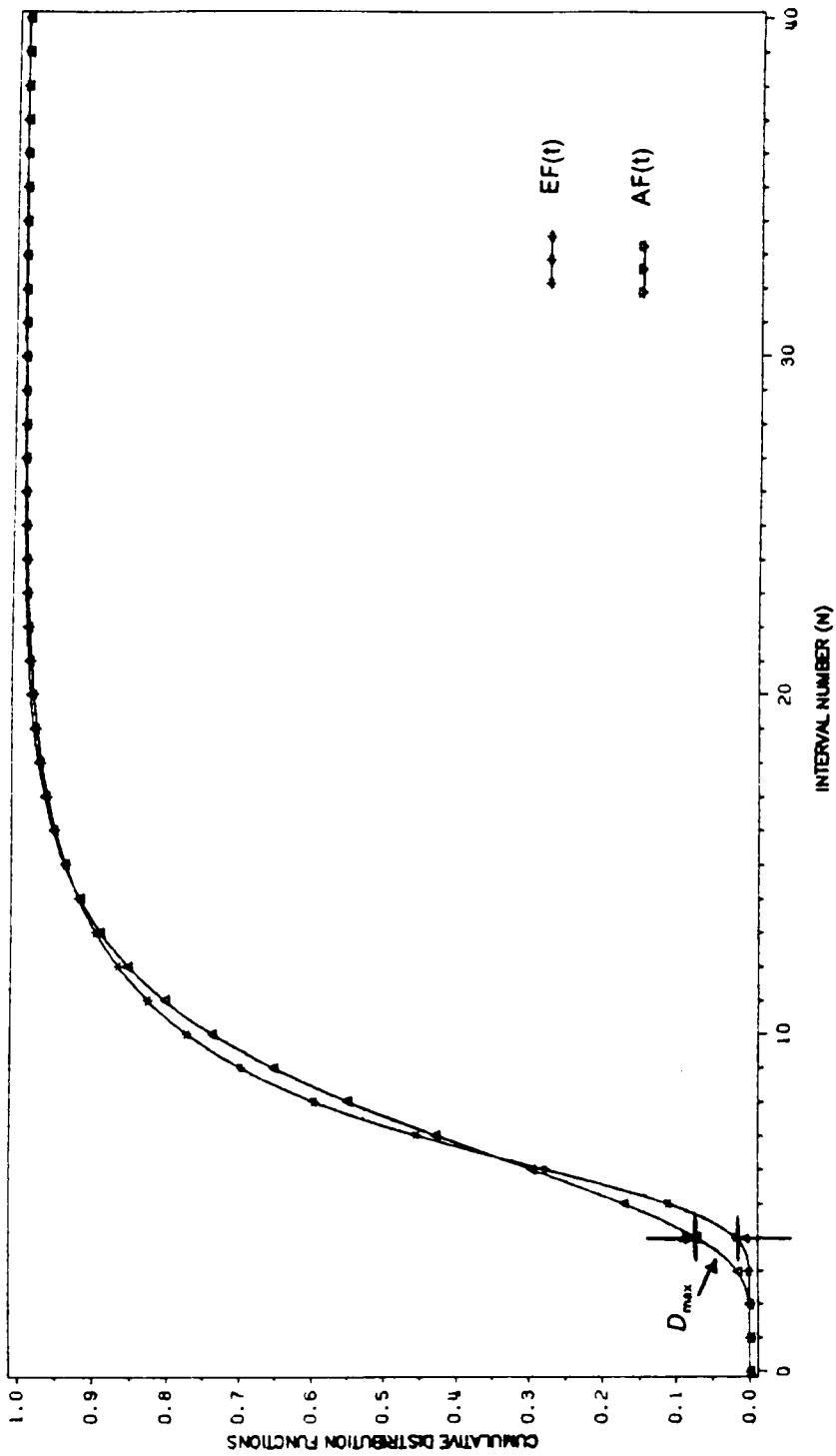


Figure 3. Determination of Error Between Approximating and Exact Distributions

Kolmogorov-Smirnov test. Strictly speaking, the values of  $D_{\max}$  are not determined from random samples but are actual values derived analytically. However, the statistic does provide a useful indication of the closeness of the approximation on a standardized scale of 0 to 1.

Tests have been conducted using  $n = 40$  for the range of parameter values mentioned above. The mean absolute deviation value ( MAD ) was also computed to estimate the average error in the approximation. Table 1 shows the cases resulting in the largest deviations ( greater than 0.05 ). It also shows a frequency distribution of the error over the 128 combinations considered.

It can be seen that both the lognormal and the gamma distributions result in better fits of the maximum than the normal distribution. The lognormal distribution performs marginally better than the gamma distribution in approximating the maximum. For the normal distribution the largest value of  $D_{\max}$  over the range considered is 0.08163, with a MAD of 0.0094. The corresponding values for the lognormal distribution are 0.06069 and 0.01005 respectively. The gamma distribution results in values of 0.06142 and 0.01162 respectively. The error frequency also shows the same trend. For the sake of comparison, the critical value for random samples corresponding to this value of  $n$  for a 90% significance level is 0.213, and  $D_{\max}$  values below this imply that the approximation cannot be rejected at that level.

It can be seen that the largest deviations occur for all distributions for the same combinations of parameter values. Specifically, they occur when the following conditions occur simultaneously: (i) the variances differ widely, by factors of 5 to 6 or more, and (ii) the means are of similar magnitude. This situation occurs when one of the random variables is relatively invariant compared to the other. Consequently, the re-



**Table 1. Comparison of Deviations of Actual and Approximating Distributions for the Maximum**

Error Frequency Distribution							
Error Range			Normal	Lognormal	Gamma		
$D_{max} \geq 0.05$			12	6	8		
$0.05 > D_{max} \geq 0.04$			10	0	3		
$0.04 > D_{max} \geq 0.03$			18	8	8		
$0.03 > D_{max} \geq 0.02$			20	11	14		
$0.02 > D_{max} \geq 0.01$			33	34	28		
$D_{max} < 0.005$			35	69	69		
Total No. of Combinations			128	128	128		
Combinations with $D_{max} \geq 0.05$							
	Variable 1		Variable 2		Maximum Deviation ( $D_{max}$ )		
No.	Mean	Var.	Mean	Var.	Normal	Lognormal	Gamma
1	10	6.25	14	59.29	0.08163	0.06069	0.06142
2	6	2.25	10	30.25	0.07159	0.05320	0.06033
3	10	6.25	14	39.69	0.06463	0.03990	0.04736
4	6	4.41	10	30.25	0.06142	0.03647	0.04130
5	6	2.25	10	20.25	0.05986	0.03467	0.03841
6	10	12.25	14	59.29	0.05931	0.03628	0.03939
7	2	0.25	2	1.21	0.05892	0.05699	0.05520
8	10	6.25	10	30.25	0.05892	0.05699	0.05520
9	6	2.25	6	10.89	0.05892	0.05699	0.05520
10	14	12.25	14	59.29	0.05892	0.05699	0.05520
11	6	2.25	14	59.29	0.05527	0.02308	0.04089
12	6	4.41	14	59.29	0.05168	0.01994	0.03213

sulting distribution of the maximum is of a truncated shape and the quality of fit deteriorates if the means are of similar magnitude. The impact of differences in variances reduces as the differences between the means increases, that is, when one of the variables starts to dominate and effectively determines the maximum.

### **3.7 Moments and Distribution of the Sum of Two Random**

#### **Variables**

In this section, we discuss the derivation of the moments of the sum of two random variables, and study the resulting distribution to verify if it can be approximated by the same form as the constituent distributions.

Let  $X_1$  and  $X_2$  be two random variables with means  $\mu_1$  and  $\mu_2$  respectively, standard deviations  $\sigma_1$  and  $\sigma_2$  respectively, and correlation coefficient  $\rho$ . Then, according to statistical theory, the mean and standard deviation of  $X = X_1 + X_2$  are given by

$$\mu = \mu_1 + \mu_2$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho$$

In the context of the assembly system problem, we are assuming that the processing times are independent of the station and part ready times. Consequently, we need to be concerned only with sums of independent random variables.

The correlation coefficient of the operation finish time,  $F_{ij} = S_{ij} + P_{ij}$ , with any random variable of interest, say  $Y$ , is related to the correlation coefficient of the start time,  $S_{ij}$ , with  $Y$  by the following relationship.

$$r(F_{ij}, Y) = r(S_{ij}, Y) \frac{\sigma_{S_{ij}}}{\sigma_{F_{ij}}} = r(S_{ij}, Y) \frac{\sigma_{S_{ij}}}{\sqrt{\sigma_{S_{ij}}^2 + \sigma_{P_{ij}}^2}} \quad (3.20)$$

These relations can thus be used to determine the mean, variance, and correlations of the variable arising out of the summation of the operation start and processing times.

The distribution of the sum, however, depends on the nature of the constituent distributions. If  $X_1$  and  $X_2$  are normally distributed, then the distribution of their sum is known to be normal. No errors are therefore involved in this case.

If the two random variables are independent, follow the gamma distribution and share a common scale parameter  $\beta$ , then the sum of the two also follows the gamma distribution with the same scale parameter  $\beta$ , and  $\alpha = \alpha_1 + \alpha_2$ . This reproductive property is restricted to when the  $\beta$  values are equal. However, there is no guarantee that the scale parameters for the different random variables in the assembly problem will be the same. Even if the individual variables have the same scale parameters, their maximum will not. The reproductive property therefore needs to be verified for the case of unequal scale parameters.

Likewise, the distribution of the sum of two lognormals does not follow a known distribution. The reproductive property therefore needs to be verified for this distribution as well.

The strategy followed to verify the property for the lognormal and gamma distributions is the same as used for the maximum. The moments of the sum are computed using the standard relations above. These parameters are used to construct an approximating distribution of the same form as the constituents. The actual distribution of the sum of two independent random variables can be determined as follows.

$$Pr(X_1 + X_2 \leq t) = \int_{L_1}^t \int_{L_2}^{t-X_1} f_1(x_1)f_2(x_2)dX_2dX_1 \quad (3.21)$$

Evaluation of the inner integral results in the expression

$$Pr(X_1 + X_2 \leq t) = \int_{L_1}^t f_1(x_1)Pr(X_2 \leq t - X_1)dX_1 \quad (3.22)$$

This integral cannot be evaluated analytically for either of the distributions. However, it can be evaluated numerically using numerical integration techniques for a given value of  $t$ .

The numerically determined values can now be compared with the approximating distribution obtained by modelling the sum as another distribution of the same form. The tests have been conducted using the same ranges of means and coefficients of variation as used in the study of the maximum. As before, the Kolmogorov-Smirnov statistic of maximum deviation between the actual and approximating distributions is used as a measure of the error involved in the approximation. The range is divided into 40 intervals at which the values are compared.

Analysis of the results of the comparison shows that the errors involved in taking the sum to be of the same form as the constituent distributions are much lower than for the maximum. The cases resulting in the highest errors (greater than 0.02) and the error frequency distribution are shown in Table 2. The gamma distribution generally results in lower errors than the lognormal distribution.

For the lognormal distribution the largest value of  $D_{\max}$  over the range of values considered is 0.02624, with a mean absolute deviation of 0.00671. For the gamma distribution, the largest value of  $D_{\max}$  is 0.01861, with a mean absolute deviation of 0.00512. This shows that the sum of two random variables can be reasonably accurately modelled as belonging to the same family of distributions.

It is interesting to note that the largest deviations in the sum as well as the maximum occur for both distributions at similar combinations of parameter values. The deviations are most pronounced when the variances of the constituent distributions differ considerably and the means are similar. As the difference between the means increases, one of the variables dominates and determines the nature of the distribution, and the errors are much lower. In the assembly system problem, it is expected that the part or subassembly ready times will be larger than the processing times (Wilhelm, 1986d), so the assumption can be reasonably invoked.

**Table 2. Comparison of Deviations of Actual and Approximating Distributions for the Sum**

Error Frequency Distribution						
Error Range			Lognormal		Gamma	
$D_{max} \geq 0.02$			8		0	
$0.02 > D_{max} \geq 0.015$			11		8	
$0.015 > D_{max} \geq 0.01$			19		12	
$0.01 > D_{max} \geq 0.005$			28		36	
$D_{max} < 0.005$			62		72	
Total No. of Combinations			128		128	
Combinations with $D_{max} \geq 0.02$						
	Variable 1		Variable 2		Maximum Deviation ( $D_{max}$ )	
No.	Mean	Var.	Mean	Var.	Lognormal	Gamma
1	10	30.25	6	2.25	0.02624	0.01861
2	6	2.25	10	59.29	0.02623	0.01752
3	10	6.25	14	59.29	0.02499	0.01848
4	2	0.25	6	10.89	0.02466	0.01664
5	2	0.25	2	1.21	0.02048	0.01644
6	6	2.25	6	10.89	0.02048	0.01644
7	10	6.25	10	30.25	0.02048	0.01644
8	14	12.25	14	59.29	0.02048	0.01644

### 3.8 Expectation of the Maximum of a Random Variable and a Constant

The last quantity that needs to be determined for computing the expected total cost function is the expectation of the maximum of the finish time and the due date. This is required for computing the expected waiting and tardiness costs for the finished assembly. The expressions for this quantity are derived below for the three distributions being considered, using the following relation.

$$E(\max\{T, X\}) = TPr(X \leq T) + \int_T^U xf(x)dx \quad (3.23)$$

The integral term on the right hand side is the partial expectation of a random variable, and may be determined for the different distributions along the lines of Winkler et al. (1972), with  $U = \infty$ .

For the normal distribution,

$$E(\max\{T, X\}) = T\phi\left(\frac{T-\mu}{\sigma}\right) + \mu\left[1 - \phi\left(\frac{T-\mu}{\sigma}\right)\right] + \sigma\psi\left(\frac{T-\mu}{\sigma}\right) \quad (3.24)$$

For the lognormal distribution,

$$E(\max\{T, X\}) = T\phi\left(\frac{\ln(T) - \mu_n}{\sigma_n}\right) + \mu_n\phi\left(\frac{\mu_n - \ln(T) + \sigma_n^2}{\sigma_n}\right) \quad (3.25)$$

where the subscript n refers to parameters of the associated normal.

For the gamma distribution,

$$E(\max\{T, X\}) = T \frac{\Gamma(\alpha, T/\beta)}{\Gamma(\alpha)} + \alpha\beta \left( 1 - \frac{\Gamma(\alpha + 1, T/\beta)}{\Gamma(\alpha + 1)} \right) \quad (3.26)$$

where  $\Gamma(\alpha, T/\beta)$  is the incomplete gamma integral evaluated at  $T/\beta$ , and  $\Gamma(\alpha)$  is the Gamma function.

For the normal and lognormal distributions, the expected value of the maximum of a random variable and a constant can equivalently be derived, using equations (3.12) and (3.14-3.15) respectively, by treating the constant as a degenerate random variable with mean =  $T$  and standard deviation = 0. This method also yields the same relations as above. Therefore, the same modelling method as that used at any assembly station can be applied by treating the due date as a dummy station with zero variance and processing time.

For the gamma distribution, however, the same approach cannot work since setting the standard deviation equal to zero requires making the value of  $\alpha = \infty$ , which cannot be handled in the equations developed. Accordingly, the due date setting station has to be handled using the formula above.

### **3.9 Summary**

In this chapter, the basic descriptive model used to represent the behavior of the assembly system has been developed. The performance measures based on which the system will be optimized have been identified, and relationships necessary for



their evaluation have been developed. Analysis of the errors involved in approximating the distributions resulting from the maximum and summation operators to be of the same form as the constituent distributions has been done. The prescriptive part of the model, based on the descriptive model developed in this chapter, is discussed in Chapters 4 and 5.

## **4.0 Solution of Single Job or Lot-for-lot Problem**

### **4.1 Introduction**

This chapter describes the development of the objective function and solution techniques for the single job or lot-for-lot problem. This problem is concerned with determining optimal part delivery dates when only one job flows through the assembly stages, or when the entire lot is produced as one job without distinguishing between completions of individual jobs. It is therefore not necessary to consider the availability of assembly stations from previous jobs. Since the part available times are assumed to be independent of each other as well as of all finish times, there is no need to compute any correlations in deriving the various measures of system performance. The objective function can thus be computed in a straightforward way by recursive application of the fundamental relationships from Chapter 3.

The single job problem solution is useful to the overall problem solution in three ways. First, in case the entire batch is produced as one lot, the solution directly yields the optimal delivery dates for the parts and planned completion times of the subassemblies. Second, the solution can be used to generate insights into system behavior and thus facilitate a better understanding of the multiple job problem. Finally, as will be seen in the next chapter, it can be imbedded in heuristic methods to solve the

multiple job problem, in view of the size of the problem and the difficulty of a direct solution method.

## **4.2 Single Station Optimization**

A typical assembly station in the single job problem can be represented as shown in Figure 4a. The inputs to the station are the subassembly from the previous station and the part required at the station. The start and finish times of the assembly operation at station  $i$  are given by the fundamental relationships in equations (3.1) and (3.2), except that the term representing station availability need not be considered. The subscript  $j$  identifying the job can be omitted in the single job case.

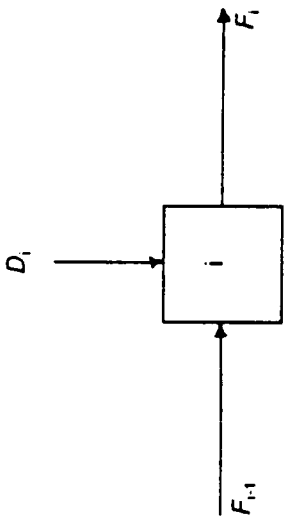
The waiting time for the subassembly at station  $i$  is given by

$$WS_i = S_i - F_{i-1}$$

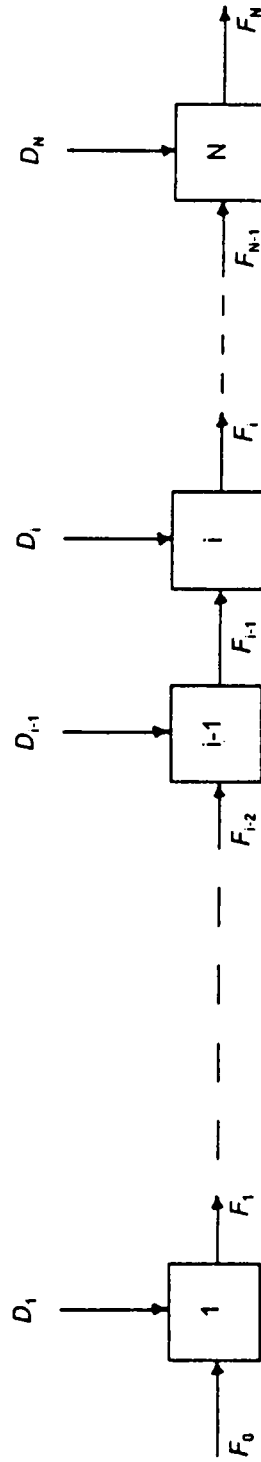
where  $S_i = \max\{F_{i-1}, D_i\}$ . If the subassembly arrives before the part, this term represents the time that it waits for the part. If it arrives after the part and thus determines the assembly start time, this term is zero. The corresponding waiting time for the part is given by

$$WP_i = S_i - D_i$$

The expected total waiting cost at station  $i$  can be determined by taking expectations of the above waiting times, and multiplying them by the respective costs.



a) Single Station Assembly



b) Single Job Multi-Station Assembly

Figure 4. Single Job Assembly System

$$\begin{aligned}
E(TC_i) &= CS_i E(WS_i) + CE_i E(WP_i) \\
&= CS_i [E(\max\{F_{i-1}, D_i\}) - \mu F_{i-1}] + CE_i [E(\max\{F_{i-1}, D_i\}) - \mu D_i] \quad (4.1)
\end{aligned}$$

Substitution of the formulae for the moments of the maximum function, for the distribution being considered, yields direct relations for the expected waiting cost at the station.

#### **4.2.1 Single Station Solution for Normal Distribution**

The total expected cost for the single station problem for the normal distribution is obtained by substitution of the moments of the maximum of two normals from equation (3.12). The expected assembly start time is therefore given by

$$\mu S_i = E(\max\{F_{i-1}, D_i\}) = \mu F_{i-1} \phi(\alpha_i) + \mu D_i \phi(-\alpha_i) + a \psi(\alpha_i)$$

where  $a = \sqrt{(\sigma F_{i-1}^2 + \sigma D_i^2)}$ ,  $\alpha_i = \frac{\mu F_{i-1} - \mu D_i}{a}$ , and  $\psi$  and  $\phi$  are the normal density and the cumulative probability distribution function respectively.

Substituting this formula in equation (4.1) results in the following relation after simplification.

$$E(TC_i) = (CS_i + CE_i) a \psi(\alpha_i) + (CS_i + CE_i) \phi(\alpha_i) (\mu F_{i-1} - \mu D_i) - CS_i (\mu F_{i-1} - \mu D_i) \quad (4.2)$$

Sarin and Das (1987) derived relations for the optimal delivery date ( $\mu D_i^*$ ) and the optimal cost ( $TC_i^*$ ) by setting the first derivative of the expected station cost equal to zero.

$$\alpha_i^* = \phi^{-1}\left(\frac{CS_i}{CS_i + CE_i}\right) \quad (4.3)$$

$$\mu D_i^* = \mu F_{i-1} - \alpha_i^* \sqrt{(\sigma F_{i-1}^2 + \sigma D_i^2)} \quad (4.4)$$

$$E(TC_i^*) = [(CS_i + CE_i)\{\alpha_i^* \phi(\alpha_i^*) + \psi(\alpha_i^*)\} - CS_i \alpha_i^*] \sqrt{(\sigma F_{i-1}^2 + \sigma D_i^2)} \quad (4.5a)$$

Substituting the value of  $\phi(\alpha_i^*) = \frac{CS_i}{CS_i + CE_i}$  this expression can be further simplified to

$$E(TC_i^*) = (CS_i + CE_i) \psi(\alpha_i^*) \sqrt{(\sigma F_{i-1}^2 + \sigma D_i^2)} \quad (4.5)$$

Evaluation of the second derivative of the expected station cost shows that the function is convex and therefore the solution obtained above is the global minimum. The sensitivity of the solution to the problem parameters can be understood based on these relations, and is as follows.

1. If the part and subassembly holding costs are equal ( $CE_i = CS_i$ ), the optimal solution is always at  $\mu D_i^* = \mu F_{i-1}$ , regardless of the values of the other parameters. However, the optimal cost increases as the variances increase.
2. If the inventory cost of the subassembly is greater than that of the part ( $CS_i > CE_i$ ), then  $\alpha_i^* > 0$ , so that  $\mu D_i^* < \mu F_{i-1}$ . This means that the part is expedited to reduce the probability of the more expensive subassembly having to wait for the part. The part delivery time decreases as the cost of the subassembly increases. Corresponding relationships hold if  $CS_i < CE_i$ .
3. If  $CS_i > CE_i$ , the amount by which the part is expedited changes as a function of the variances. Increasing the variances increases the safety time provided for the delivery of the part. Similar relationships hold if  $CS_i < CE_i$ .

## **4.2.2 Single Station Solution for Non-Normal Distributions**

The total expected cost for the single station problem for the lognormal distribution is obtained along the same lines as for the normal distribution. Substitution of the moments of the maximum of two lognormals from equations (3.14-3.15) in equation (4.1) results in the following relation after simplification.

$$E(TC_i) = (CS_i + CE_i)\mu F_{i-1}\phi(z_1) + (CS_i + CE_i)\mu D_i\phi(z_2) - CE_i\mu D_i - CS_i\mu F_{i-1} \quad (4.6)$$

where  $z_1 = \frac{\mu_{nF} - \mu_{nD} + \sigma_{nF}^2}{\sqrt{(\sigma_{nF}^2 + \sigma_{nD}^2)}}$ , and  $z_2 = \frac{\mu_{nD} - \mu_{nF} + \sigma_{nD}^2}{\sqrt{(\sigma_{nF}^2 + \sigma_{nD}^2)}}$ , and the subscript n refers to the parameters of the associated normals.

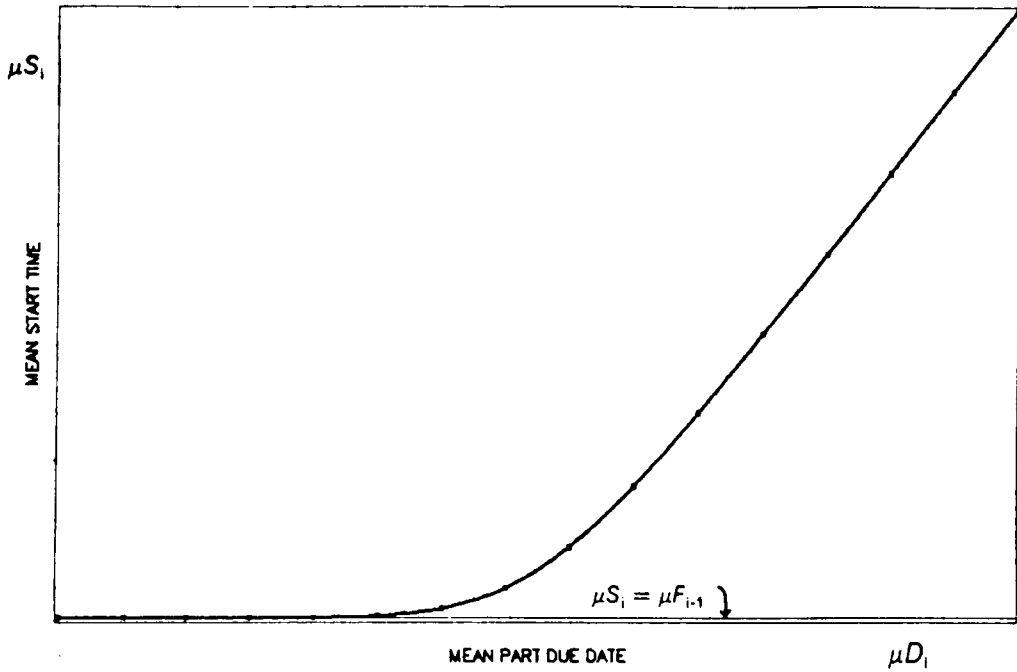
Differentiating and setting the derivative of the expected cost equal to zero results in an extremely complex equation that is not possible to solve in closed form. The gamma distribution results in even more complex forms, due to the series summations in the formulae in (3.16-3.18) for the moments of the maximum. Consequently other means have to be explored for non-normal distributions even for single stage optimization. However, it can be shown that the objective function remains convex. To that end, differentiating the general expression in equation (4.1) yields

$$\frac{dE(TC_i)}{d\mu D_i} = (CE_i + CS_i) \frac{d\mu S_i}{d\mu D_i} - CE_i = 0 \quad (4.7)$$

$$\frac{d^2E(TC_i)}{d^2\mu D_i} = (CE_i + CS_i) \frac{d^2\mu S_i}{d^2\mu D_i} \quad (4.8)$$

where  $S_i = \max\{F_{i-1}, D_i\}$

a)  $\mu S_i$  vs.  $\mu D_i$



b)  $\frac{d\mu S_i}{d\mu D_i}$  vs.  $\mu D_i$

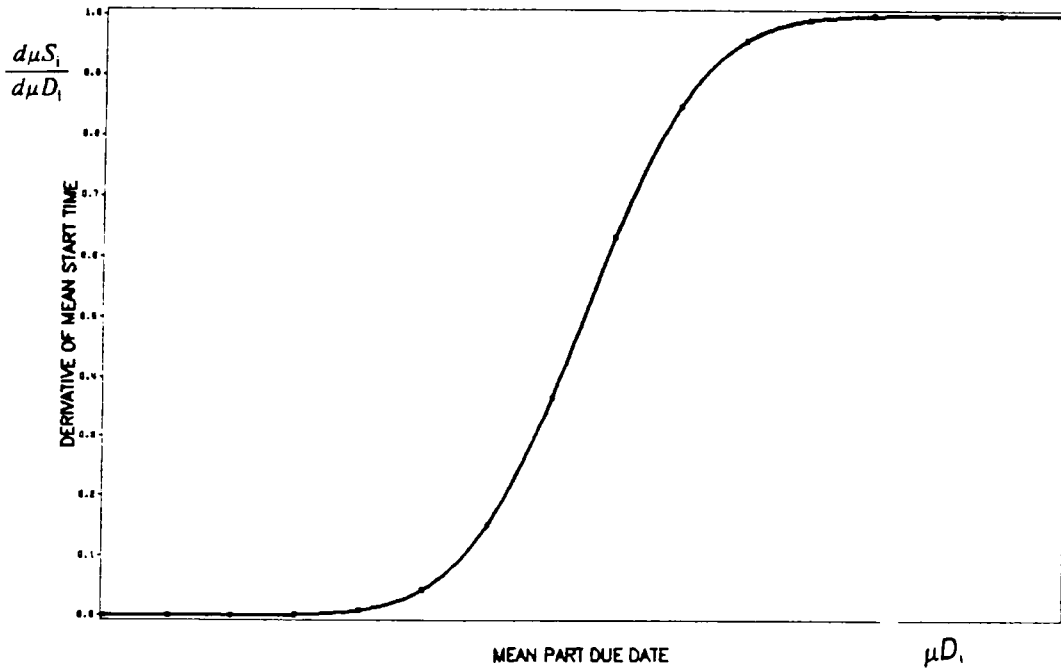


Figure 5. Behavior of Mean Start Time vs. Mean Part Due Date



To illustrate the behavior of the function, the shape of  $\mu S_i$  vs.  $\mu D_i$  is graphed in Figure 5a for a general distribution. It takes on a value close to  $\mu F_{i-1}$  when  $\mu D_i$  is very small and increases monotonically, tending to  $\mu D_i$  as  $\mu D_i$  becomes much larger than  $\mu F_{i-1}$ . The derivative of  $\mu S_i$  starts from a limiting value of 0 when  $\mu F_{i-1}$  dominates, and increases monotonically until it reaches a value of 1 in the limit. The second derivative of  $\mu S_i$  is therefore always positive, attaining a value of 0 only in the limit. This behavior is illustrated in Figure 5b.

Since the derivative is monotonic and takes on values between 0 and 1, and the costs are positive, equation (4.7) has a unique solution at  $\frac{d\mu S_i}{d\mu D_i} = \frac{CE_i}{CE_i + CS_i}$ . For positive costs, the value of the right hand side of equation (4.8) is always positive, and the function is convex for general distributions as well. Therefore, the single variable unconstrained optimization methods can be applied to find the optimal solution for a given set of parameters. This has been done for the lognormal and gamma distributions using the DSC-Powell method of unconstrained single variable search. For the normal distribution, the optimal solution is obtained using the closed form expressions in equations (4.3)-(4.5).

Table 3 shows the results of single station optimization for a set of sample problems for each of the three distributions. For each problem number, the parameters are assumed to be for the distributions under consideration. It can be seen that the greatest differences between the optimal decisions for the different distributions occur when the variances are large (problems 4,8,10,12), since that is when the differences between the distributions are most marked. The costs associated with the optimal policy vary by as much as 10% even in the other cases. The effect of distributions on decisions and costs will be discussed further when the solutions to the

**Table 3. Comparison of Single Station Optimal Values for Different Distributions**

Initial Problem:  $\mu F_0 = 10, \sigma F_0 = 2, \sigma D_1 = 2, CS_1 = 1, CE_1 = 1$

Pr. No.	Variation from Initial Problem	Normal		Lognormal		Gamma	
		$\mu D^*$	$TC^*$	$\mu D^*$	$TC^*$	$\mu D^*$	$TC^*$
1	--	10.0	2.257	9.99	2.227	9.96	2.245
2	CS = 5	7.26	4.239	7.40	4.275	7.35	4.206
3	CS = 10	6.22	5.089	6.38	5.282	6.47	5.076
4	$\sigma D = 4$	10.0	3.568	10.38	3.415	10.41	3.512
5	$\sigma D = 0.5$	10.0	1.645	9.82	1.62	9.82	1.62
6	$\sigma F = 4, \sigma D = 4$	10.0	4.513	9.88	4.293	9.94	4.424
7	$\sigma F = 0.5, \sigma D = 0.5$	10.0	0.564	10.0	0.564	10.0	0.564
8	$\sigma D = 4, CS = 5$	5.67	6.703	6.42	7.424	6.30	7.014
9	$\sigma D = 0.5, CS = 5$	8.01	3.090	8.03	2.800	8.03	2.800
10	$\sigma F = 4, CS = 5$	5.67	6.703	6.00	5.861	5.83	6.119
11	$\sigma F = 0.5, CS = 5$	8.01	3.090	8.13	3.372	8.13	3.372
12	$\sigma F = 4, \sigma D = 4, CS = 5$	4.52	8.479	5.38	8.451	5.22	8.697

multi-station problem are discussed in Section 4.7. The behavior of the solutions with changes in problem parameters follows essentially similar patterns to the normal distribution.

### 4.3 Multi-station Assembly Problem

The single job multi-station assembly system consists of a series of individual stations of the type described in Section 4.2 and represented in Figure 4b. The total expected cost function for the single job problem can be determined by applying defining equations (3.1) and (3.2) recursively and, summing up the part and subassembly waiting costs as per equations (3.3) and (3.4) over all stations.

$$E(TC) = \sum_{i=1}^N [CS_i[E(\max\{F_{i-1}, D_i\}) - \mu F_{i-1}] + CE_i[E(\max\{F_{i-1}, D_i\}) - \mu D_i]] \quad (4.9)$$

The cost pertaining to the makespan of the job is implicitly considered in the subassembly waiting costs. Equivalently, should this cost need to be considered separately, it is simply added on to the subassembly cost at each station.

The determination of optimal finished assembly delivery dates can be modelled within the same framework as above by creating a dummy  $(N + 1)^{th}$  station at the end of the line, with  $\sigma D_{N+1} = 0$ ,  $\mu P_{N+1} = 0$ ,  $\sigma P_{N+1} = 0$  and  $CE_{N+1} =$  the tardiness cost per unit time. The cost calculation for the due date setting station has to be done using the expressions derived in Section 3.8 for the maximum of a random variable and a con-

stant. The optimal delivery date for the dummy part is then the optimal assembly delivery date. If the delivery date is customer specified, then all other decisions are translated by the amount required to make the optimal delivery date equal to the specified date.

### **4.3.1 Initial Solution Approaches**

The single station solution derived in Section 4.2 considers each station in isolation. It does not consider the effect the decisions taken at that station have on the costs at later stations. In order to be optimal, the solution technique must consider the interrelations between the stations, specifically the effect of  $\mu D_i$  on the expected cost function at later stations.

#### **Dynamic Programming**

Sarin and Das (1987) developed a Dynamic Programming approach to solve this problem with normally distributed delivery times and deterministic processing times. In this approach, each assembly station is modelled as a DP decision stage. The individual return function for each stage is the expected cost for that stage alone, and the overall return function is the cumulative sum of all individual return functions up to that stage. In order to capture the interrelation between the stages, the derivative of the return at the previous stage has to be included when deriving the optimal policy for the current stage.

The ability to determine the optimal single stage return in closed form for the normal distribution enables determination of the derivative of return transferred from

the previous stage in closed form as well. However, this proves to be of an extremely cumbersome form, and therefore a regression model is utilized to approximate the derivative and retain a form that is similar to the individual stage derivative, in order to enable further derivation in closed form. Ranges of parameter values are considered in developing the regression model, and the results are shown to be effective compared to previously reported methods of scheduling such as the safety buffer rule.

Since only deterministic processing times are considered in the DP approach,  $\sigma F_i^2 = \sigma S_i^2$ . If processing times are random, then  $\sigma F_i^2 = \sigma S_i^2 + \sigma P_i^2$ . Therefore, in order to incorporate random processing times in the DP approach, the equation defining the return transferred from the previous stage has to be altered to include the effect of this change in variance. Consequently, a new regression model must be developed to approximate its derivative. This will cause further errors in the approximation and affect the quality of the solution.

For non-normal distributions, even the single stage return cannot be optimized in closed form, so there is no way to analytically represent the optimal policy and the derivative of the transferred return. Attempts to obtain approximations or transformations that could be used in conjunction with the DP methodology were not successful. Therefore, other means of solving the problem need to be explored.

### **Nonlinear Programming**

The problem on hand is that of minimizing a nonlinear objective function without any constraints, since the restrictions due to precedence are already built into the objective function. It was therefore decided to explore unconstrained nonlinear opti-

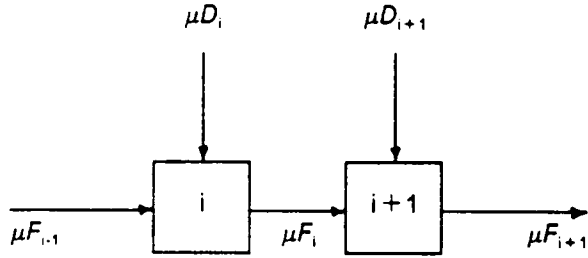
mization techniques as a solution approach, with the hope that the insights resulting from the solution may aid in developing specially tailored solution methods.

Prior to application of the optimization technique, attempts were made to understand the behavior of the objective function, specifically with respect to convexity and unimodality, since the effectiveness of the technique depends on these factors. The objective function for the two station problem, which follows the structure shown in Figure 6a, is given by  $E(TC) = E(TC_i) + E(TC_{i+1})$ . Here  $E(TC_i)$  and  $E(TC_{i+1})$  represent the contributions from stations  $i$  and  $i + 1$  respectively, and the optimal values of  $\mu D_i$  and  $\mu D_{i+1}$  need to be determined. Since  $\mu D_{i+1}$  appears only in the second term, its optimal value is the same as the single station solution, given the input  $\mu F_i$  and  $\sigma F_i^2$  arising out of the decision at station  $i$ . The corresponding optimal objective function value for station  $i + 1$  is obtained from equation (4.5), and is

$$E(TC_{i+1}^*) = (CS_{i+1} + CE_{i+1})\psi(\alpha_{i+1}^*)\sqrt{(\sigma F_i^2 + \sigma D_{i+1}^2)}$$

This expression can replace the second term in the two station objective function  $E(TC) = E(TC_i) + E(TC_{i+1})$ , since the optimal decision at station  $i + 1$  will be translated according to the decision at station  $i$ . Since the costs are positive, this term is a monotonically increasing function of  $\sigma F_i^2$ , the finish time variance of the previous stage. Both terms in the objective function are functions of  $\mu D_i$ , where the first term is simply the single station objective function for station  $i$ , and the dependency of the second term is through the variance of finish time  $\sigma F_i^2$ . The first term as a function of  $\mu D_i$  is known to be convex (Sarin and Das, 1987), but the behavior of the second term needs to be examined. The Hessian matrix even for the normal distribution is too complex to draw any conclusions analytically, and that for non-normal distrib-

a) Single Job Two Station Problem



b)  $\sigma S_i^2$  vs.  $\mu D_i$

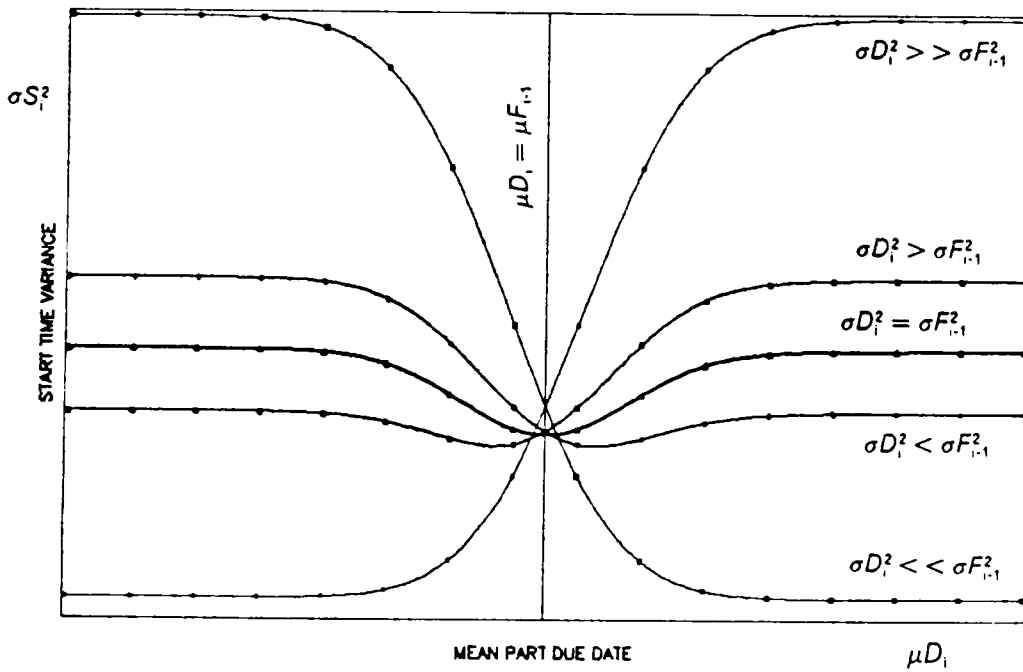


Figure 6. Relationships in Two Station Problem

utions is even more intractable. It was therefore necessary to use other means to understand this behavior.

As mentioned earlier, the second term in the objective function is a monotonically increasing function of the variance of the operation finish time  $\sigma F_i^2$ , which in the deterministic processing case is simply the variance of  $S_i = \max(F_{i-1}, D_i)$ . The behavior of  $\sigma S_i^2 (= \sigma F_i^2)$  as a function of  $\mu D_i$  is shown in Figure 6b for different combinations of  $\sigma F_{i-1}^2$  and  $\sigma D_i^2$ . If these variances are equal,  $\sigma S_i^2$  attains its minimum value at  $\mu F_{i-1} = \mu D_i$ , and tends asymptotically to either variance value on each side. As the difference between the variances increases, the minimum variance point shifts towards the smaller variance and becomes less distinct, and the function goes asymptotically to the variance of the dominating variable,  $\sigma F_{i-1}^2$  for small  $\mu D_i$  and  $\sigma D_i^2$  for large  $\mu D_i$ . Eventually, it becomes altogether asymptotic between the two extremes. It can be concluded that  $\sigma S_i^2$  and  $\sigma F_i^2$  are not convex functions of  $\mu D_i$ .

The second term in the objective function is a monotonically increasing function of  $\sigma F_i^2$ , which as shown above is not a convex function of  $\mu D_i$ . Therefore this term in the objective function is not a convex function of  $\mu D_i$ . It does have bounds, however, determined by the local and asymptotic extrema. The total cost as a function of  $\mu D_i$  is the summation of this term and the first term which is convex, and will have a finite minimum since both terms have finite minima. However, it could not be shown to be unimodal in general. The exact properties depend on the relative magnitudes of costs and variances at the stations.

In order to further understand the behavior of the function, a number of numerical tests of convexity were performed for a variety of cost and variance structures, using the linear combination method. The tests were limited to situations where the part



has a non-zero cost but is less expensive than the subassembly, so that the region of interest is limited to the left of  $\mu F_{i,1}$ . In the problems tested, the 2 station objective function turned out to be reasonably well behaved.

The function satisfies the convexity conditions in problems that are balanced in costs and variances, but when the values get considerably unbalanced, convexity gets violated in some problems. However the tests showed that the function retains strong quasi-convexity in cases where convexity is violated. To illustrate the behavior of the objective function, the contours for a sample 2 station problem are graphed in Figure 7. This problem is unbalanced in variances and the objective function is not convex in the region marked. Some of the contours appear to consist of piecewise linear segments, implying only strict quasi-convexity. However, that is due to the grid accuracy limitations in plotting the contours, and numerical tests verify that the function is strongly quasi-convex over the entire range. The optimal solution is shown by the intersection of the two reference lines.

Extensive numerical tests have been done for up to three station problems, and the same property holds for all the distributions considered. Consequently, unconstrained nonlinear programming methods can be applied to determine the minimum cost decisions. Of course, these properties may not continue to hold true for larger problems, but in the absence of analytical methods it is difficult to draw conclusions either way. However, since the quasi-convexity holds true for up to three station problems, it is expected to hold true for larger problems as well. This property has been subsequently confirmed by the fact that the solutions using this approach are never inferior to those obtained by other methods, and multiple starts always resulted in solutions in the same vicinity.

Problem parameters:  $\mu F_0 = 25$ ,  $\sigma F_0 = 0.5$ ,  $\sigma D_1 = 4$ ,  $\sigma D_2 = 0.001$ ,  $\mu P_1 = 5$ ,  $\sigma P_1 = 0$

$CE_1 = 1$ ,  $CE_2 = 5$ ,  $CS_1 = 1$ ,  $CS_2 = 2.5$

Optimal Solution:  $\mu D_1^* = 21.12$ ,  $\mu D_2^* = 30.86$ ,  $TC^* = 7.79$

Region of Non-convexity

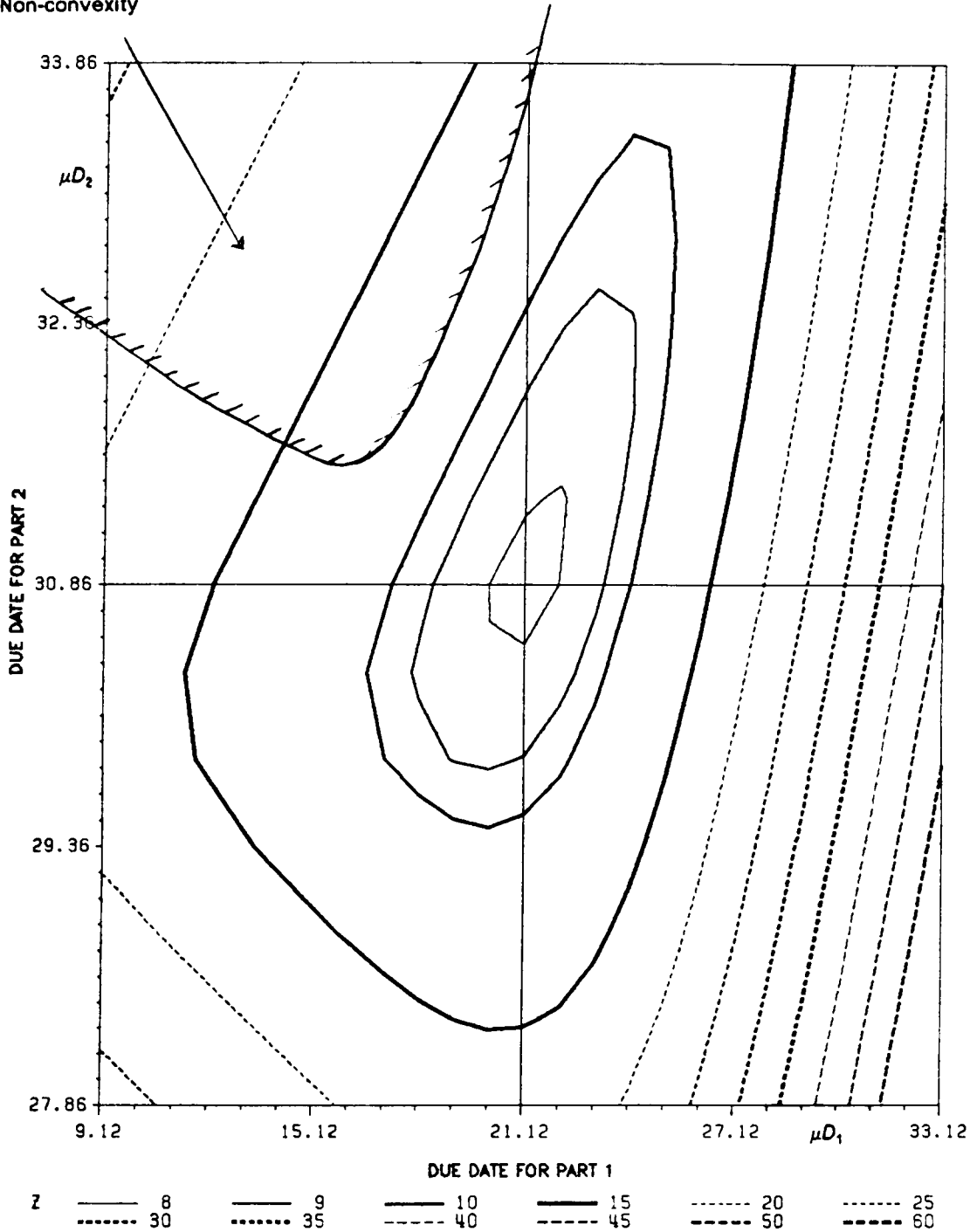


Figure 7. Objective Function Contours for a 2 Station Problem

Considering the difficulty of obtaining expressions for the first derivatives even for small problems, the choices to search for the optimal solution are restricted to methods that do not require derivatives. Note that the numerical evaluation of the function for a given set of decisions can be done without difficulty by recursive application of equations (3.1), (3.2) and (4.9) together with the formulae for the moments of the maximum function for the distribution being considered. Consequently, two approaches were tried to search for the optimal solution. The first one is a function evaluation based nonlinear programming (NLP) search approach, and the second one is based on the optimization of individual stages in isolation.

As part of the first approach, two different methods have been experimented with. The Hooke and Jeeves method was first applied since it is a simple, easily implemented and reliable method (Avriel, 1976). Subsequently, Powell's method was implemented, since it is regarded as one of the best unconstrained optimization methods without the use of derivatives (Himmelblau, 1972, Avriel, 1976). Multiple starts were used to check if the solution always converged to the same solution, in an attempt to verify if there was multimodality in the objective function.

### **Independent Solution Method**

The other solution approach attempted is based on the single station solution, optimizing each successive station in isolation. This method is subsequently referred to as the Independent Solution method (ISM). The idea is to repetitively apply the method of Section 4.2 to each station, starting at station 1, according to the following scheme.

1. At station  $i$ , determine the independent optimal policy  $\mu D_i^*$ , given the input to the station  $\mu F_{i-1}$ ,  $\sigma F_{i-1}^2$  and  $\sigma D_i^2$ , and the associated cost parameters  $CS_i$  and  $CE_i$ .
2. Given the independent optimal policy for station  $i$ , determine the mean  $\mu S_i$  and the variance  $\sigma S_i^2$  of operation start time, using the formulae for the moments of the maximum.
3. Determine the mean  $\mu F_i$  and the variance  $\sigma F_i^2$  of operation finish time using the means and variances of operation start times and the processing times at station  $i$ . These finish times become the input to the next station.
4. Continue steps 1-3 until decisions are determined for all stations.

### **4.3.2 Comparison of Solutions**

In order to assess the quality of the solution obtained and the relative performance of different methods, the DP methodology was also implemented using the computer programs from Das (1985). A number of sample 2 and 3 station problems was run using these three methods, both without and with prespecified due dates (by setting  $\sigma D_{N+1} = 0$ ). The tests were done for normally distributed delivery times with deterministic processing times, since the DP solution methodology has been developed for that scenario only. To ensure consistency, the total expected costs for the decisions obtained from each method were determined by a common cost calculation routine. The results for the sample problems are shown in Table 4.

It can be seen that the solution from the nonlinear optimization technique is always superior to all the other methods. In no case does it result in a solution inferior to the other methods. Experimentation with multiple starts occasionally resulted in

**Table 4. Comparison of NLP, ISM and DP Methods for 3 Station Problems**

Initial Problem:

$$\mu F_0 = 15, \sigma F_0 = \sigma D_1 = \sigma D_2 = 2, \sigma D_3 = 0.001, \mu P_1 = 5, \sigma P_1 = 0$$

$$CE_1 = CE_2 = 1, CE_3 = 8, CS_1 = 1, CS_2 = 2.5, CS_3 = 4$$

No.	Variation	Meth	TC	%Error	$\mu D_1$	$\mu D_2$	$\mu D_3$
1	--	NLP	11.806		15.00	19.78	27.26
		ISM	11.811	0.04	15.00	19.66	27.23
		DP	15.237	29.06	13.12	15.45	26.19
2	$CE_3 = 20$	NLP	14.224		15.00	19.81	28.07
		ISM	14.232	0.06	15.00	19.66	28.02
		DP	18.995	33.54	12.26	15.19	26.99
3	$CE_3 = 3$	NLP	9.414		15.00	19.75	26.35
		ISM	9.416	0.02	15.00	19.66	26.32
		DP	12.038	27.88	13.78	15.70	25.36
4	$\sigma D_1 = 4$	NLP	15.526		12.51	19.88	27.35
		ISM	18.815	8.31	15.00	19.92	28.29
		DP	19.834	27.75	9.57	14.94	26.12
5	$\sigma D_1 = 0.5$	NLP	8.554		16.88	19.90	27.59
		ISM	9.684	13.21	15.00	19.50	26.76
		DP	10.882	27.21	15.28	16.66	26.47
6	$\sigma F_1 = 4, \sigma D_1 = 4$	NLP	19.735		15.00	22.13	29.72
		ISM	20.614	4.45	15.00	20.07	29.08
		DP	31.376	58.99	8.14	12.38	26.92
7	$CS_1 = 10, CS_2 = 11.5,$ $CS_3 = 13, CE_3 = 26$	NLP	34.327		12.49	17.50	26.20
		ISM	35.469	3.33	11.22	16.26	25.99
		DP	38.465	12.05	9.68	14.77	25.89
8	$CS_1 = 10, CS_2 = 11.5,$ $CS_3 = 13, CE_3 = 26,$ $\sigma D_1 = 4, \sigma D_2 = 4$	NLP	44.136		8.73	13.67	26.15
		ISM	44.175	0.09	9.03	13.92	26.20
		DP	45.830	3.84	6.59	11.66	25.95
9	$CE_1 = 2, CS_2 = 3.5,$ $CS_3 = 5, CE_3 = 10$	NLP	14.810		15.76	19.86	27.60
		ISM	14.912	0.69	16.22	19.84	27.83
		DP	17.271	16.62	14.91	16.20	26.81
10	$CE_1 = 2, CS_2 = 3.5,$ $CS_3 = 5, CE_3 = 10,$ $\sigma D_1 = 4$	NLP	21.091		13.88	20.08	27.85
		ISM	23.634	12.06	16.93	20.14	29.47
		DP	24.772	17.45	13.35	15.36	27.17

different values for the  $\mu D$ 's, but the values were in the same vicinity, and the differences in objective function value were insignificant. This suggests that the objective function could be relatively insensitive in the neighborhood of the optimum. None of the tests produced any significant indications of multi-modality. The number of evaluations required ranged from about 100 for 2 station problems to about 1000 for 10 station problems. It must be noted, however, that the computational effort required in a function evaluation also increases with the number of stations.

While the DP methodology performed reasonably well without due date setting, its performance deteriorated considerably when due dates were introduced. Examination of the approximation reveals that one of the terms in the regression equation is  $(\sigma D_i)^C$ , with  $C = -0.58349$ , so setting  $\sigma D_{N+1} = 0$  to fix due dates drives that term to infinity. Even when very small values such as 0.001 were used for  $\sigma D_{N+1}$  to set due dates, the performance remained poor, in some cases resulting in errors of over 50%. Later when larger problems of up to 10 variables were tried, it was found that even without due date setting, the performance of the DP approach deteriorated with problem size. The results of some 10 station problems are shown in Table 5.

A surprising observation from the runs was the remarkable performance of the independent station solution method. In many of the 2 and 3 station problems it gave results almost identical to the nonlinear programming based solution, being within about 5% of the optimum in many of the problems, and it requires only a fraction of the computational effort. Compared to the DP solution method, it was superior in most cases even without due dates, and at worst, it was only marginally inferior. This ranking in performance continued to hold even when larger problems of up to 10

**Table 5. Comparison of NLP, ISM and DP Methods for 10 Station Problems**

Initial Problem:

$$\mu F_0 = 15, \sigma F_0 = \sigma D_1 = 2, i = 1, \dots, 9, \mu P_1 = 5, i = 1, \dots, 9, \mu P_{10} = 0, \sigma P_1 = 0, i = 1, \dots, 10$$

$$CE_1 = 1, i = 1, \dots, 9, CE_{10} = 30, CS_1 = 1, CS_2 = 2.5, CS_3 = 4, CS_4 = 5.5, CS_5 = 7,$$

$$CS_6 = 8.5, CS_7 = 10, CS_8 = 11.5, CS_9 = 13, CS_{10} = 15$$

Pr. No.	Variation	Method	$\sigma_{10} = 2$		$\sigma_{10} = 0.001$	
			TC	% Error	TC	% Error
1	--	NLP	70.887		51.598	
		ISM	70.931	0.06	51.698	0.19
		DP	92.056	29.86	75.131	45.61
2	$CE_{10} = 60$	NLP	81.481		56.722	
		ISM	81.539	0.07	56.858	0.24
		DP	102.58	25.89	82.760	45.91
3	$CE_{10} = 10$	NLP	55.561		44.181	
		ISM	55.588	0.05	44.235	0.12
		DP	74.898	34.80	64.095	45.07
4	$\sigma D_1 = 4$	NLP	74.943		56.119	
		ISM	77.649	3.61	60.143	7.17
		DP	95.631	27.60	78.806	40.43
5	$\sigma D_1 = 0.5$	NLP	67.086		45.353	
		ISM	68.495	2.10	48.586	7.13
		DP	72.912	8.68	47.850	5.51
6	$\sigma F_1 = 4, \sigma D_1 = 4$	NLP	79.703		61.423	
		ISM	83.526	4.80	67.316	9.59
		DP	140.10	75.78	128.05	108.5
7	$CS_1 = 10, CS_2 = 12.5, \dots, CS_{10} = 25, CE_{10} = 50$	NLP	111.37		80.810	
		ISM	113.87	2.24	85.712	6.07
		DP	121.63	9.21	95.891	18.66
8	$CS_1 = 10, CS_2 = 12.5, \dots, CS_{10} = 25, CE_{10} = 50$	NLP	116.24		86.327	
		ISM	118.66	2.09	91.200	5.64
		DP	125.51	7.98	99.959	15.79
9	$CE_1 = 2, CS_2 = 3.5, \dots, CS_{10} = 16, CE_{10} = 32, \sigma D_1 = 4$	NLP	82.505		62.723	
		ISM	88.420	7.17	71.373	13.79
		DP	108.24	31.20	89.337	42.43

variables were run, although the decisions from the independent solution method were not always as close to the optimal values as for smaller problems.

The safety buffer rule ( Das and Sarin, 1987) using a safety factor of 1 standard deviation was also tested along with the other methods, and it consistently performed the worst. Because it applies a fixed safety factor to each problem regardless of the costs involved, it yields the same decisions even if the costs change. The errors ranged from 12% to 96% in the 3 station problems and from 17% to 97% in the 10 station problems.

From the results of the sample problems, it appears that the DP methodology cannot be used effectively in its present form in the presence of due dates or when variances are small. The errors in the DP model arise out of the regression approximation for the transferred derivative. The form of the equation is constrained by the necessity of a form similar to the stage derivative, to enable closed form solution. Attempting to improve the capability of the regression equation, to enable setting due dates, is likely to result in more complicated forms that may not be possible to optimize in closed form, thus defeating its purpose. In any case, there is no guarantee of finding a suitable form. The difficulty of extending the DP approach to non-normal distributions and random processing times is also a serious limitation on its generality.

In view of the clear rankings in performance and the other difficulties outlined above, the DP approach was not found attractive enough to pursue further. The approach taken subsequently has been to initially use NLP to solve the problem. These solutions can be used to generate understanding and insights about the single job problem and its variations, and possibly help in developing simpler methods later.



The remarkable performance of the single station solution approach (ISM) suggested that reasonably good solutions could be obtained with little computational effort. Moreover, if some suitable correction factors could be determined, that would bring its solution even closer to the optimal solution. The determination of these correction factors and the approximate solution method is addressed in Section 4.5. The solution of the multi-station problem with random processing times is discussed next.

#### **4.4 Incorporation of Random Processing Times**

Variability in processing times could affect the decisions because of the changes that occur in the variances from station to station, since now  $\sigma F_i^2 = \sigma S_i^2 + \sigma P_i^2$ . In order to assess the changes in costs and decisions due to this variation, the same set of problems as before was run with processing time variances added. The solution methods work in exactly the same way as before, the only changes being in the calculation of finish time variances. Table 6 compares the results for the sample 3 station problems for different levels of processing time variance using the normal distribution.

According to the results of these runs, random processing affects the costs and decisions in the following ways.

**Table 6. Effect of Processing Time Variability on Costs and Decisions for 3 Station Problems**

Initial Problem:

$$\mu F_0 = 15, \sigma F_0 = \sigma D_1 = \sigma D_2 = 2, \sigma D_3 = 0.001, \mu P_i = 5,$$

$$CE_1 = CE_2 = 1, CE_3 = 8, CS_1 = 1, CS_2 = 2.5, CS_3 = 4$$

No.	Variation	$CV_i$	TC	$\mu D_1$	$\mu D_2$	$\mu D_3$
1	--	0	11.806	15.00	19.78	27.26
		0.2	13.999	15.00	19.95	27.58
		0.4	18.684	15.00	20.22	28.25
2	$CE_3 = 20$	0	14.224	15.00	19.81	28.07
		0.2	17.140	15.00	20.05	28.64
		0.4	23.293	15.00	20.47	29.85
3	$CE_3 = 3$	0	9.414	15.00	19.75	26.35
		0.2	10.885	15.00	19.83	26.36
		0.4	14.087	15.00	19.92	26.40
4	$\sigma D_1 = 4$	0	15.526	12.51	19.88	27.35
		0.2	17.346	12.75	20.05	27.69
		0.4	21.549	13.13	20.34	28.43
5	$\sigma D_1 = 0.5$	0	8.554	16.88	19.90	27.59
		0.2	11.819	16.26	20.01	27.66
		0.4	17.172	15.85	20.20	28.19
6	$\sigma F_1 = 4, \sigma D_1 = 4$	0	19.735	15.00	22.13	29.72
		0.2	21.133	15.00	22.11	29.87
		0.4	24.677	15.00	22.07	30.31
7	$CS_1 = 10, CS_2 = 11.5,$ $CS_3 = 13, CE_3 = 26$	0	34.327	12.49	17.51	26.20
		0.2	41.321	12.26	17.45	26.41
		0.4	56.431	11.96	17.29	26.89
8	$CS_1 = 10, CS_2 = 11.5,$ $CS_3 = 13, CE_3 = 26,$ $\sigma D_1 = 4, \sigma D_2 = 4$	0	44.136	8.73	13.67	26.15
		0.2	50.442	8.78	13.82	26.37
		0.4	64.711	8.83	14.05	26.86
9	$CE_1 = 2, CS_2 = 3.5,$ $CS_3 = 5, CE_3 = 10$	0	14.810	15.76	19.86	27.60
		0.2	17.584	15.83	20.03	27.94
		0.4	23.503	15.93	20.26	28.63
10	$CE_1 = 2, CS_2 = 4.5,$ $CS_3 = 5, CE_3 = 10,$ $\sigma D_1 = 4$	0	21.091	13.88	20.08	27.85
		0.2	23.248	14.16	20.24	28.21
		0.4	28.321	14.62	20.49	28.97

1. The operation finish time variances are significantly higher with random processing times than with deterministic processing times. This is due to the fact that now  $\sigma F_i^2 = \sigma S_i^2 + \sigma P_i^2$ , whereas with deterministic processing times  $\sigma F_i^2 = \sigma S_i^2$ .
  
2. There is a clearly rising trend in costs as processing time variability increases, due basically to the increase in finish time variances. Increasing the processing time coefficients of variation from 0 to 0.4 for all operations increased the optimal policy costs by as much as 100% in problem 5 which had the lowest variance to start with, and by 60% in problems 1, 2, 3, 7 and 9. The effect on the costs in problems that had high part variances to start with is of the order of 30-40%, because the marginal increase in variances is lower.
  
3. The interval between the part due dates and between the completion times for successive stations increases for all the problems. For example, in problem 5, the interval between the part due dates for the first and third stations,  $(\mu D_3 - \mu D_1)$ , for the deterministic processing case is 10.71 units. This interval increases to 11.44 units when all processing time standard deviations are increased to 1 unit ( $CV = 0.2$ ), and to 12.34 units when they are further increased to 2 units ( $CV = 0.4$ ). This increase arises due to the building in of safety allowances for the processing times as well. Again, the same problems as above showed the highest increases in safety times for processing operations.
  
4. The addition of processing time variances appears to damp the effect of decisions taken at previous stations, since the stages are linked together mainly by the finish time variances. The addition of a very large processing time variance may allow very little of the effect of a decision to be transmitted to variance at the next

station. By contrast, deterministic processing times transmit totally the effect of a change in the variance.

In the context of the approximation technique developed in the next section, this means that the correction factors may need to be further adjusted to account for the processing time variances. These adjustments are discussed and derived in Section 4.5.4.

## **4.5 Development of Approximate Solution Method for**

### **Normal Distribution**

The optimal solution to the problem is a function of relative variances of parts and subassemblies, the relative costs of parts and subassemblies, the number of stations on the line, and the relative costs between stations all along the line. The ISM method ignores downstream factors, specifically the effect a decision taken at a station has on the finish time variance, which is the input to the subsequent station and affects later ones as well. In order to obtain good solutions this effect must be considered to give the method some limited look-ahead capability. In this section, an approximate solution technique based on the independent solution method is developed to correct for some of these effects. While it may not be possible to capture all the factors, at least the most dominant ones must be considered so that reasonably good solutions may be obtained.

One method of incorporating some look-ahead capability in a piecewise solution approach is to solve a set of overlapping sub-problems. For example, rather than solve a series of single station problems, one could solve a 2 or 3 station problem optimally, use the results to 'freeze' the decision at the first station, and solve another such problem starting from the next station. Such an approach, termed the Rolling Horizon method (ROH) would consider at least some of the downstream factors. It would definitely be closer to the optimal solution than a totally independent solution.

Initial investigation showed that the errors from the Rolling Horizon method, with a horizon length of 3 stations, fall in a much narrower range than the ISM approach. However, since all possible factors are still not considered, it is still not good enough to be used directly without any correction, especially in long lines. Therefore, even for this method we need to determine appropriate correction factors to bring the solutions closer to optimality. In view of the greater number of parameters now internal to each sub-problem, the reasons and magnitudes of the corrections are more difficult to isolate. On the other hand, with the ISM approach, the correction factors are relatively easier to understand and determine. It was therefore decided to initially base the approximate solution on the totally independent solution with correction factors, and later incorporate some features of the ROH method to improve its performance.

The approach taken has been to solve a set of identical problems using the NLP and ISM methods, and use the solutions to identify the differences and their sources. For the purpose of the study, measurement of the effects of changes has been done at the first station. The same changes can similarly be applied to the remaining stations. It has also been assumed in creating the data sets that the subassembly is

at least as expensive as the part, and that the total costs are non-decreasing down the line.

Preliminary tests showed that if the costs and variances of the part and subassembly at a station were identical, the ISM method resulted in exactly the same decision at that station as those obtained by the NLP method. However, the solutions differ if the variances or the costs are different. The magnitude of the difference is a function of the amount of change, the downstream cost ratios between stations, and the number of stations downstream of the station at which the changes are made. Since the last station cannot affect any stations, no corrections are required there. The key parameters that affect the presence and the magnitude of the error are identified as:

1. Difference between part and subassembly standard deviation  $DS_i = \sigma D_i - \sigma F_{i-1}$
2. Cost proportion of the part at the station  $CP_i = \frac{CE_i}{CS_i + CE_i}$
3. Cost ratio between stations  $CR_i = \frac{CS_{i+1} + CE_{i+1}}{CS_i + CE_i}$
4. Number of stations downstream of the station =  $N - i$

The individual sources of error are now examined individually.

#### **4.5.1 Sources of Error in Independent Solution Method**

##### 1. Difference in Variance

If the part and subassembly costs and variances are the same ( $CP_i = 0.5$  and  $DS_i = 0$ ), then the optimal value is  $\mu D_i^* = \mu F_{i-1}$ . This is because the minimum variance

of finish time ( $\sigma F_i^2$ ) is realized at that value, as can be seen in Figure 6b. If  $CP_i = 0.5$  but  $DS_i > 0$ , the ISM method still results in  $\mu D_i = \mu F_{i-1}$ . However, the optimal policy reduces the value of  $\mu D_i$  so that  $\mu F_{i-1}$  can become more dominant and thus reduce  $\sigma F_i^2$ . The greater the value of  $DS_i$ , the greater is the magnitude of this shift, in order to provide for greater protection against variability. The magnitude of the shift depends on the values of  $DS_i$  and  $CR_i$ , since a balance must be struck between the costs at successive stages. Also, since the effects of the change in variance are transmitted down the line, the error increases as the number of downstream stations increases.

## 2. Effect of Cost Proportion

If  $DS_i = 0$  and  $CP_i < 0.5$ , the single station solution reduces the value of  $\mu D_i$  by an amount depending upon the value of  $CP_i$ . The smaller the relative cost of part inventory, the greater the reduction, in order to prevent the more expensive subassembly having to wait for the part. This expediting of the part tends to increase  $\sigma F_i^2$ , since its minimum value occurs at  $\mu D_i = \mu F_{i-1}$ , so the optimal solution moves the decision back in the direction of  $\mu F_{i-1}$ . Again, the magnitude of the shift depends on the  $CP_i$  and  $CR_i$  values and the number of stations remaining.

## 3. Effect of Cost Ratio

If  $DS_i = 0$  and  $CP_i = 0.5$ , varying the values of  $CR_i$  makes no difference to the optimal decision and it remains at  $\mu D_i = \mu F_{i-1}$ . However, if either  $DS_i \neq 0$  or  $CP_i \neq 0.5$ , then the magnitude of the error increases as  $CR_i$  increases. Although all downstream ratios affect the error, the effect declines rapidly with increasing distance between the station being considered and the station where the cost ratio changes. The immediate cost ratio remains the most dominant, unless there are major changes in the cost structure down the line.

#### 4. Effect of Number of Downstream Stations

Again, if  $DS_i = 0$  and  $CP_i = 0.5$ , varying the number of downstream stations makes no difference to the decision at that station. However, if either  $DS_i \neq 0$  or  $CP_i \neq 0.5$ , then the magnitude of the error increases as the number of downstream stations increases.

To summarize the results of the preliminary tests, the relative variances and cost proportions determine whether or not a correction is required. If a correction is required, the magnitude depends on the extent of this difference, the cost ratio and the number of downstream stations.

A series of experiments were then conducted to determine these correction factors empirically. The strategy was to vary only one of the factors one at a time at the first station, and measure the effect of the change by the difference  $\Delta_1$  between the ISM decision at that station and the solution from NLP, that is,  $\Delta_1 = \mu D_1^I - \mu D_1^N$ .

The objective was to develop a set of approximating equations that enable determination of the appropriate correction factor at a station for a given set of parameter values. For simplicity, the correction factors were taken to be additive. To the extent possible, it was attempted to keep the relations linear in standard deviations, which can be interpreted as scaling factors for the time unit, and in terms of ratios in costs so that the approach is not sensitive to scale changes.

The experimental data were set up with  $\mu F_0 = 15$ ,  $\sigma F_0 = 2$ , with the values of CR ranging from 1.0 to 12.0 at the first station as well as down the line. Since costs are assumed to be increasing down the line, values of CR less than 1 have not been considered. The upper limit of CR = 12 is chosen as a reasonable estimate of the



largest magnitude of value addition expected over a single stage. The length of the lines ranged from 1 to 10 stations, and all processing times were deterministic. The experiments were first run with all  $DS_i = 0$  and  $CP_i = 0.5$  to confirm that changing the ratios and number of stations did not cause any errors by themselves. When this was found to be true, the other experiments were run by varying the  $DS_i$  and  $CP_i$  values separately to isolate the effects of the change and measuring the resulting error. The ranges of variation considered for the  $DS_i$  and  $CP_i$  values are stated in the corresponding sections.

#### **4.5.2 Correction for Variance**

The effect of changing the variance was tested for different values of  $\sigma D_i$  and  $\sigma F_0$  in the range 0.5 to 8 units. Initially, a range of 0.5 to 4 units was considered, and enlarging the range to 0.5-8 showed that the behavior remained similar. Hence, this range of variability was found to be general enough to study the system behavior changes. The  $DS_i$  values ranged from -6.0 to 6.0. Figure 8 shows the results of these runs. The tests showed that the error is almost linear in the number of stations. The dependence of the intercept (AV) and slope (BV) of this linear relationship on the magnitude of  $DS_i$  is shown in Figure 9a. For the range of variation considered, the intercept and the slope are approximately proportional to  $DS_i$ .

Further experiments were conducted to measure the effect of changing all cost ratios (CR's) down the line uniformly. These are shown in Figure 9b for a value of  $DS_i = 2$ , for CR values between 1 and 4. The results show that the error is still almost linear in the number of stations, but the intercepts and the slopes differ for varying

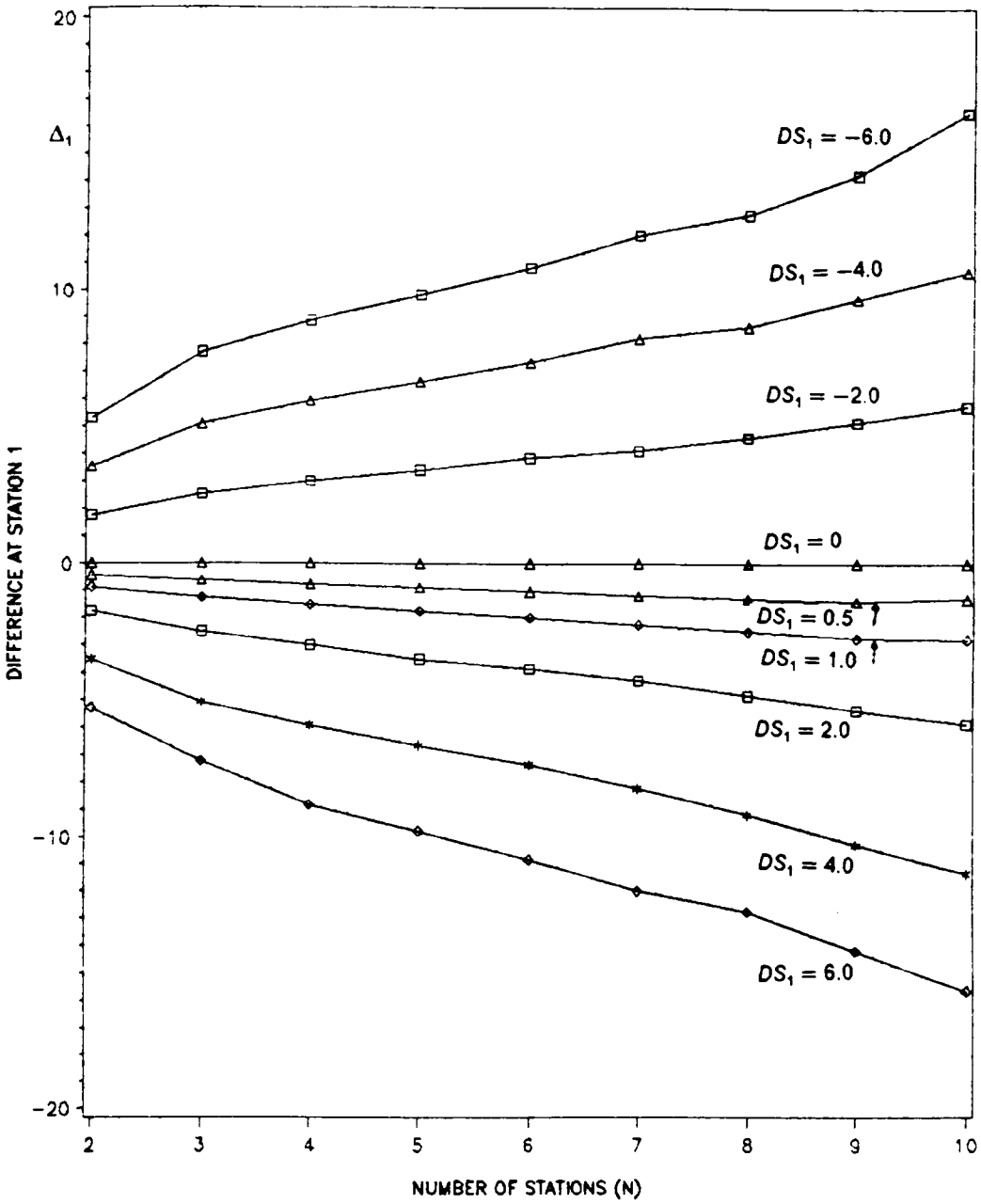
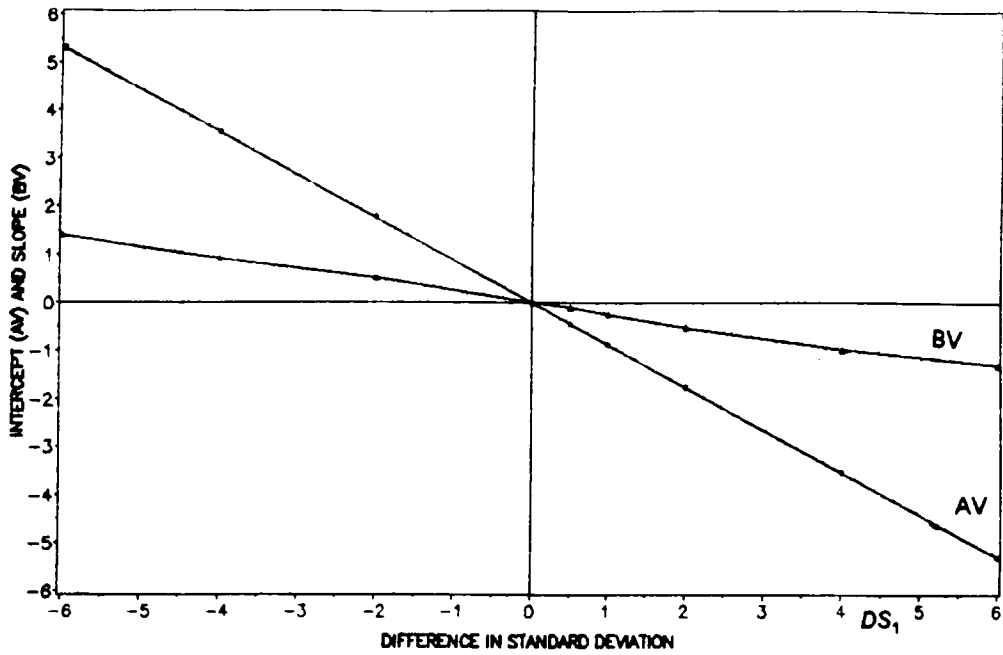


Figure 8. Effect of Change in Variance for Different DS Values

a) Intercept and Slope of Variance Correction vs  $DS_1$



b)  $\Delta_1$  vs. Number of Stations for different CR values

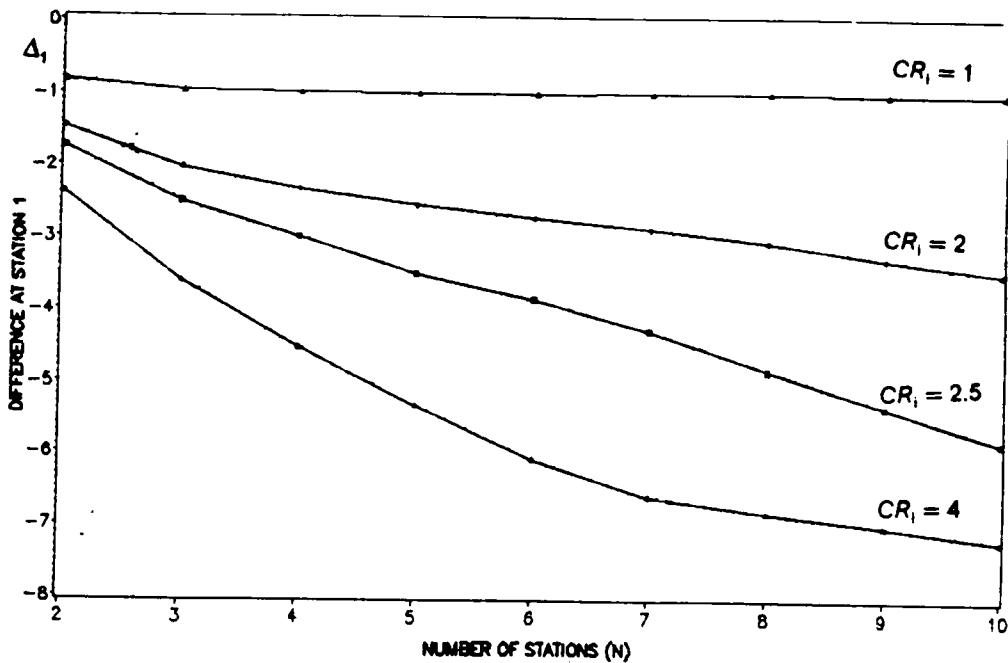


Figure 9. Effect of Change in Variance

values of  $CR_i$ . Similar trends were observed when higher cost ratios were considered. The effect of varying selected downstream ratios was also studied, and it was found that the effect of the  $CR_i$  value at that station was the most dominant. However, if there is a large increase in downstream cost ratio, the decisions are affected although the effect declines rapidly as the distance to the station where the change occurs increases. In such a case, experimentation showed that the error is governed approximately by

$$ACR_i = \max \left( \frac{CS_{i+k} + CE_{i+k}}{CS_i + CE_i} \right)^{1/k}, k = 1, \dots, N - k$$

where  $ACR_i$  is the adjusted cost ratio. This adjustment, though empirically determined, is consistent with the results in case the same cost ratio is maintained between all stations. Changes in downstream part variances did not make an appreciable difference.

Now that the most significant sources of error have been identified, equations to determine the correction factor can be formulated. To completely specify the equations, the values of the intercept ( $AV_i$ ) and the slope ( $BV_i$ ) per unit  $DS_i$  have to be determined for a particular value of  $ACR_i$ . The relationship between these parameters and the ratios was found to be nonlinear, so a quadratic approximation was fitted using the data points generated. The resulting equations for the variance correction factor for  $i < N$  and deterministic processing times are

$$\Delta_i^V = (AV_i + \sum_{k=i+1}^{N-1} BV_k) \quad (4.10)$$

$$AV_i = - (-0.01461ACR_i^2 + 0.31393ACR_i + 0.14990)DS_i \quad (4.11)$$

$$BV_i = - (-0.00927ACR_i^2 + 0.14457ACR_i - 0.09764)DS_i \quad (4.12)$$

The subscript  $k$  does not appear inside the summation in equation (4.10) because a constant value is being added  $(N-i-1)$  times. However, this form has been used for consistency with the form used in Section 4.5.4. These equations can be used to determine the necessary correction factors due to a difference in variance at the station being considered.

### **4.5.3 Correction for Cost Proportion**

The effect of changing the cost proportion at station 1 was tested at five different levels of  $CP_1$  in the range between 0.05 and 0.5. Since the subassembly is assumed to be at least as expensive as the part, values of  $CP$  greater than 0.5 have not been considered. The lower limit of  $CP = 0.05$  has been chosen because values below that correspond to part cost below 5% of the total cost at the station, which would make the impact of part cost relatively insignificant in decision making. The results of these tests are shown in Figure 10, and show that the error is almost linear in the number of stations. The relationship between the intercept ( $AP$ ) and the slope ( $BP$ ) with  $(0.5 - CP_1)$  is shown in Figure 11a, and can be approximated as being directly proportional.

Further experiments were conducted to measure the effect of changing all cost ratios ( $CR$ 's) down the line uniformly. These are shown in Figure 11b for a value of  $CP_1 = 0.3$ , which falls approximately in the middle of the range considered, for  $CR$  values between 1 and 4. The results show that the error is still almost linear in the

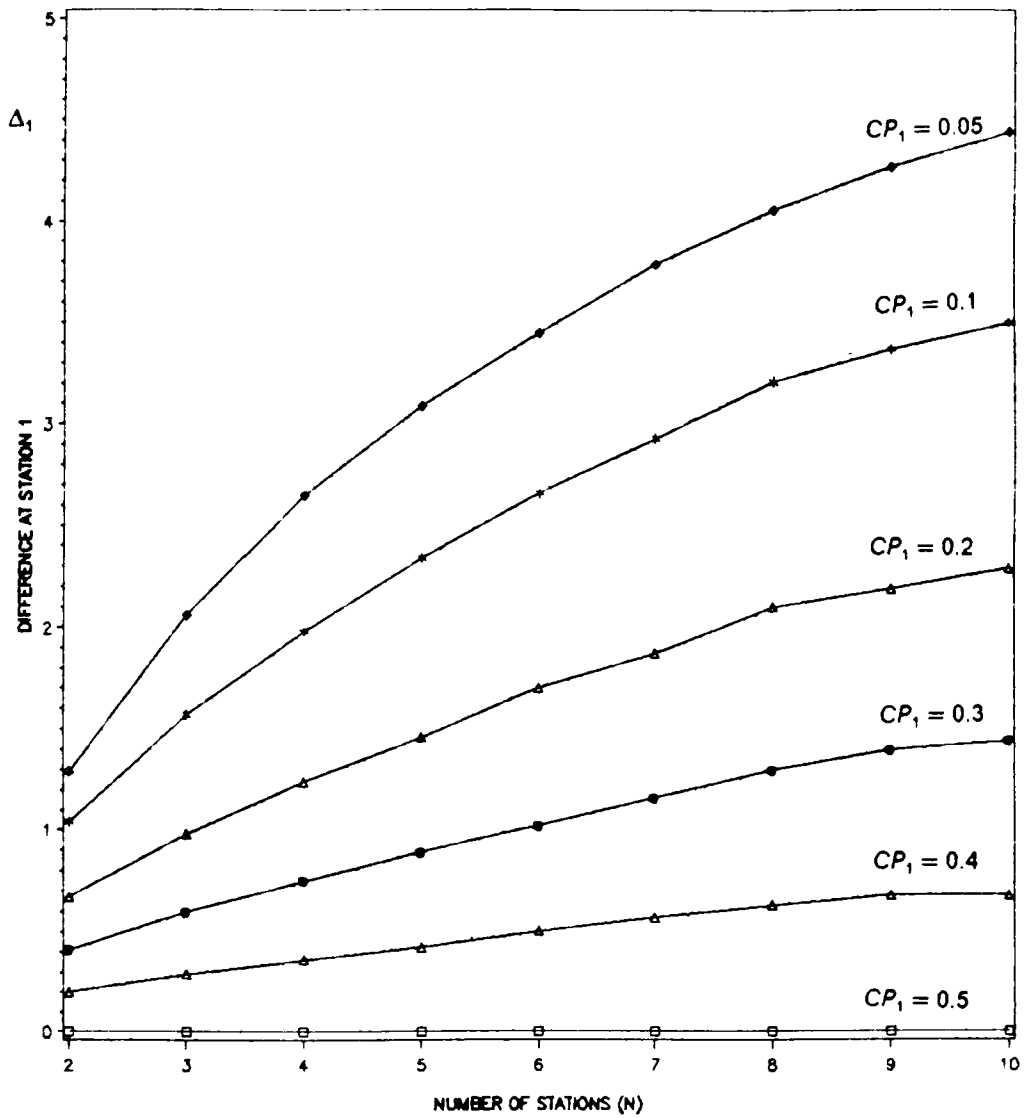
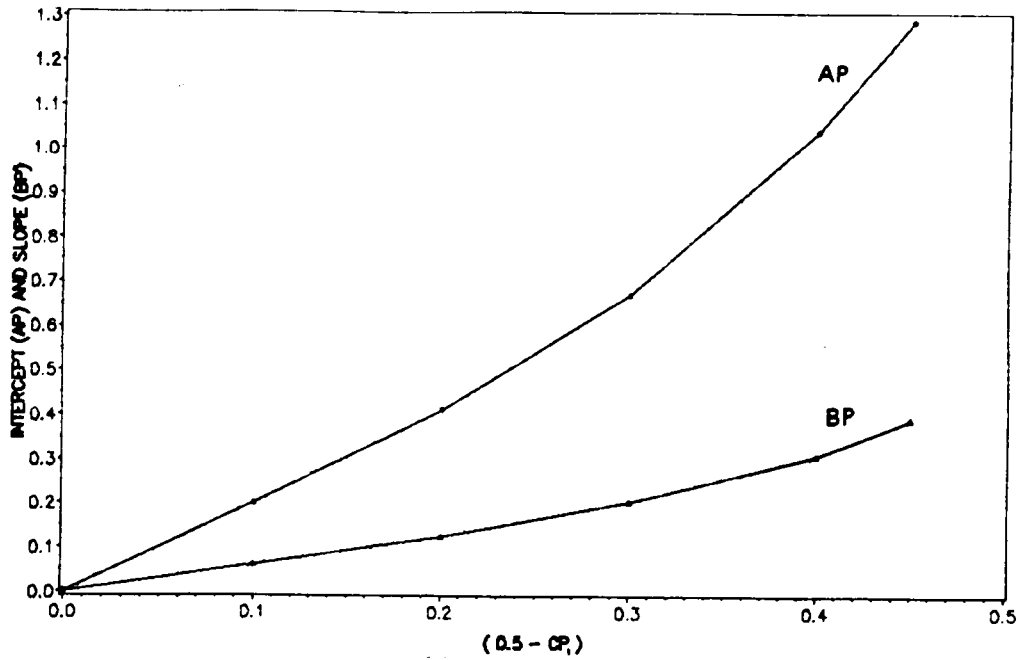


Figure 10. Effect of Change in Cost Proportion for Different CP Values

a) Intercept and Slope of Cost Proportion Correction vs  $(0.5 - CP_1)$



b)  $\Delta_1$  vs. Number of Stations for different CR values

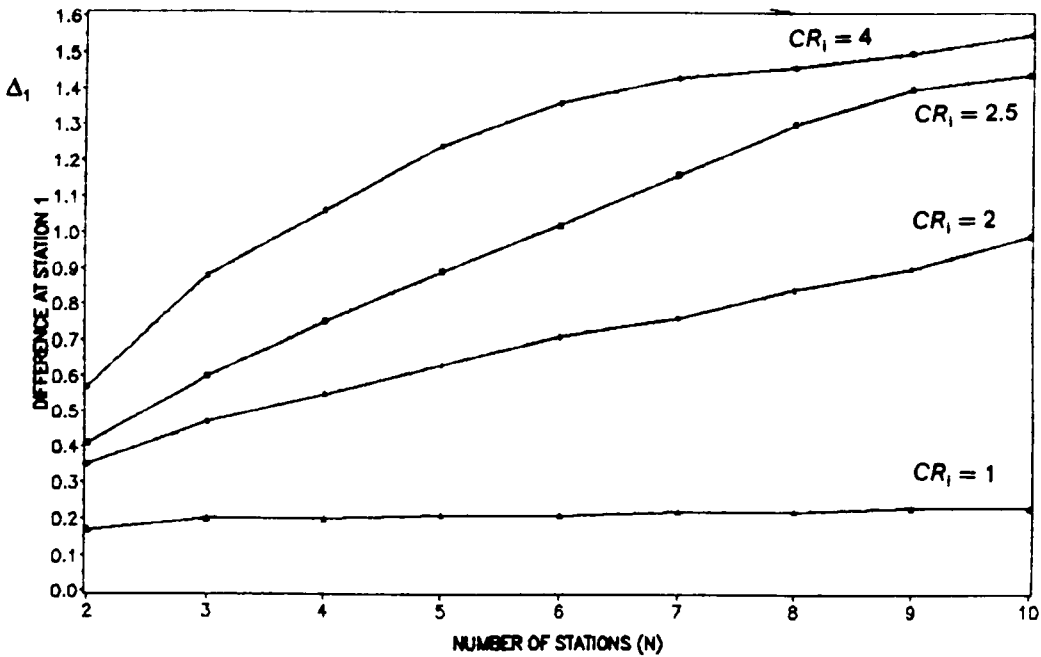


Figure 11. Effect of Change in Cost Proportion

number of stations but the intercepts and the slopes differ for varying values of  $CR_i$ . Similar trends were observed when higher cost ratios were considered. The effect of varying selected downstream ratios was also studied, and it was found that the effect of the  $CR_i$  value at that station was the most dominant. However, if there is a large increase in downstream cost ratio, the decisions are affected although the effect declines rapidly as the distance to the station where the change occurs increases. As in the variance correction, this error is governed approximately by the adjusted cost ratio,  $ACR_i = \max\left(\frac{CS_{i+k} + CE_{i+k}}{CS_i + CE_i}\right)^{1/k}$ . This adjustment is also consistent with the results in case the same cost ratio is maintained between all stations. Downstream changes in  $CP_i$  did not make an appreciable change in the error.

In order to develop equations determining the correction factor for  $CP_i$  it is necessary to determine the values of the intercept ( $AP_i$ ) and the slope ( $BP_i$ ) for a particular value of  $ACR_i$ . The relationship between these parameters and the ratios was also found to be nonlinear, so a quadratic approximation was fitted using the data points generated. The resulting equations for the proportion correction factor for  $i < N$  and deterministic processing times are

$$\Delta_i^P = (AP_i + \sum_{k=i+1}^{N-1} BP_k) \quad (4.13)$$

$$AP_i = (-0.01707ACR_i^2 + 0.38048ACR_i - 0.15472)(0.5 - CP_i)\sigma D_i \quad (4.14)$$

$$BP_i = (-0.00965ACR_i^2 + 0.13324ACR_i - 0.03230)(0.5 - CP_i)\sigma D_i \quad (4.15)$$

The subscript  $k$  does not appear inside the summation in equation (4.13) because a constant value is being added  $(N-i-1)$  times. However, this form has been used for



consistency with the form used in Section 4.5.4. These equations can be used to determine the necessary correction factors due to a difference in cost proportion at the station being considered.

#### **4.5.4 Correction for Random Processing Times**

The correction factors developed so far assume that processing times are deterministic, so that any effect of decisions on finish time variances is directly transmitted downstream. However, with random processing times, the effect at downstream station  $k$  is damped according to the relation

$$VF_k = \frac{\sigma S_k}{\sigma F_k} = \frac{\sigma S_k}{\sqrt{\sigma S_k^2 + \sigma P_k^2}} \quad (4.16)$$

where  $VF_k$  is the damping or adjustment factor required at the station being considered, due to downstream station  $k$ . Since the effect continues down the line, the net adjustment factor is the product of all such factors from station  $i$  to station  $k$ , and is given by

$$NVF_k = \prod_{l=i}^k VF_l = \prod_{l=i}^k \frac{\sigma S_l}{\sqrt{\sigma S_l^2 + \sigma P_l^2}} \quad (4.17)$$

However, during the operation of the algorithm, the start and finish time variances are not known a priori, since they are the outcome of the decisions taken. Therefore, they are estimated using the values obtained from the uncorrected independent sol-

ution. These estimates are used to adjust the correction factors computed by the formulae for each individual additional station in succession. Equations (4.10) and (4.13) are therefore modified to yield the following formulae.

$$\Delta_i^V = (AV_i \cdot VF_i + \sum_{k=i+1}^{N-1} BV_i \cdot NVF_k) \quad (4.18)$$

$$\Delta_i^P = (AP_i \cdot VF_i + \sum_{k=i+1}^{N-1} BP_i \cdot NVF_k) \quad (4.19)$$

$$\mu D_i^{CI} = \mu D_i^I + \Delta_i^V + \Delta_i^P \quad (4.20)$$

The application of these equations is done by applying the ISM approach in the same way as before, except that at each station the decisions are corrected using the above equations.

#### **4.5.5 Development of Hybrid Solution Technique**

The approximation eventually adopted is a combination of the Corrected Independent Solution Method (CISM) and direct optimization, using some insights gained from the solutions from the Rolling Horizon Method (ROH). The equations developed in the previous section are implemented in the CISM as follows.

1. At station  $i$ , determine the independent optimal policy  $\mu D_i^I$ , given the input to the station  $\mu F_{i-1}$ ,  $\sigma F_{i-1}^2$  and  $\sigma D_i^2$ , and the associated cost parameters  $CS_i$  and  $CE_i$ .

2. Determine the values of  $DS_i$ ,  $CP_i$ , and  $ACR_i$  for the station and compute the necessary correction factors using the appropriate equations. Obtain the corrected decision  $\mu D_i^{cl} = \mu D_i + \Delta_i^y + \Delta_i^p$ .
3. Given the corrected policy for station  $i$ , determine the mean  $\mu S_i$  and the variance  $\sigma S_i^2$  of operation start time, using the formulae for the moments of the maximum.
4. Determine the mean  $\mu F_i$  and the variance  $\sigma F_i^2$  of operation finish time using the means and variances of operation start times and the processing times at station  $i$ . These finish times become the input to the next station.
5. Continue steps 1-4 until decisions are determined for all stations.

The results from the CISM showed that the approximation improves the ISM solution significantly, yielding near-optimal solutions in most problems. However, it leaves errors of over 5% in problems with an extremely large change in cost ratio at the end. This situation is encountered during the heuristic solution of the multiple job problem, when the finished goods and tardiness costs for the entire batch are applied to the last station. The difference is more serious when processing times are random, leading to high operation finish time variances, suggesting that the corrections are not adequate in some cases.

To understand this situation further, the results were compared with those from the ROH method using a horizon of 3 stations. Though generally the CISM errors were significantly lower, the ROH produced better results in a few situations with high tail-end costs. The exact values obtained in these cases showed that though the CISM values at the early stations were consistently closer to the optimal values, the tail-end performance of the ROH was superior since it considers more stations simultaneously, yielding better overall results. The performance of the CISM remained

satisfactory if these cost changes occurred mid-way. In general, however, major changes in the costs are expected to occur only at the end, since that is where the finished goods holding costs and tardiness costs are imposed.

Since the costs are assumed to be increasing down the line, the importance of decisions at later stations also increases. Also, if the cost ratios at the end are large, decisions at preceding stations need to be taken more accurately. The performance of the CISM approach can thus be improved further by solving a larger sub-problem at the end.

These errors in the CISM approach arise mainly out of two approximations used in implementing the algorithm. The first is due to the approximation used to determine the intercept and the slope of the correction factors for a particular value of cost ratio. The second is due to the adjustment of downstream cost ratios by the formula  $ACR_i = \max\left(\frac{CS_{i+k} + CE_{i+k}}{CS_i + CE_i}\right)^{1/k}$ . These errors become more critical as the cost ratios increase. Consequently, the length of the final sub-problem is determined depending on the values of the cost ratios.

Simple decision rules have been evolved by experimentation to determine this final length. If the final cost ratio ( $CR_{N-1}$ ) exceeds a certain value, the length of the final sub-problem is incremented by an amount depending on the value of the cost ratio. Likewise, if the adjusted cost ratio ( $ACR_i$ ) due to the final stations exceeds the original ratio ( $CR_i$ ) by more than a certain factor, the final sub-problem is enlarged to include that station. In the problems tested, values of 3 and 4 respectively were used for the above cut-off levels to determine the final sub-problem length. This resulted in final sub-problem lengths of about 20-30% of the total number of stations. This adjustment removed most of the errors remaining after the application of CISM.

The new procedure, termed the Hybrid Method (HYM), operates as follows.

1. Determine the length (NF) of the final sub-problem using the decision rules based on cost ratio values,  $CR_i$  and  $ACR_i$ , along the line. Carry out steps 2-5 for the first (N-NF) stations.
2. At station  $i$ , determine the independent optimal policy  $\mu D_i^I$ , given the input to the station  $\mu F_{i-1}$ ,  $\sigma F_{i-1}^2$  and  $\sigma D_i^2$ , and the associated cost parameters  $CS_i$  and  $CE_i$ .
3. Determine the values of  $DS_i$ ,  $CP_i$  and  $ACR_i$  for the station and compute the necessary correction factors using the appropriate equations. Obtain the corrected decision  $\mu D_i^{HY} = \mu D_i^I + \Delta Y + \Delta P$ .
4. Given the corrected policy for station  $i$ , determine the mean  $\mu S_i$  and the variance  $\sigma S_i^2$  of operation start time, using the formulae for the moments of the maximum.
5. Determine the mean  $\mu F_i$  and the variance  $\sigma F_i^2$  of operation finish time using the means and variances of operation start times and the processing times at station  $i$ . These finish times become the input to the next station.
6. Determine the  $\mu D_i^{HY}$  values for the last NF stations simultaneously, using NLP, given the parameters for these stations and the inputs from station (N-NF).

This approach combines the approximation factors developed in the previous section with limited direct optimization. It improves the overall performance of the algorithm, at very little additional effort, by solving a larger problem at the end. The performance of the methods is compared in the next section.

## **4.6 Implementation of the Approximation Method and**

### **Results**

The NLP, ISM, Corrected ISM (CISM), and Hybrid (HYM) methods have been applied to the problems that have been used to compare different methods earlier. These problems are distinct from the ones used in the experiments to develop the approximations, and provide different combinations of differences in variances, cost proportions and ratios. The results of these runs are shown in Table 7 for 3 and 10 station problems. The total costs and the relative errors are shown to demonstrate the original error in the ISM and the amount corrected by the HYM method. The figures in parentheses show the error remaining after application of the CISM method.

The results show clearly the significant reduction in the errors by application of the HYM method. In the 3 station problems the largest remaining error is 0.40%, compared to 13.21% which was the largest original error. All the errors are within 1%, and the approximation did not worsen the performance in any case. In the 10 station problems, the largest remaining error is 1.62%, down from 7.13% without correction, while the previously largest error of 13.79% is reduced to 0.11%.

Table 8 shows the results for some additional test problems, with and without random processing times, and high cost ratios down the line, to demonstrate the effectiveness of the variance and cost adjustments. These problems are ones that are solved during the application of the heuristic methods in the multiple job problem. In these problems, the CISM approach leaves some significant errors, but the HYM

**Table 7. Comparison of Results from Approximate Solution Method**

3 Stations				10 Stations			
No.	Method	TC	% Error	No.	Method	TC	%Error
1	NLP	11.806	0.00 0.04 (0.01)	1	NLP	51.598	0.06 0.19 (0.07)
	HYM	11.806			HYM	51.632	
	ISM	11.811			ISM	51.698	
2	NLP	14.224	0.00 0.06 (0.05)	2	NLP	56.722	0.05 0.24 (0.06)
	HYM	14.224			HYM	56.750	
	ISM	14.232			ISM	56.858	
3	NLP	9.414	0.00 0.02 (0.00)	3	NLP	44.181	0.10 0.12 (0.11)
	HYM	9.414			HYM	44.227	
	ISM	9.416			ISM	44.235	
4	NLP	15.526	0.18 8.31 (0.18)	4	NLP	56.119	0.05 7.17 (0.05)
	HYM	15.553			HYM	56.144	
	ISM	16.815			ISM	60.143	
5	NLP	8.554	0.40 13.21 (0.50)	5	NLP	45.353	1.62 7.13 (1.64)
	HYM	8.589			HYM	46.086	
	ISM	9.684			ISM	48.586	
6	NLP	19.735	0.00 4.45(0.03)	6	NLP	61.423	0.17 9.59 (0.17)
	HYM	19.735			HYM	61.527	
	ISM	20.614			ISM	67.316	
7	NLP	34.327	0.28 3.33(0.35)	7	NLP	80.810	0.21 6.07 (0.23)
	HYM	34.423			HYM	80.983	
	ISM	35.469			ISM	85.712	
8	NLP	44.136	0.04 0.09(0.07)	8	NLP	86.327	0.54 5.64 (0.56)
	HYM	44.155			HYM	86.792	
	ISM	44.175			ISM	91.200	
9	NLP	14.810	0.01 0.69(0.02)	9	NLP	62.723	0.11 13.79 (0.11)
	HYM	14.812			HYM	62.791	
	ISM	14.912			ISM	71.373	
10	NLP	21.091	0.13 12.06(0.14)				
	HYM	21.118					
	ISM	23.634					

(.) : Error from CISM

approach improves them considerably by solving a larger problem at the end. These results confirm the effectiveness of the approximation over a wide range of problems.

The significant improvements verify the effectiveness of the idea of using correction factors in conjunction with the ISM method to generate good solutions. Although the individual  $\mu D_i$ 's after correction are not exactly at the optimum values, that is to be expected in an approximation technique. More importantly, the corrections are always in the right direction, and the resulting point is sufficiently close to the optimum to be at least in the region of indifference. The magnitude of resulting errors is quite satisfactory for an approximation, especially since only the most dominant factors have been considered for the sake of simplicity and understanding.

The equations developed in this section are an attempt at empirically approximating the errors in the ISM method, with respect to the optimal solution determined by NLP. They are obviously dependent upon the quality of the NLP solution, and are limited in their look ahead capability. However, they enable quick solution to the single job problem using the characteristics of the problem. The results of the tests carried out have confirmed the ability to obtain good solutions using the approximations with very little computational effort. The approximations have been refined to incorporate reasonable ranges in parameter values, some look-ahead features for downstream changes, and nonlinear relationships, to maintain their effectiveness over a wide range of problem variations. Another important benefit of this study and the resulting relations is the insight provided of the behavior of the process and the optimal policy due to changes in problem parameters. The consideration of other distributions in the solution of the single job problem is addressed in the next section.



**Table 8. Comparison of Results from Approximation**

No.	Stns.	P.T.	Method	TC	% Error
1	10	R	NLP	332.05	0.43 46.18 (5.78)
			HYM	333.48	
			ISM	485.38	
2	10	D	NLP	101.81	0.46 8.28 (0.53)
			HYM	102.28	
			ISM	110.25	
3	14	R	NLP	476.36	0.32 61.14 (5.99)
			HYM	477.91	
			ISM	767.61	
4	14	D	NLP	158.52	0.18 9.51 (0.26)
			HYM	158.80	
			ISM	173.59	
5	15	R	NLP	584.53	0.46 90.8 (8.50)
			HYM	587.23	
			ISM	1115.3	
6	15	D	NLP	167.59	0.05 7.95 (0.07)
			HYM	167.68	
			ISM	180.92	
7	20	R	NLP	960.47	1.29 90.93 (9.01)
			HYM	972.88	
			ISM	1833.83	
8	20	D	NLP	264.96	0.16 7.51 (0.17)
			HYM	265.39	
			ISM	284.87	
9	20	R	NLP	1136.52	0.95 82.32 (9.27)
			HYM	1147.33	
			ISM	2072.09	
10	20	D	NLP	431.35	2.60 7.45 (3.39)
			HYM	442.55	
			ISM	463.51	

(.) : Errors from CISM

## **4.7 Solution of Multi-station Problem for Non-normal**

### **Distributions**

The prescriptive methods were then applied to the lognormal and gamma distributions. This was done with a view to understand the differences between the decisions and costs with different distributions, and therefore to assess the importance of the choice of distribution. Also, it needs to be confirmed if the attractiveness of the ISM approach continues to hold, so that similar approximations can be developed for the other distributions as well.

#### **4.7.1 Optimal and Independent Solutions**

In order to implement the optimization methods, the objective function defined in equation (4.9) needs to be evaluated. This is done by applying the defining equations (3.1) and (3.2) and using the formulae for the moments of the maximum in equations (3.14-3.15) for the lognormal distribution, and (3.16-3.18) for the gamma distribution. The application of the methods proceeds in exactly the same way as for the normal distribution, except that now even the single stage solution required for the ISM method has to be determined by the single variable search.

The same sets of problems as run for the normal distribution were also run for the lognormal and gamma distributions using the NLP and ISM methods to compare the effect of distributions on costs and decisions, and assess the performance of the

ISM method. Tables 9 and 10 show the results of these runs for the sample 3 and 10 station problems.

The results of the test problems for different distributions show clearly that the optimal policy costs for the lognormal and gamma distributions are significantly higher than those for the normal, ranging between 10-15% for 3 station problems. This is an outcome of the skewed nature of the distributions, which causes the makespan values to be larger. The gamma distribution generally results in somewhat lower estimates of cost than the lognormal.

Also of interest is the effect of the distribution on the optimal decisions, since that is one of the reasons for considering different distributions. As expected, the most significant differences occur when the variances are large, as in problems 4,6,8 and 10. The decisions are, however, quite similar when the variances are small, since then the shapes of the distributions are quite similar. In general, the  $\mu D$  values for the lognormal distribution are the most different from the normal, and for the gamma distribution a little less so.

It is also necessary to examine the potential errors arising out of the choice of distribution. In other words, a measure is needed for the impact on cost if decisions were taken using the normal distribution, when in fact the actual distribution was one of the other two. To measure this potential error, the decisions from the normal distribution were measured in terms of their expected cost using the cost calculation routines of the other distributions. The difference between this cost and the actual optimal cost for that distribution provides a measure of the significance of the choice of distribution.

**Table 9. Comparison of Normal, Lognormal and Gamma Distributions for 3 Station Problems (using NLP method)**

Initial Problem:

$$\mu F_0 = 15, \sigma F_0 = \sigma D_1 = \sigma D_2 = 2, \sigma D_3 = 0.001, \mu P_i = 5, \sigma P_i = 0$$

$$CE_1 = CE_2 = 1, CE_3 = 8, CS_1 = 1, CS_2 = 2.5, CS_3 = 4$$

No.	Variation	Distr	TC	$\mu D_1$	$\mu D_2$	$\mu D_3$
1	--	N	11.806	15.00	19.78	27.26
		L	12.695	15.03	19.81	27.34
		G	12.520	15.02	19.80	27.31
2	$CE_3 = 20$	N	14.224	15.00	19.81	28.07
		L	15.572	15.06	19.90	28.29
		G	15.339	15.03	19.83	28.22
3	$CE_3 = 3$	N	9.414	15.00	19.75	26.35
		L	9.926	15.03	19.77	26.32
		G	9.808	15.02	19.75	26.31
4	$\sigma D_1 = 4$	N	15.526	12.51	19.88	27.35
		L	17.725	12.64	20.21	27.73
		G	17.040	13.27	20.00	27.72
5	$\sigma D_1 = 0.5$	N	8.554	16.88	19.90	27.59
		L	9.250	16.96	19.97	27.74
		G	9.250	16.96	19.97	27.74
6	$\sigma F_1 = 4, \sigma D_1 = 4$	N	19.735	15.00	22.13	29.72
		L	22.600	15.46	22.56	30.27
		G	21.996	15.33	22.38	30.08
7	$CS_1 = 10, CS_2 = 11.5,$ $CS_3 = 13, CE_3 = 26$	N	34.327	12.49	17.50	26.20
		L	36.615	12.14	17.22	26.14
		G	36.098	12.49	17.25	26.17
8	$CS_1 = 10, CS_2 = 11.5,$ $CS_3 = 13, CE_3 = 26,$ $\sigma D_1 = 4, \sigma D_2 = 4$	N	44.136	8.73	13.67	26.15
		L	54.277	8.40	12.96	26.41
		G	51.231	9.81	14.42	26.52
9	$CE_1 = 2, CS_2 = 3.5$ $CS_3 = 5, CE_3 = 10$	N	14.810	15.76	19.86	27.60
		L	15.871	15.92	19.92	27.74
		G	15.658	15.93	19.89	27.72
10	$CE_1 = 2, CS_2 = 3.5,$ $CS_3 = 5, CE_3 = 10,$ $\sigma D_1 = 4$	N	21.091	13.88	20.08	27.85
		L	23.757	14.26	20.35	28.25
		G	22.872	13.92	19.90	27.91

**Table 10. Performance of ISM Method for Lognormal and Gamma Distributions for 10 Station Problems**

Initial Problem:

$$\mu F_0 = 15, \sigma F_0 = \sigma D_i = 2, i = 1, \dots, 9, \mu P_i = 5, i = 1, \dots, 9, \mu P_{10} = 0, \sigma P_i = 0, i = 1, \dots, 10$$

$$CE_i = 1, i = 1, \dots, 9, CE_{10} = 30, CS_1 = 1, CS_2 = 2.5, CS_3 = 4, CS_4 = 5.5, CS_5 = 7,$$

$$CS_6 = 8.5, CS_7 = 10, CS_8 = 11.5, CS_9 = 13, CS_{10} = 15$$

Pr. No.	Variation	Method	Lognormal		Gamma	
			TC	% Error	TC	% Error
1	--	NLP	55.187		54.925	
		ISM	55.594	0.56	55.144	0.40
2	$CE_{10} = 60$	NLP	60.996		60.706	
		ISM	61.413	0.68	61.005	0.49
3	$CE_{10} = 10$	NLP	46.855		46.626	
		ISM	47.009	0.33	46.739	0.24
4	$\sigma D_1 = 4$	NLP	61.578		60.464	
		ISM	68.307	10.93	66.765	10.42
5	$\sigma D_1 = 0.5$	NLP	48.322		48.322	
		ISM	52.025	7.66	52.025	7.66
6	$\sigma F_1 = 4, \sigma D_1 = 4$	NLP	67.883		66.992	
		ISM	78.945	16.30	76.190	13.73
7	$CS_1 = 10, CS_2 = 12.5, \dots$ $CS_{10} = 25, CE_{10} = 50$	NLP	86.856		86.144	
		ISM	90.673	4.39	90.832	5.44
8	$CS_1 = 10, CS_2 = 12.5, \dots$ $CS_{10} = 25, CE_{10} = 50$ $\sigma D_1 = 4$	NLP	98.489		95.633	
		ISM	106.23	7.86	105.87	10.71
9	$CE_1 = 2, CS_2 = 3.5, \dots$ $..CS_{10} = 16, CE_{10} = 32$ $\sigma D_1 = 4$	NLP	68.939		67.622	
		ISM	80.342	16.54	78.715	16.41

The largest difference using this measure for 3 station problems was 5% for the gamma distribution in problem 8, while for problems 4 and 10, the error was about 1%. Even though the optimal decisions for the lognormal were the most different, the cost errors were within 1% in all the problems. However, when the 10 station problems were run the errors for the lognormal became more dominant. On problems 4, 6, 8 and 9, all of which had a high part delivery time variance, the errors were over 1%, with problem 4 recording the highest error of 1.7% for the lognormal, and problem 8 the highest error of 2.4% for the gamma.

This appears to suggest that for low variances, the errors arising out of the use of a wrong distribution may be small, but increase as the variability increases. The choice of how much error is significant is dependent on the absolute values of the costs involved. However, regardless of the effect on the decisions, the use of the normal distribution always gives a considerably optimistic estimate of the total cost, so if cost estimation errors are a factor then the choice of distribution can be quite significant.

The values of the errors resulting from the ISM method for the lognormal and gamma distributions are of the same order of magnitude as in the normal case. It is therefore expected that correction factors similar to those for the normal distribution can be developed for these distributions as well. The determination of these factors is discussed in the next section.

## **4.7.2 Approximate Solution Technique for Non-normal Distributions**

The extension of the approximate solution technique to non-normal distributions follows the same approach as used for the normal distribution. For simplicity, and using the understanding of system behavior generated earlier, the same factors and ranges have been considered. As an initial approximation, the correction equations used for the normal distributions were used for the lognormal and gamma distributions as well. The results using this approximation showed errors of similar magnitude to those observed for the normal distribution, further confirming the viability of the approach. However, the individual decisions are marginally different due to the skewed nature of the new distributions. The new correction factors have been determined in order to eliminate these differences to the extent possible. The equations are listed in Appendix C.

The procedure operates in the same manner, except that the independent station decisions are now determined using single variable search. These are corrected using the necessary equations incorporating variances, cost proportions, cost ratios, and random processing. Table 11 shows the results using the different methods for the lognormal and gamma distributions, using the same problems considered for testing the approximation for the normal distribution. The performance is of the same quality as for the normal distribution. As before, the HYM technique corrects most of the errors remaining after application of the CISM approach. It is also considerably less time consuming than the NLP approach, increasing its preferability in view of the much greater computational requirements for these distributions due to the complexity in their formulae.

**Table 11. Performance of HYM Method for Lognormal and Gamma Distributions for Test Problems**

Pr. No.	Stns.	P.T.	Method	Lognormal		Gamma	
				TC	% Error	TC	% Error
1	10	R	NLP	346.16		345.83	
			HYM	348.61	0.71	347.49	0.48
			ISM	508.63	46.9 (6.25)	508.68	47.1 (3.81)
2	10	D	NLP	108.79		107.76	
			HYM	109.24	0.42	109.21	1.35
			ISM	116.39	6.99 (0.89)	116.29	7.92 (1.69)
3	14	R	NLP	488.47		488.10	
			HYM	490.33	0.38	490.61	0.52
			ISM	806.42	65.09 (6.23)	805.73	65.07 (6.15)
4	14	D	NLP	168.91		168.23	
			HYM	168.93	0.01	169.20	0.58
			ISM	184.28	9.10 (0.90)	184.98	9.96 (0.98)
5	15	R	NLP	600.31		599.29	
			HYM	602.40	0.35	606.02	1.02
			ISM	1190.8	98.4 (6.42)	1191.1	98.8 (8.31)
6	15	D	NLP	176.48		177.17	
			HYM	176.82	0.19	177.49	0.18
			ISM	191.67	8.04 (0.66)	191.68	8.19 (0.53)
7	20	R	NLP	975.62		981.17	
			HYM	987.99	1.19	1004.6	2.39
			ISM	1975.1	102.4 (8.34)	1977.2	101.5 (9.87)
8	20	D	NLP	276.56		277.20	
			HYM	281.51	1.79	279.51	0.84
			ISM	305.96	10.63 (2.91)	304.68	9.91 (2.63)
9	20	R	NLP	1157.0		1161.3	
			HYM	1165.4	0.73	1176.6	1.32
			ISM	2216.3	91.6 (8.54)	2214.7	90.7 (9.44)
10	20	D	NLP	456.75		456.42	
			HYM	466.18	2.07	458.19	0.39
			ISM	494.33	8.23 (3.60)	491.37	7.66 (1.84)



## **4.8 Summary**

In this chapter the problem of determining optimal part delivery dates in the single job or lot-for-lot assembly line has been addressed. The objective function has been developed for the single station and multi-station scenarios. Different solution approaches have been examined and compared, and it has been found that the nonlinear programming approach consistently yields the best solutions and is general enough to be used for different distributions and random processing times. Even the independent solution approach out-performs the DP approach developed in earlier research for normally distributed part delivery times and deterministic processing times.

An approximation technique based on this approach has been developed for the single job problem. The approximate solution method is shown to give results comparable in quality to the nonlinear programming method with very little computational effort. The effect of introducing random processing times has been examined. The differences between the distributions considered have been examined from the point of view of their impact on decisions and costs. The approximate solution method has been generalized to include random processing times and the different distributions considered, and retains its effectiveness over these generalizations.

## **5.0 Solution of Multiple Job Problem**

### **5.1 Introduction**

This chapter describes the development of the objective function and solution techniques for the multiple job problem. In this problem, a set of identical jobs is assembled in sequence, flowing through the assembly stations in the same order, and delivered to the customer as a batch. The parts required are delivered separately for each job and each station. The system operates according to the scenario described in Chapter 1 and shown in Figure 1.

The first set of decision variables in this problem are the mean delivery dates for parts at each assembly station for each job. An additional set of decision variables is introduced in the model to determine the optimal mean assembly job launch times into the system. This is necessary because having all the jobs on hand at the start will incur unnecessary waiting costs, since assembly start times are limited by availability of stations. As in the single job case, the delivery date for the finished batch of assemblies is also a decision variable. If  $N$  is the number of stations and  $J$  is the number of jobs to be produced, then the total number of decision variables is  $J \times (N + 1)$ .

In this problem, the availability of stations from previous jobs must also be considered in determining operation start times. Although the part delivery times are still assumed to be independent, the different operation start and finish times are correlated. Therefore, for analytical evaluation of operation start and finish times and obtaining measures of system performance, it is necessary to compute correlations between the operation times in the network.

## **5.2 Precedence Network Representation of Multiple Job**

### **Assembly**

The descriptive relationships used in deriving the objective function can be better visualized with the aid of a precedence network. This network is a generalization of the series structure in Section 4.3. Each node in the network represents an operation being performed on a particular job at a particular station. The configuration of the network for the unlimited and limited buffer situations is described next.

#### **5.2.1 Assembly Systems with Unlimited Buffers**

An assembly system with unlimited buffers and part input can be represented as a precedence network, with node  $(i, j)$  representing an operation performed at station  $i$  on job  $j$ . An illustrative network for a 5 station, 5 job network is shown in Figure 12. Each column in the network represents a different assembly station, and is identified

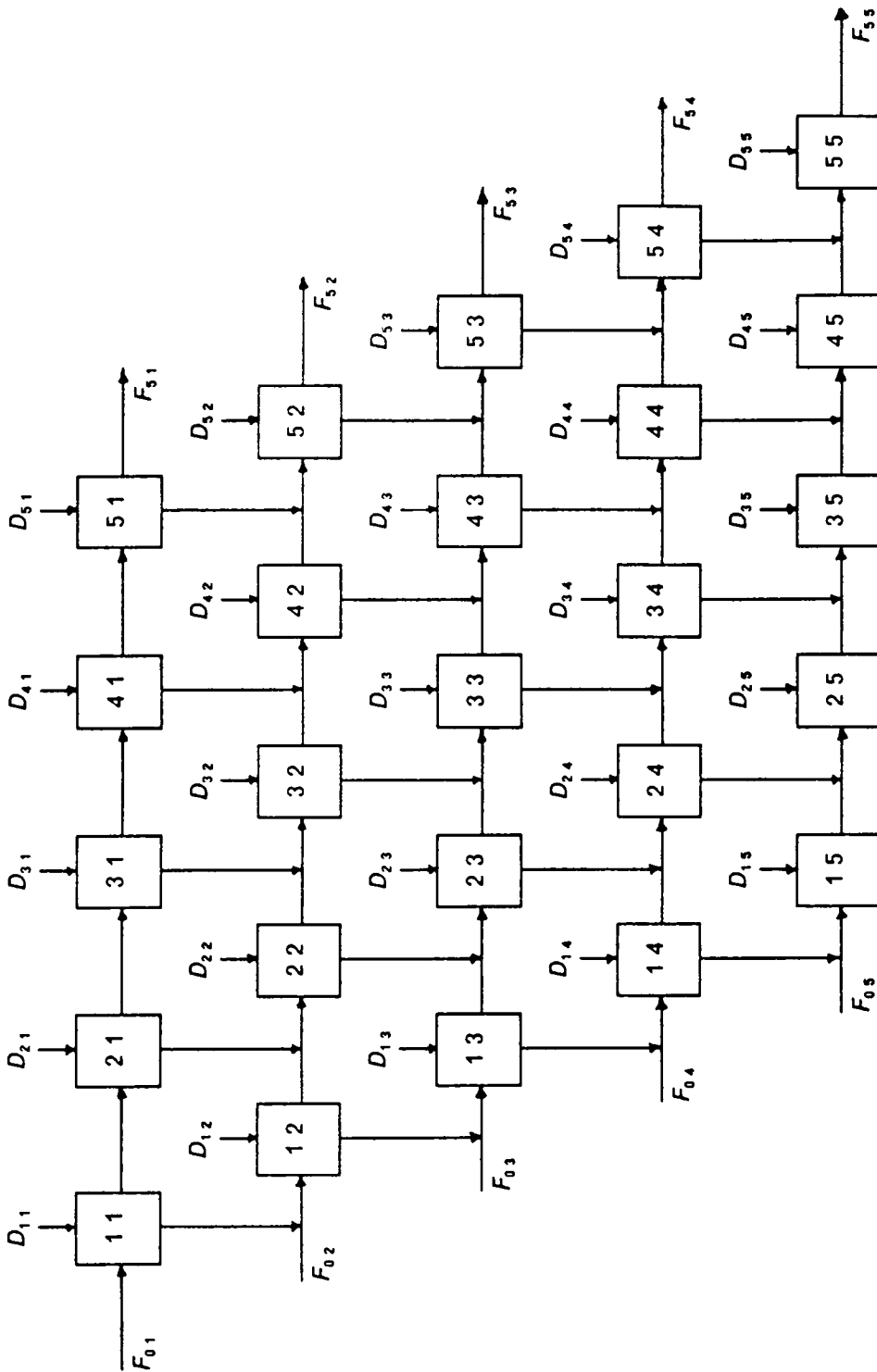


Figure 12. Precedence Network Representation of a Multiple Job Assembly System with Unlimited Buffer Capacities

by the first number in the node label. The restriction on the start time of operation (i,j) due to the availability of station i from job (j-1) is represented by the vertical lines in the network. Each row in the network represents a different job, identified by the second number in the node label. The availability of job j from station (i-1) is represented by the horizontal lines in the network. The availability of the part required is shown by the arrow  $D_{i,j}$  directly incident upon the node. This network represents the basic descriptive relationship for the start time of the assembly operation.

$$S_{i,j} = \max\{F_{i,j-1}, F_{i-1,j}, D_{i,j}\}$$

$$F_{i,j} = S_{i,j} + P_{i,j}$$

### **5.2.2 Assembly Systems with Limited Buffers**

Until now, it has been assumed that the only restrictions on operation start time are the part availability, completion of job j at station (i-1), and completion of job (j-1) at station i. If the buffer capacities between the stations are limited, even the completion of jobs before job (j-1) at stations after i can affect the start time of operation (i,j) due to the possibility of blocking of jobs at stations. This can occur if an operation is complete, but the job cannot leave the station because its downstream station and buffers are full. For example, if there are no buffers between stations 1 and 2, job 2 cannot leave station 1 unless job 1 is also completed at station 2. Consequently, job 3 cannot move into station 1 unless both operations (1,2) and (2,1) are completed. Operation (2,1) thus becomes an additional predecessor for operation (1,3). Similar relationships can be derived for other buffer capacities as well.

In general, the start time of an operation may be limited by any or all downstream stations in addition to those in the basic model. An additional parameter  $B(i)$  is used to denote the buffer capacity before station  $i$ . The start of operation  $(i,j)$  is then limited by the completion time at station  $(i + 1)$  of job  $[j-2-B(i + 1)]$ . This ensures that space is available in the buffer before station  $(i + 1)$  for job  $(j-1)$  to move into after it completes processing at station  $i$ . Likewise, it may also be limited by the completion time at station  $(i + 2)$  of job  $[j-3-B(i + 1)-B(i + 2)]$ . These restrictions continue until the last station, or until there are no more jobs ahead of job  $j$  that can cause any additional restrictions. If the buffer capacities are unlimited, these dependencies do not add any precedence restrictions on the start time.

The modified descriptive relationship for assembly start time incorporating this feature was developed by Wilhelm (1986a) in his study of flow shops with limited storage, and is shown below. The queueing spaces for parts, initial mainframes, and finished goods are still assumed to be of unlimited capacity.

$$S_{i,j} = \max[F_{N,j-N(i)}, \dots, \max\{F_{i+1,j-2-B(i+1)}, \max(F_{i,j-1}, F_{i-1,j}, D_{i,j})\}] \quad (5.1)$$

where  $N(i) = \sum_{k=i+1}^N B(k) + (N - i + 1)$  is the net effect at operation  $(i,j)$  of buffer capacities up to station  $N$ .

The buffer capacities between stations can therefore be converted into the equivalent additional precedence restrictions due to subsequent jobs, which can be shown on the network. An example network for a 5 station, 5 job network with buffer capacities 1,2,1, and 2 before stations 2, 3, 4 and 5 respectively is shown in Figure 13. The broken lines represent the additional precedence restrictions due to limited buffer capacities.

Buffer Capacities: (1-2-1-2)

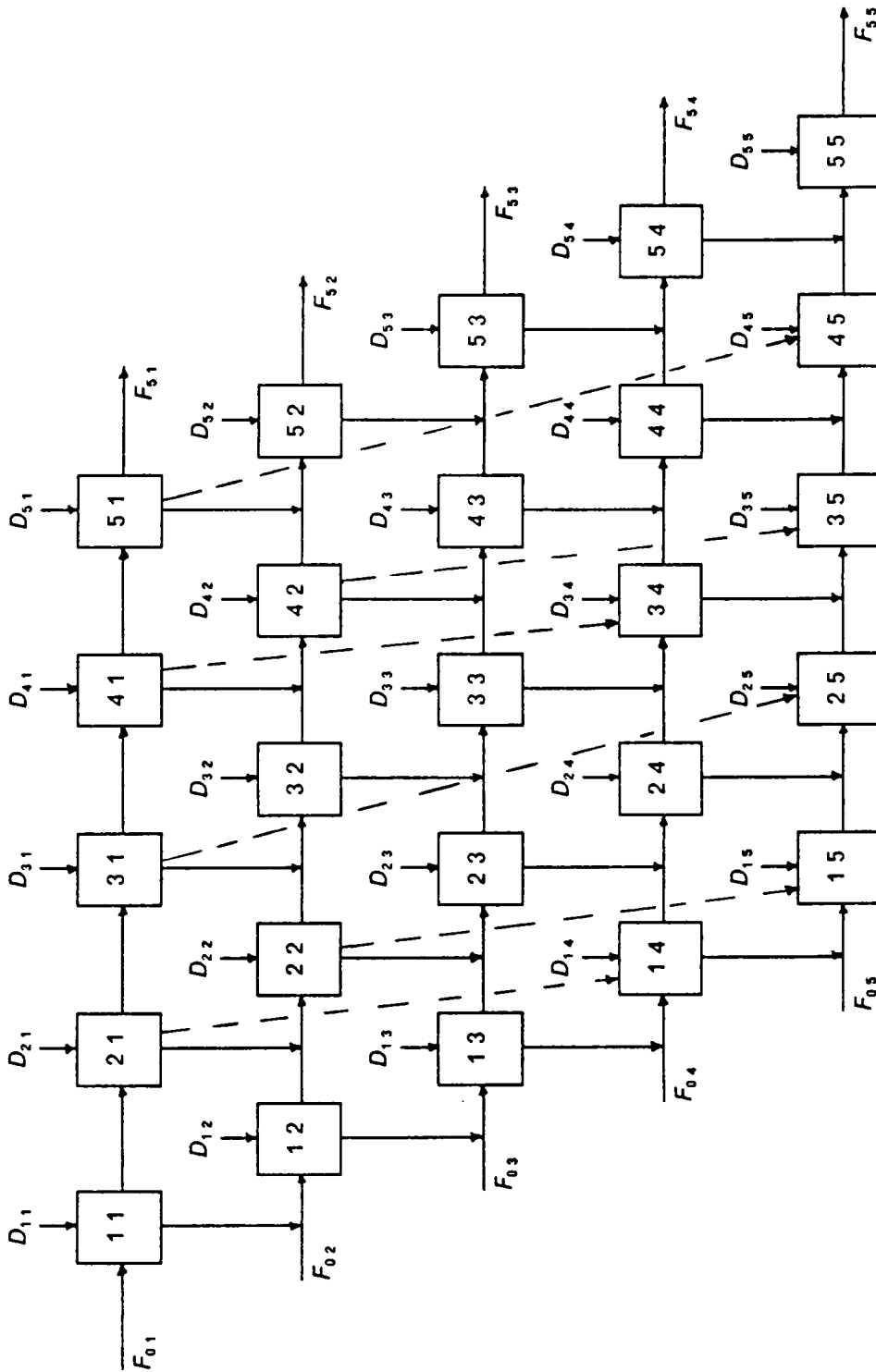


Figure 13. Precedence Network Representation of a Multiple Job Assembly System with Limited Buffer Capacities

The computations of expectations and correlations now must consider each of the inputs to a particular node, and the maximum operator must be applied as many times as there are predecessors of that node. This can be done in a nested fashion from right to left, updating the correlations each time the maximum operator is applied.

### 5.2.3 Objective Function for Multiple Job Assembly

The objective function for the multiple job problem for either of the above system configurations can be determined by applying the descriptive equations (3.1) and (3.2) recursively and summing the part and subassembly waiting costs over all stations for each job. Additional costs introduced in this model include the makespan cost, the finished goods holding cost, and the tardiness penalty for time taken beyond the customer delivery date. The terms in the objective function are as defined in equations (3.3) to (3.9).

$$\begin{aligned}
 E(TC) &= \sum_{j=1}^J \sum_{i=1}^N [CS_i E(WS_i) + CE_i E(WP_i)] + E(FT)CM + \sum_{j=1}^J [E(WF_j)CF + E(TF)CT] \\
 &= \sum_{j=1}^J \sum_{i=1}^N [CS_i(\mu S_{i,j} - \mu F_{i-1,j}) + CE_i(\mu S_{i,j} - \mu D_{i,j})] + [\mu F_{N,j} - \mu F_{0,1}]CM \\
 &\quad + \sum_{j=1}^J [[E(\max\{T, F_{N,j}\}) - \mu F_{N,j}]CF + [E(\max\{T, F_{N,j}\}) - T]CT] \tag{5.2}
 \end{aligned}$$



If the parts required at a particular station for different jobs are identical, then strictly speaking the actual delivery time of a part for the first job is the minimum of the  $J$  actual realizations of the delivery times of the part. Likewise, actual delivery time for the second job is the second smallest, and so on. Strictly speaking, new relationships are required to determine the moments of actual delivery times as ordered samples from distributions with non-identical means. In view of the difficulty of incorporating this in the model, it is therefore assumed that the parts required for different jobs at a particular station follow the same delivery time distribution, but can be taken to be distinct. If the parts are not taken to be distinct, an assumption must be made that the part with the smaller mean time will arrive first. This is similar to the assumption made by Wilhelm (1986b), using the principle of stochastic dominance, in his study of workcells with parallel machines. The assumption is reasonable if the interval between mean part delivery times at a station for successive jobs is large compared to the standard deviation of delivery time.

#### **5.2.4 Evaluation of the Objective Function**

The evaluation of the objective function can be done either by simulation or by applying the analytical relations above. The simulation approach does not make any assumptions about the distributions of the maximum and the sum since the actual realizations are considered. However, for use in conjunction with a function evaluation based optimization method, an exact value is required, whereas simulation can result in different values for the same set of decisions due to randomness. Therefore, for direct optimization it is necessary to use the analytical relations.

Since the computational requirements in terms of function evaluations are likely to be fairly heavy in the multiple job problem, especially for large problems, before attempting to implement a direct optimization technique attempts must be made to reduce this effort. The main computational effort in this problem arises due to the need to determine pairwise correlations between finish times.

Determining the correlations one operation at a time for each job results in the computation of a number of correlations never used in the calculations. In order to reduce the number of correlations computed, a scheme based on the structure of the network has been developed. The idea behind the scheme is to carry out the computations according to precedence levels in the processing network, which are similar to product levels used in product structure diagrams. A level runs diagonally across the nodes as shown in Figure 14, and contains those nodes all of whose predecessors have been included at previous levels. It thus has the feature that a node cannot be a predecessor of any other node at the same level in the network. To illustrate, in a  $5 \times 5$  network of the type in Figures 12-14, the first level contains only node (1,1), the second contains (2,1) and (1,2), the third contains (3,1), (2,2), and (1,3), and so on. The total number of such levels in the network is  $(N + J - 1)$ .

Since the predecessors of a node must be at lower levels in the network, the total number of levels in the network is also the greatest number of nodes any path can traverse through the network. The nodes at a particular level in the network can be determined using the logic of the network. To simplify the record keeping, all nodes are identified by a single identifier rather than by a station-job pair. For node (i,j) this identifier is  $(N(j-1) + i)$ , and the station-job pair can be easily calculated from this identifier whenever required.

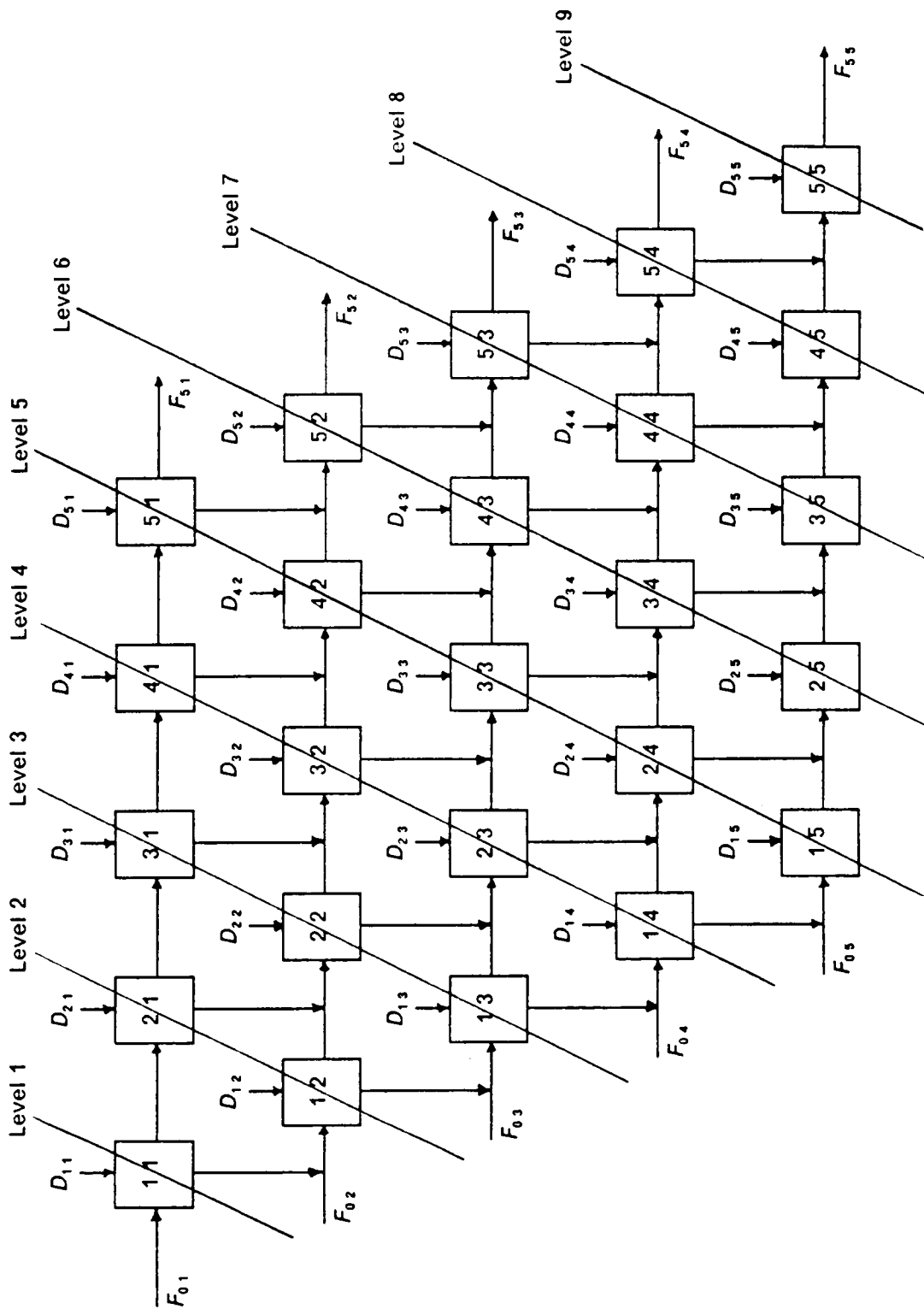


Figure 14. Precedence Levels in a Multiple Job Assembly System

The computations are carried out one level at a time. When a particular level is being considered, the expectations and correlations for all nodes at previous levels have been computed. Since predecessors can only be at previous levels, all the necessary expectations and correlations are therefore already available and the necessary computations can be carried out. For each node, the correlations are computed with the necessary nodes at that level and at as many earlier levels as desired. These correlations have to be recalculated each time the Maximum operator is applied. Also, if processing times are random, the correlations of finish times have to be calculated as per equation (3.20).

A feature of this scheme is that all predecessors for the limiting cases of unlimited buffers and 0 buffers are at the immediately earlier level. Therefore, even if correlations are computed with only one preceding level, no needed correlations would have been ignored. For limited, non-zero buffers, predecessors can come about from any preceding level, depending on the buffer capacities between specific stations. However, actual results from the computations showed that the correlations more than one level apart are in fact very small (of the order of  $10^{-3}$ ). The correlations get further damped as per equation (3.20) when random processing times are introduced, so it has been decided to ignore correlations of operations more than one level apart. Note that this does not mean that the precedence related dependencies are ignored, only the correlations have been taken to be zero. This assumption is similar to the scheme used in Wilhelm (1982), that nodes two jobs and two stations away from the one being evaluated can be assumed to be remote and the correlations ignored. In the present scheme, however, no correlations are ignored for the cases of unlimited and zero buffers.

This scheme for computing correlations and expectations has been implemented, and the expected costs are evaluated by substituting these expectations in equation (5.2). To assess the accuracy of the scheme, the total costs derived by this method, for a given set of values of decision variables, have been compared with those obtained by simulation, using 1000 replications, for problems of up to 10 stations  $\times$  10 jobs. In the unlimited and zero buffer cases, where no correlations are ignored, the errors in total cost ranged between 0.5-2%. These errors arise out of the approximation in the analytical procedure of the distribution of the maximum as a normal. In problems with limited, non-zero buffers where some correlations are ignored, the errors were also of the same order of magnitude. However, should further accuracy be desired, additional levels may be added to the scheme. However, in view of the magnitudes of the errors, it did not seem necessary to do so.

## **5.3 Determination of Optimal Solution for Multiple Job**

### **Problem**

#### **5.3.1 Solution Approach**

Considering the encouraging results obtained by applying unconstrained nonlinear programming techniques to the single job problem, it was decided to attempt the approach for the multiple job problem as well. The DP methodology used in Sarin and Das (1987) to solve the single job problem was difficult to generalize to the multiple job problem because of difficulties with the dimensionality of the problem. Con-

sequently, only heuristic solution could be attempted for the multiple job problem. On the other hand, the non-linear programming approach can be generalized readily as long as the objective function can be evaluated for a set of decisions. This can be done only for the normal and lognormal distributions, since formulae incorporating correlations are available for these distributions only.

Of course, it is even more difficult to draw any conclusions about the convexity and unimodality of function in this problem in view of the large solution space. The computational requirements are likely to be considerably higher than the single job problem, even for the same number of variables, because of the need to compute correlations between operation finish times. However, even if suboptimal solutions are obtained, they will provide insights about system behavior that can be used to guide the development of heuristic methods. Additionally, this method provides a good yardstick to compare the solutions with any heuristic solution methods implemented later.

Using the objective function evaluation scheme described above, the unconstrained nonlinear programming method was implemented for the multiple job case with limited and unlimited buffers, as well as with deterministic and random processing times. Different function tolerance levels were used for termination of the algorithm to examine the trade-off between solution quality and computation time. As expected, the computation times were consistently much higher, even with relaxed tolerances, than for the single job problems. A number of small problems were run initially to understand the optimal policy behavior in different scenarios.

Specific factors for which the solutions were examined include:

1. The role of the bottleneck processing station.
2. The role of the critical path of the processing network, which passes along job 1 up to the bottleneck station, continues along the bottleneck station for the remaining jobs, and then through the remaining stations along the last job.
3. The launch intervals between successive jobs and the spacing between the decisions before and after the bottleneck station.
4. The effect of limited buffers on decisions.
5. The effect of random processing times on decisions.

The results for a sample problem under different scenarios are shown in Tables 12 and 13, for deterministic and random processing times respectively, both for unlimited and zero buffers.

### **5.3.2 Analysis of Results**

The observations and insights obtained regarding system behavior from the examination of the optimal solutions for a number of problems under different scenarios are discussed below. Figures 15-17 show comparisons of some indicators of optimal policy behavior in the sample problem under different scenarios. The scenarios are identified on the figures as follows.

1. Deterministic Processing, Unlimited Buffers: D/U
2. Deterministic Processing, Zero Buffers: D/Z
3. Random Processing, Unlimited Buffers: R/U
4. Random Processing, Zero Buffers: R/Z

**Table 12. Comparison of Optimal Solutions for Sample 5 Station, 5 Job Problems**

Mean Operation Times: 5-8-10-9-7, Bottleneck station: 3

First assembly arrival time: 15.0

**a) Deterministic Processing, Unlimited Buffers**

Job	Start Time	Part Delivery Times at Station				
		1	2	3	4	5
1	15.00	12.90	17.95	25.90	41.05	57.70
2	23.65	23.60	27.90	36.20	49.95	64.70
3	33.85	33.30	38.10	46.10	58.65	71.40
4	43.50	43.50	48.25	56.20	67.30	78.35
5	53.35	53.15	58.40	66.10	75.50	84.85

Delivery Date 96.65      Total Cost 1799.3

**b) Deterministic Processing, Zero Buffers**

Job	Start Time	Part Delivery Times at Station				
		1	2	3	4	5
1	15.00	13.40	18.43	25.90	37.50	49.70
2	23.65	23.65	28.40	35.70	47.85	59.60
3	33.80	32.81	38.60	46.10	57.00	69.70
4	43.95	43.00	48.75	56.20	67.80	78.95
5	54.10	53.15	57.85	66.20	75.95	84.95

Delivery Date 96.65      Total Cost 1831.3



**Table 13. Comparison of Optimal Solutions for Sample 5 Station, 5 Job Problems**

Mean Operation Times: 5-8-10-9-7, Standard Deviations 2-2-4-2-2

a) Random Processing, Unlimited Buffers

Job	Start Time	Part Delivery Times at Station				
		1	2	3	4	5
1	15.00	11.60	16.65	24.45	43.45	62.65
2	21.40	21.65	27.10	36.05	53.10	69.85
3	32.60	32.65	38.35	47.30	63.10	77.05
4	44.40	44.35	49.95	58.95	73.25	84.90
5	56.30	56.35	62.00	70.75	83.60	93.45

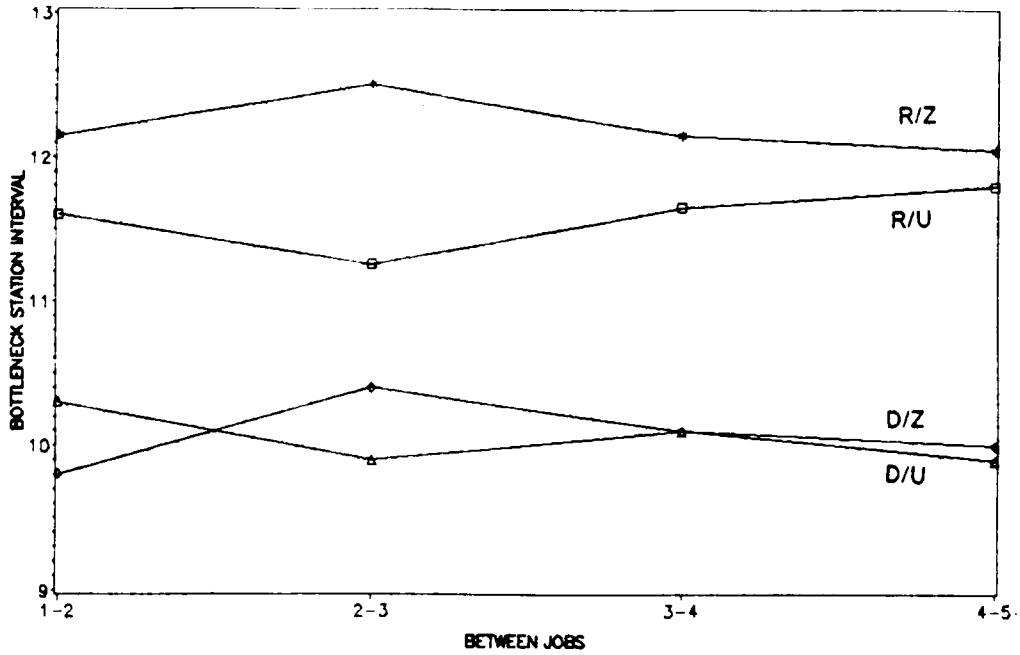
Delivery Date 107.25      Total Cost 2318.5

b) Random Processing, Zero Buffers

Job	Start Time	Part Delivery Times at Station				
		1	2	3	4	5
1	15.00	11.60	16.80	24.80	39.25	53.55
2	22.50	22.55	27.90	36.95	50.90	65.70
3	34.50	35.00	40.35	49.45	63.25	77.60
4	46.95	46.95	52.80	61.60	75.15	87.40
5	59.25	59.20	65.00	73.65	85.95	95.75

Delivery Date 109.50      Total Cost 2386.6

a) Intervals at Bottleneck Station



b) Mainframe Launch Intervals

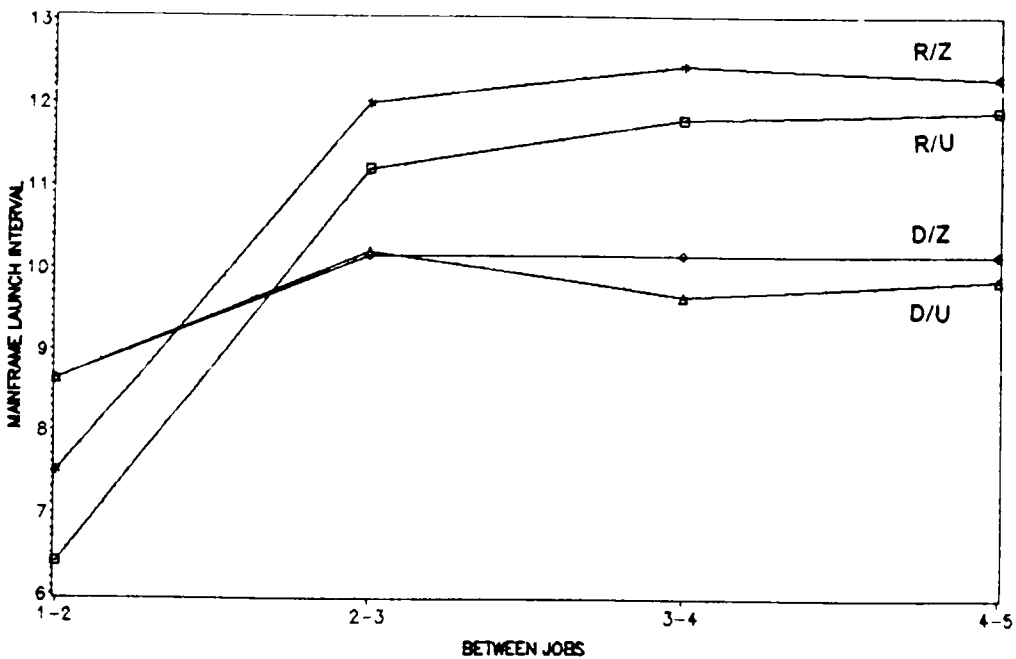


Figure 15. Optimal Intervals Between Successive Jobs

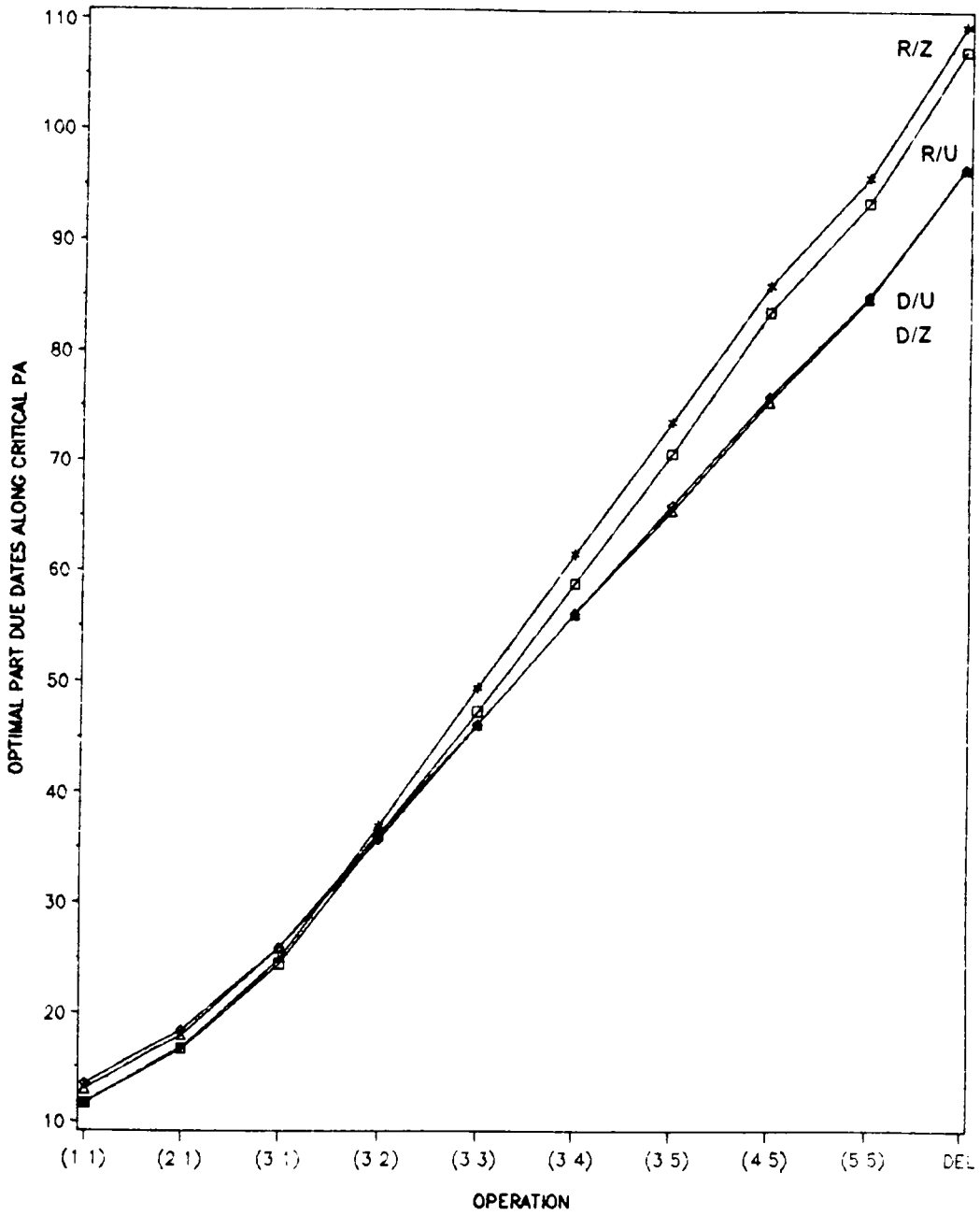


Figure 16. Optimal Part Due Dates Along Critical Path

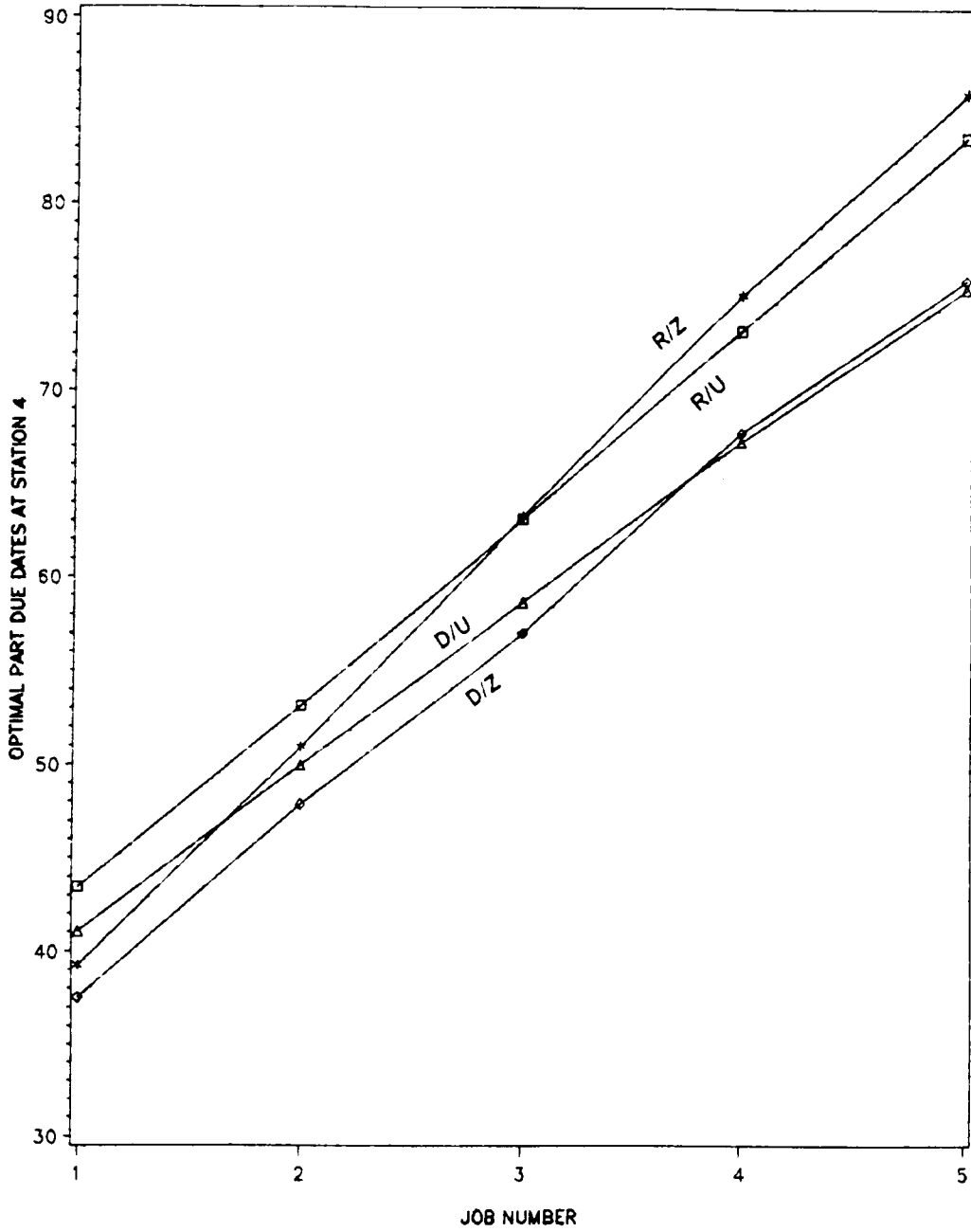


Figure 17. Optimal Part Due Dates at Station 4

### 1) Deterministic Processing, Unlimited Buffers

The interval between successive jobs is governed almost exactly by the duration of the bottleneck station, in this case 10 units for station 3, as shown in Figure 15a. This confirms the intuitive expectation that there is no advantage to expediting parts or jobs if they are going to have to wait at the bottleneck. If  $B$  is the bottleneck station, all decisions for successive jobs up to station  $B$  are spaced by this constant interval.

Subassembly costs have been assumed to be increasing down the line. Therefore, after the bottleneck station, non-critical operations are delayed as much as possible to avoid incurring additional subassembly and finished goods cost. This delay is by the maximum possible amount that can be sustained without delaying the completion time of the batch. The decisions after the bottleneck depend on the relative magnitudes of downstream operation times. If  $B_2$  has the largest operation time downstream of  $B$ , decisions for successive jobs between  $B$  and  $B_2$  are spaced approximately by the duration of  $B_2$ , in this case station 4. Similar relationships hold for all downstream stations after  $B_2$ .

The arrival of the first job initiates the measurement of the makespan and the system starts to incur makespan cost. Accordingly, the first part delivery time is expedited so that the job does not have to wait for the part and incur unnecessary makespan cost. The launch of the second job is somewhat earlier because the system has not filled up by that time, and expediting the second job reduces idle time of stations and blocking of subsequent jobs. This behavior can be seen in Figure 15b.

### 2) Deterministic Processing, Zero Buffers

The interval between successive jobs is still governed almost exactly by the duration of the bottleneck station. Although the exact values are marginally different, all decisions for successive jobs at and before station B are still spaced by this constant interval. Zero buffer capacities do not appear to make a difference to decisions before the bottleneck, since the launch interval is spaced by the bottleneck operation time, and on the average, queues do not develop at earlier workstations with shorter operation times. Even very close operation times (95% of the bottleneck station time) preceding the bottleneck station did not affect this phenomenon.

The batch completion and delivery times are also very marginally different from the unlimited buffer situation. In fact, decisions all along the critical path of the processing network are almost the same, as can be seen in Figure 16. However, decisions for non-critical operations after the bottleneck station may change significantly. Earlier, it had been attempted to delay these operations as much as possible. That cannot be done now, since some subsequent operation might get blocked if an earlier job is holding up a station. Accordingly, the finish times of these operations are now constrained by the smallest start time among all their successors in the network, and thus can be delayed only by a limited amount. This can be seen in the relative decisions for stations 4 and 5 in Table 12, and is shown for station 4 in Figure 17.

### 3) Random Processing, Unlimited Buffers

The interval between successive jobs is still controlled by the bottleneck station. However, the spacing between the decisions for successive jobs at and before station B is greater than the mean bottleneck duration (in this case by about 15%), and depends on the variability as well. This can be seen in the results in Table 13 and in

Figure 15a. Safety times are thus provided for each operation. Also, the initial decisions are expedited a little to reduce the likelihood of interference between jobs. The customer delivery time is increased considerably (by about 11 time units) as a result of the cumulative safety times along the line. The total optimal cost is increased by about 30% with the addition of processing time variability.

After the bottleneck, non-critical jobs are delayed as much as possible without affecting final completion. The time values used to offset these decisions from the critical path decisions are higher than the respective expected operation times because of the safety times.

#### 4) Random Processing, Zero Buffers

The interval between successive jobs continues to be controlled by the bottleneck station. However, the spacing between the decisions for successive jobs at and before station B is even greater than the intervals in the case of random processing with unlimited buffers. This happens because the number of precedence connections increases with limited buffers, and many other paths can become critical. The interval is greater than the mean bottleneck duration by about 20%, thus providing additional safety time for variability. The customer delivery time also increases, by about 2 units compared to the random processing-unlimited buffer case, and by about 13 units compared to the deterministic case, as a result of the additional safety times along the line. The total optimal cost is also correspondingly higher.

After the bottleneck, jobs cannot be delayed as much as in the unlimited buffer case, because of the effect these delays might have on the more critical paths and the batch completion time. Again, the time values used to offset these decisions from the

more critical paths are higher than the values used with unlimited buffers because of the additional safety times.

These observations and insights about optimal policy behavior will be used to guide the development of the heuristic procedure. While it may not be possible to incorporate all of the features, it will be attempted to consider at least the general ideas.

## **5.4 Development of Heuristic Procedure for Multiple Job**

### **Problem**

As mentioned earlier, the computational requirements for direct optimization of the multiple job problem are fairly heavy, and grow rapidly with problem size, especially with limited buffers. Therefore it is necessary to explore heuristic procedures to enable determination of good, near-optimal solutions with lower computational effort. The underlying ideas and the development of the heuristic procedures are described in the remainder of this section.

#### **5.4.1 Bottleneck Station and Critical Path**

The bottleneck station is the station with the largest processing time. It controls the capacity of the line since introducing new jobs into the system at shorter intervals will only cause queues to develop in front of the bottleneck. The critical path is the



path that determines the makespan of the processing network and influences the completion time of the batch. It passes along the first job up to the bottleneck station, then along the bottleneck station for the rest of the jobs, and finally through the remaining stations for the last job. The critical path therefore passes through  $(N + J - 1)$  operations in all.

Das and Sarin (1988) developed a heuristic procedure for the determination of part delivery dates for the multiple job problem in stochastic assembly systems. The heuristic is based on the critical path in the processing network, and breaks down the overall problem into a series of single job problems. In the context of the assembly system with part input, the complete network contains the activities related to part delivery as well. However, those are the result of some decisions that are presently unknown. In view of the importance of the critical path, the heuristic therefore determines the decisions based on the processing network only.

The idea is to optimize decisions along the critical path first, and optimize the other stations with respect to the critical path so that they feed the activities on the critical path. Stations following the bottleneck station are scheduled as late as possible using similar single job optimization procedures, by determining the bottleneck station among the remaining stations, until all stations have been scheduled. The idea is similar to the OPT principle of forward scheduling the bottleneck, and backward scheduling the non-bottlenecks with respect to the bottleneck. The difference in this problem is that the criterion is overall cost rather than the output rate of the system.

The idea behind the heuristic is derived out of the insights gained from the behavior of the sample problems with deterministic processing and unlimited buffers

between station. In order to generalize the above idea to random processing networks and limited buffers, the concept of critical path and the scheduling of non-bottleneck stations may need to be modified.

#### **5.4.2 Incorporating Buffers in Deterministic Networks**

If the processing times are deterministic and limited buffers are introduced, there is no change to the critical path. This can be understood with reference to the network diagrams. As discussed in section 5.2, the largest number of nodes any path can take is  $(N + J - 1)$  due to the levelwise structure of the network. The critical path also passes through the same number of nodes. All precedence restrictions due to buffers must skip at least one job, since the earliest job that can cause operation  $(i, j)$  to be blocked is job  $(j - 2)$ . Therefore any path created due to limited buffers must skip at least one bottleneck operation (for job  $j - 1$ ), and replace it by another operation of smaller duration. In a deterministic network, the bottleneck operation is always the longest, so the length of the path cannot increase, and the critical path remains the same. This is corroborated by the observations regarding optimal policy behavior in the previous section.

However, the scheduling of the non-bottleneck activities changes since now there are many more connections between operations. Accordingly, operations cannot be delayed as much as desired. Therefore, the backward scheduling procedure has to be modified in a way similar to the latest start scheduling in activity networks. For each decision at a node, it must now be checked if it satisfies the minimum of the

starting times of each of its successors already scheduled. If the requirement is not satisfied, the decisions have to be adjusted accordingly.

### **5.4.3 Incorporating Random Processing Times**

When random processing times are introduced, the individual activity durations change from cycle to cycle. The critical path and its length therefore are not confined to the critical path, CP, based on expected durations alone. In fact, every activity now has a finite probability of being critical. The expected completion time,  $E(C)$ , of the network is generally greater than the sum of the expected durations along the critical path,  $\sum_{k \in CP} \mu P_k$ , based on expected durations alone. Likewise, the variance of completion time,  $V(C)$ , will generally be less than the sum of the variances along the critical path,  $\sum_{k \in CP} \sigma P_k^2$ , based on expected durations (Elmaghraby, 1977).

Correction for randomness in processing times could be done in two ways. One way is to identify the changes in the critical path, if any, and carry out the heuristic procedure based on the new critical path. A new method of identifying the critical path therefore has to be found. Alternatively, the same method of identifying the critical path can be used along with some suitable correction factors to capture the effects of randomness on the completion time of the network.

In order to understand the behavior of this process, a number of simulation experiments were conducted. One measure of the criticality of a specific operation is the proportion of times it shows up on the critical path in a certain number of actual realizations. This measure is termed the criticality index by Elmaghraby (1977). The experiments using this measure showed that while the bottleneck station still ap-

peared on the critical path the most frequently, other activities also appeared on the critical path. The relative frequency of the occurrence of different activities on the critical path depends on the relative durations and variabilities. If the critical path is interpreted as the path with the highest occurrence of critical activities, it does not always traverse the network in the same way as before. Even though it passes through the bottleneck station most frequently, it may pass through previous stations even after the first job and cross over to the subsequent stations before the last job. This behavior is shown in Table 14.

In the context of the assembly situation, this suggests that the launch intervals of some later jobs could be governed by operation times smaller than the bottleneck station. This was not borne out by the observations from the optimal solutions in Section 5.3. In any case, the sum of the expected durations and the variances along the critical path identified by this method will not add up to the expected duration and variance of completion time respectively. Correction factors will therefore still need to be determined. Consequently, it was decided to continue with the old definition of the critical path and look for other means to correct for randomness.

As discussed,  $E(C) \geq \sum_{k \in CP} \mu P_k$ , and  $V(C) \leq \sum_{k \in CP} \sigma P_k^2$ , with the equality holding only if the network is deterministic or if the durations along the CP considerably dominate all the other paths. In setting up a single job problem along the critical path, the entire processing network is sought to be replaced by an equivalent chain of activities. If the path based on expected durations is used to solve a single job problem and optimize the critical path, the actual average duration will be underestimated, while the variance would be overestimated. This would result in incorrect decisions.

**Table 14. Frequency of Critical Activities in a Random Processing Network**

Number of Stations: 8, Number of Jobs: 6

Mean Operation Times: 4-9.5-7-10-9-5-8-6, Bottleneck station: 4

Standard Deviations : 1-3-2-3-2.5-1-2-2

Number of replications: 1000

Job	Station							
	1	2	3	4	5	6	7	8
1	<b>1000</b>	<b>972</b>	387	351	73	11	11	0
2	28	<b>613</b>	<b>241</b>	<b>479</b>	145	19	30	0
3	0	408	162	<b>538</b>	251	52	81	3
4	0	286	139	<b>540</b>	366	108	186	15
5	0	167	122	<b>490</b>	<b>496</b>	208	377	123
6	0	57	62	315	<b>603</b>	<b>608</b>	<b>877</b>	<b>1000</b>

Critical path based on critical activities:

(1 1)-(2 1)-(2 2)-(3 2)-(4-2)-(4-3)-(4 4)-(4 5)-(5 5)-(5 6)-(6 6)-(7 6)-(8 6)

To correct for variability, the activity durations have to be corrected so that the resulting durations along the critical path yield the same expected duration and variance as the entire network. Analytical methods have been developed for the determination of the expected project duration in PERT networks, but the determination of the variance is a more difficult problem. It was therefore decided to estimate these quantities by simulation, using 1000 replications of the processing network. Some experimentation was done to determine a reasonable method of allocating the actual expectation and variance along the critical path. The best and most convenient method was found to be to perform the allocation in proportion to the expectations and variances of the respective activities along the critical path. That is,

$$\overline{\mu P_i} = \mu P_i \frac{E(C)}{E(CP)}$$

$$\text{and } \overline{\sigma P_i^2} = \sigma P_i^2 \frac{V(C)}{V(CP)}.$$

The resulting values are then used in the single job optimization procedure.

This method of allocation, though not unique, serves as a reasonable approximation for the heuristic. Its actual performance will be assessed when the results are compared with the optimal solutions. The larger activity times generated by this method for each station are corroborated by the larger intervals observed in the optimal solutions with random processing times. In case the durations are deterministic, the expected duration will be the same as that obtained by the deterministic estimate, so no correction would be required.

Consideration of limited buffers with random processing times requires a combination of the methods discussed above. The expected duration of the network will be different from the unlimited buffer case because of the many more precedence con-

nections. However, the correction proceeds in a similar fashion, and the resulting durations confirm the even larger intervals found in the optimal solutions. The consideration of buffers during backward scheduling of the non-bottleneck activities also proceeds in the same way. As discussed earlier, decisions at each node have to be checked for violation of start times of each of its successors already scheduled, and translated accordingly if they do not satisfy the restrictions.

#### **5.4.4 Heuristic Procedure for the Multiple Job Problem**

Based on the discussions above and the insights gained by examination of the solutions from the nonlinear programming method, the heuristic procedure has been developed and is described below. A flowchart of the procedure is shown in Figure 18.

##### **Step 1:**

Identify the bottleneck station (B) in the network. The critical path (CP) is then the chain of operations (1,1), (2,1),....., (B,1), (B,2),....., (B,J), (B + 1,J),....., (N,J).

Determine the sum of the expectations E(CP) and variances V(CP) along the critical path.

Simulate the processing network to determine the mean E(C) and variance V(C) of its completion time.

Adjust the expected durations and variances of processing times to add up to E(C) and V(C) respectively along the CP, using the relations  $\overline{\mu P_i} = \mu P_i \frac{E(C)}{E(CP)}$  and  $\overline{\sigma P_i^2} = \sigma P_i^2 \frac{V(C)}{V(CP)}$ . These adjusted values are used for the remaining steps.

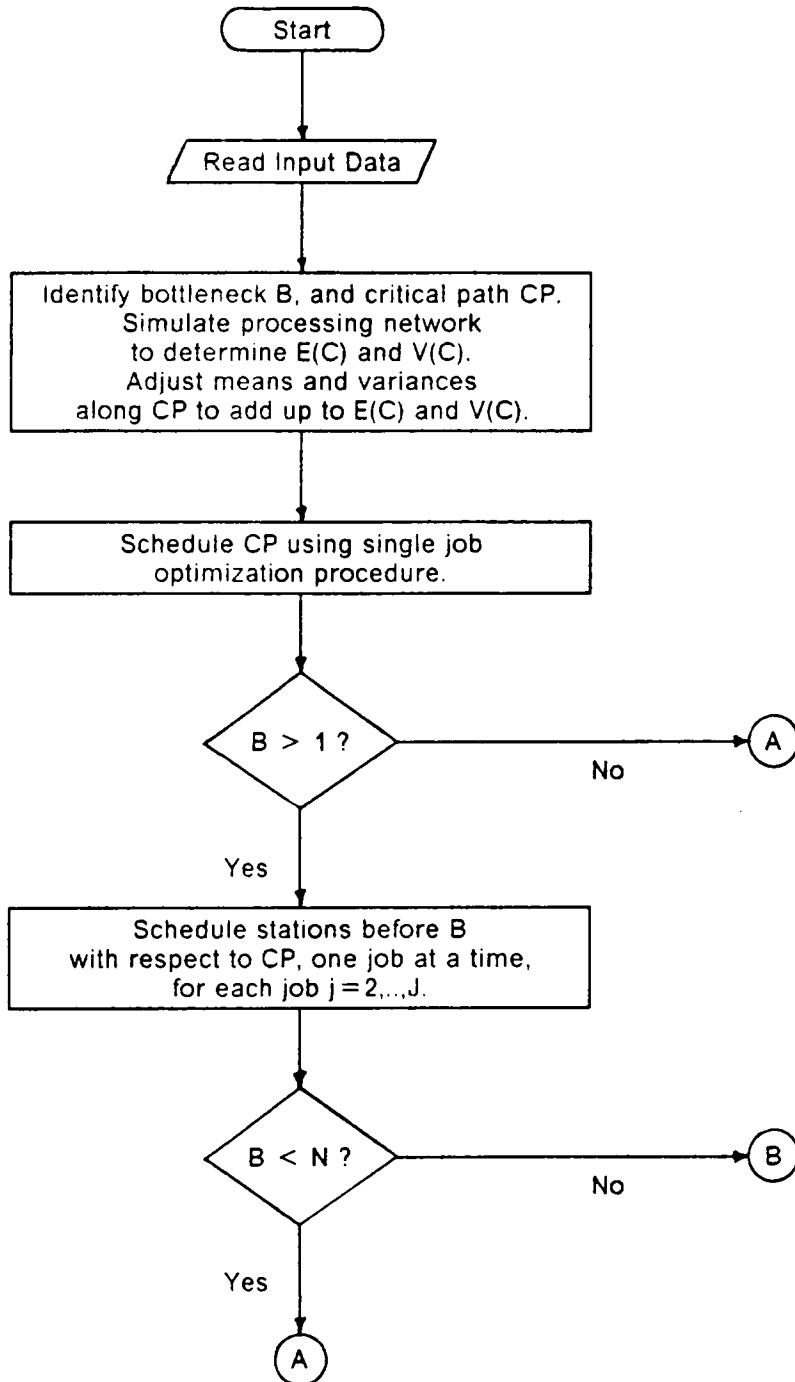


Figure 18. Flowchart for Implementation of Heuristic



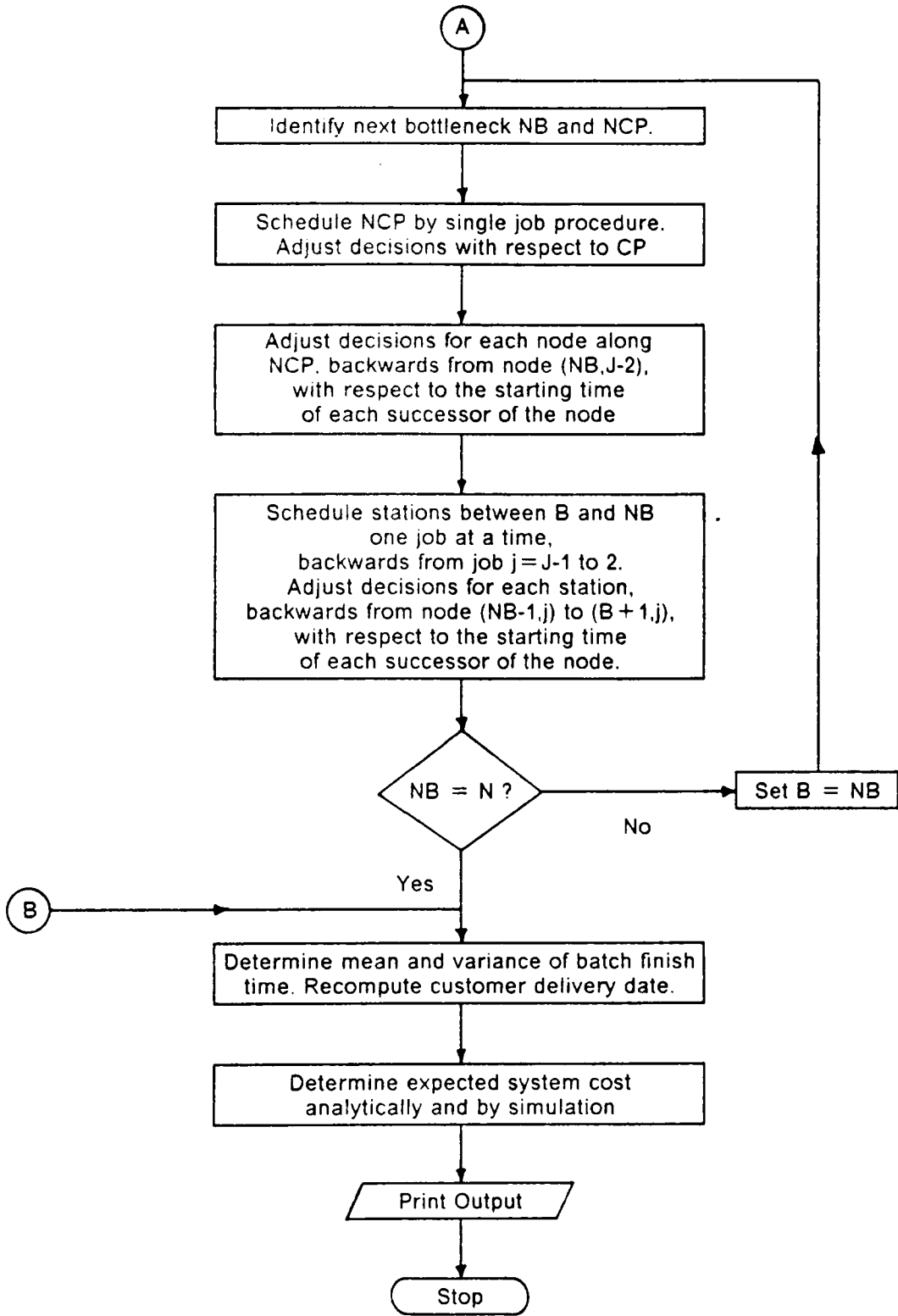


Figure 18 (Continued). Flowchart for Implementation of Heuristic

In deterministic networks,  $E(C) = E(CP)$  and  $V(C) = V(CP) = 0$ , so no adjustment is required.

Step 2:

Optimize the chain of stations along the CP using the single job optimization procedure from Chapter 4. A dummy station is added at the end to emulate the customer delivery date, with associated inventory and tardiness costs. Since the CP determines the makespan, the makespan cost CM is added on to each subassembly cost along the chain.

Step 3:

Schedule the nodes before the bottleneck station for each job  $j = 2, \dots, J$ . The chain of nodes optimized for job  $j$  is  $(1,j), (2,j), \dots, (B-1,j)$ .

To ensure that the completion time of this chain does not interfere with the CP, the start time of node  $(B,j)$  is treated as the due date for this chain.

Step 4:

Identify the largest operation time downstream of the bottleneck station. This becomes the new bottleneck station (NB) and the new chain being optimized (NCP) becomes  $(B + 1,1), (B + 2,1), \dots, (NB,1), (NB,2), \dots, (NB,J-1)$ .

Optimize this chain (NCP) using the single job optimization procedure, treating the start time of node  $(NB,J)$  as the due date.

Adjust the decisions along the NCP so that the finish time of node  $(NB,J-1)$  is equal to the start time of node  $(NB,J)$ .

### Step 5:

For each remaining node along the NCP, starting backwards from (NB,J-2) check if the finish time is no more than the start time of each successor of that node. That is, if  $SU_k$  is the set of successors of node k already scheduled, check if  $\mu F_k \leq \mu S_m$ , for each  $m \in SU_k$ .

If the condition is satisfied, leave the decisions unchanged. If not, translate the timings of node k and all its predecessors along the NCP so that its finish time is no more than the smallest of the start times of the successors.

Repeat this step for each node (NB,J-3),..., (B + 1,1), using the adjusted timings from the previously considered node.

If the network has unlimited buffers, this step will not cause any adjustments to be made.

### Step 6:

Schedule the set of stations between the old and the new bottlenecks for each job  $j=2,\dots,J-1$ . The chain being optimized is (B + 1,j),..., (NB-1,j), using the start time of node (NB,j) determined in the previous step as the due date for the chain.

For each node along this chain, starting backwards from node (NB-1,j), adjust the decisions as in Step 5, so that the finish time of that node is no more than the smallest start time of each of its successors.

### Step 7:

After completion of Step 6, treat the largest operation time downstream of the current bottleneck treated as the new bottleneck, and the current bottleneck as the old bottleneck.

Continue Steps 4-6 until the last station and all nodes have been scheduled.

### Step 8:

Determine the mean  $\mu F_{N,J}$  and variance  $\sigma F_{N,J}^2$  of the batch using the decisions determined. Recalculate the optimal customer delivery date by solving a single station optimization problem.

## **5.5 Implementation of Heuristic and Comparison of**

### **Results**

The stepwise heuristic outlined above has been implemented in a computer program in two different procedures. The first, referred to as Heur-N, uses the Non-linear Programming approach to solve the single job problems in the heuristic. The second, referred to as Heur-A, uses the Approximation technique developed in Chapter 4 to solve the single job problems. After all the decisions have been determined, the expected total cost for the set of decisions is measured by the analytical and simulation routines used in Section 5.3.

Table 15 shows the results from the heuristic algorithm Heur-N for the same sample problems whose optimal solutions were shown in Table 13. The values obtained for the decisions are quite similar, and the cost is within 1.3% of the optimal solution, which is a reasonable performance considering the computational effort saved. The solution using Heur-A results in an error of only 1.6%, at even lower computational effort.

**Table 15. Heuristic Solutions for Sample 5 Station, 5 Job Problems**

Mean Operation Times: 5-8-10-9-7, Standard Deviations 2-2-4-2-2

Bottleneck Station: 3

a) Random Processing, Unlimited Buffers

Job	Start Time	Part Delivery Times at Station				
		1	2	3	4	5
1	15.00	11.60	16.90	25.45	41.45	61.80
2	21.85	21.70	27.10	36.05	51.10	69.60
3	32.60	32.65	37.85	46.80	61.10	76.60
4	43.65	43.60	48.85	58.20	71.00	83.50
5	55.05	55.10	60.25	70.25	83.10	94.30

Delivery Date 107.10      Total Cost 2339.0

b) Random Processing, Zero Buffers

Job	Start Time	Part Delivery Times at Station				
		1	2	3	4	5
1	15.00	11.60	16.80	25.80	39.00	53.50
2	22.00	22.05	27.40	36.20	50.40	65.20
3	33.00	33.00	38.35	47.50	62.23	76.50
4	44.45	44.45	49.80	59.60	72.65	85.40
5	56.25	56.20	61.50	71.90	85.00	96.25

Delivery Date 109.30      Total Cost 2423.5

A modification of the safety buffer rule with a safety factor of 1 standard deviation has also been implemented. The modification includes the backward scheduling logic to delay jobs after the bottleneck. It also incorporates the adjustments necessary to consider limited buffers between stations.

A number of test problems have been run to investigate the performance of the heuristic procedure, and also examine the effect of changes in system structure and parameters. These problems include deterministic and random processing times, and also unlimited and zero buffers. In order to study the effect of the size and placement of buffers, some problems have also been run with a few buffer positions in selected locations along the line.

For the purpose of comparison of the quality of the solution, the total costs for the decisions from the optimal, heuristic and the modified buffer rule solution approaches have to be evaluated. To avoid making any errors due to assumptions about the sum or the maximum of the probability distributions, and any errors due to ignoring of correlations, the costs have been estimated by simulation using 1000 replications. Because of the heavy computational time requirements of the direct optimization method, the largest problem considered for the comparison was 10 stations  $\times$  10 jobs. The results for the test problems are shown in Table 16.

The relative errors in the different methods are shown in Table 17, along with the computation time (CPU) required for each method on the IBM 3090 mainframe computer. These values exclude the time required to determine the system cost using the analytical and simulation routines, which are common to all the methods. For the nonlinear programming method, a function tolerance level of 0.5% has been used for termination of the algorithm for a reasonable compromise between solution quality

and computational effort. Lower accuracy levels resulted in poorer solutions without significant reductions in computation times, while increasing the accuracy levels was found to increase the computation times considerably without corresponding improvements in the objective function.

The results show clearly the superiority of the heuristic procedures over the modified buffer rule. Both heuristics perform better than the buffer rule in every case by 4-10%. Moreover, the buffer rule does not distinguish between changes in cost structures, so the differences become greater as the problems get unbalanced in costs. Interestingly, the solutions from heuristic Heur-A are only marginally inferior to those from Heur-N, confirming the effectiveness of the single job Approximate solution technique, especially when embedded in the multiple job problem.

The errors in the heuristic solutions compared to the optimal values ranged between 0.2-3.4% . These errors must be viewed in light of the computational time requirements. For example, the  $10 \times 10$  problem with random processing times and zero buffers required 405 seconds of CPU time on the mainframe computer to obtain the optimal solution, while Heur-N took only 3.8 seconds and Heur-A took only 1.1 seconds. The computation time for the nonlinear programming approach is generally greater for problems with limited buffers, since the application of the Maximum operator and the calculation of correlations has to be done a much greater number of times.

The problems tested have demonstrated the effectiveness of the heuristics and the adaptations made to incorporate randomness and limited buffers. The heuristics yield near-optimal solutions with only a fraction of the computational effort required by the unconstrained optimization procedure. Moreover, the heuristics yield good

**Table 16. Results for Multiple Job Problems for Normal Distribution**

Pr. No.	Stns. × Jobs	Proc. Times	Buffer Capacities	Optimal Cost	Buf. Rule Cost	% Err.	Heur-N Cost	% Err.	Heur-A Cost	% Err.
1	5 × 5	Det.	Unlim.	1799.3	1937.6	7.7	1804.5	0.3	1803.4	0.2
2	5 × 5	Det.	0-1-1-0	1817.4	1956.0	7.6	1823.1	0.3	1821.7	0.2
3	5 × 5	Det.	All 0	1831.3	1967.9	7.5	1837.2	0.3	1835.6	0.2
4	5 × 5	Ran.	Unlim.	2318.5	2567.4	10.7	2339.0	0.9	2341.0	1.0
5	5 × 5	Ran.	0-1-1-0	2338.9	2585.2	10.5	2357.6	0.8	2360.8	0.9
6	5 × 5	Ran.	All 0	2386.6	2628.3	10.1	2423.5	1.5	2427.9	1.7
7	5 × 10	Det.	Unlim.	5297.6	5658.2	6.8	5323.2	0.5	5319.7	0.4
8	5 × 10	Det.	0-1-1-0	5416.1	5769.2	6.5	5433.3	0.3	5431.3	0.3
9	5 × 10	Det.	All 0	5526.9	5862.6	6.1	5533.6	0.1	5532.8	0.1
10	5 × 10	Ran.	Unlim.	6707.9	7461.8	11.2	6760.9	0.8	6763.1	0.8
11	5 × 10	Ran.	0-1-1-0	6858.3	7627.1	11.2	6903.8	0.7	6904.5	0.7
12	5 × 10	Ran.	All 0	7211.7	7861.7	9.0	7298.6	1.2	7299.1	1.2
13	8 × 6	Det.	Unlim.	3267.3	3641.1	11.4	3296.8	0.9	3297.6	0.9
14	8 × 6	Det.	0-1-1- -2-0-0-1	3271.4	3642.2	11.3	3297.9	0.8	3299.4	0.8
15	8 × 6	Det.	All 0	3332.8	3698.6	11.0	3352.5	0.6	3352.9	0.6
16	8 × 6	Ran.	Unlim.	4301.5	4638.8	7.8	4367.4	1.5	4366.2	1.5
17	8 × 6	Ran.	0-1-1- -2-0-0-1	4308.3	4650.9	8.0	4379.3	1.6	4376.3	1.6
18	8 × 6	Ran.	All 0	4417.5	4710.1	6.6	4495.5	1.8	4498.2	1.8
19	10 × 10	Det.	Unlim.	9254.2	10243.1	11.2	9498.4	2.6	9516.4	2.8
20	10 × 10	Det.	0-1-2-0- -0-1-0-0-0	9355.5	10373.4	10.9	9574.5	2.3	9592.8	2.5
21	10 × 10	Det.	All 0	9533.7	10472.4	9.9	9855.3	3.4	9853.4	3.4
22	10 × 10	Ran.	Unlim.	12177.3	13465.1	10.6	12379.7	1.7	12385.7	1.7
23	10 × 10	Ran.	0-1-2-0- -0-1-0-0-0	12313.8	13554.3	10.1	12533.7	1.8	12542.5	1.9
24	10 × 10	Ran.	All 0	12685.8	13861.8	9.3	13078.3	3.1	13101.9	3.3



**Table 17. Comparison of Techniques for Multiple Job Problems for Normal Distribution**

Pr. No.	Stns. x Jobs	Proc. Times	Buffer Caps.	NLP CPU (secs.)	Buf. Rule CPU (secs.)	% Err. Err.	Heur-N CPU (secs.)	% Err. Err.	Heur-A CPU (secs.)	% Err. Err.
1	5 x 5	Det.	Unlim.	6.0	0.01	7.7	0.59	0.3	0.36	0.2
3	5 x 5	Det.	All 0	6.7	0.01	7.6	0.60	0.3	0.36	0.2
4	5 x 5	Ran.	Unlim.	7.8	0.01	10.7	0.63	0.9	0.37	1.0
6	5 x 5	Ran.	All 0	9.5	0.01	10.1	0.71	1.5	0.27	1.7
7	5 x 10	Det.	Unlim.	35.0	0.01	6.8	1.3	0.5	0.40	0.4
9	5 x 10	Det.	All 0	117.1	0.01	6.1	1.3	0.1	0.41	0.1
10	5 x 10	Ran.	Unlim.	44.3	0.01	11.2	2.3	0.8	0.43	0.8
12	5 x 10	Ran.	All 0	52.7	0.01	9.0	1.9	1.2	0.65	1.2
13	8 x 6	Det.	Unlim.	22.6	0.01	11.4	1.2	0.9	0.23	0.9
15	8 x 6	Det.	All 0	46.1	0.01	11.0	1.2	0.6	0.23	0.6
16	8 x 6	Ran.	Unlim.	32.8	0.01	7.8	1.3	1.5	0.24	1.5
18	8 x 6	Ran.	All 0	44.5	0.01	6.6	1.9	1.8	0.28	1.8
19	10 x 10	Det.	Unlim.	189.6	0.01	11.2	1.8	2.6	0.67	2.8
21	10 x 10	Det.	All 0	427.6	0.01	9.9	3.6	3.4	0.66	3.4
22	10 x 10	Ran.	Unlim.	242.6	0.01	10.6	4.1	1.7	0.92	1.7
24	10 x 10	Ran.	All 0	405.8	0.01	9.3	3.8	3.1	1.1	3.3

solutions without computation of correlations between operation finish times. Therefore there is no need for the elaborate procedures necessary to explicitly determine and keep track of the correlations in order to describe and measure the performance of the system.

## **5.6 Generalization of Optimal and Heuristic Methods for Non-normal Distributions**

The solution approaches developed in the previous sections have been generalized to the lognormal and gamma distributions. The optimization method based on nonlinear programming has been applied to the lognormal distribution using the descriptive relationships for correlated random variables discussed in Section 3.4 and detailed in Appendix A. Since the necessary descriptive relationships for correlated random variables are not available for the gamma distribution, optimal solutions cannot be determined.

The heuristic procedure developed in the previous sections has been implemented for both distributions, using the single job optimization technique as used in Section 4.7.1, and the approximate solution method described in Section 4.7.2. The results from these methods and their relative performance is shown in Tables 18 and 19. For the  $10 \times 10$  problems, optimal solution was not attempted for the lognormal distribution because of the heavy computation time requirements. The figures in parentheses represent the errors with respect to Heur-N when optimal solutions are not available for comparison.

**Table 18. Results for Multiple Job Problems for Different Distributions**

Pr. No.	Stns. × Jobs	P.T.	Buffer Caps.	Dis.	Optimal Cost	Buf. Rule Cost	% Err.	Heur-N Cost	% Err.	Heur-A Cost	% Err.
4	5 × 5	R	Unlim.	N	2318.5	2567.4	10.7	2339.0	0.9	2341.0	1.0
				L	2392.0	2621.7	9.6	2419.5	1.1	2419.9	1.1
				G	-	2592.4	(8.7)	2384.6		2390.2	(0.2)
6	5 × 5	R	All 0	N	2386.6	2628.3	10.1	2423.5	1.5	2427.9	1.7
				L	2482.4	2685.1	8.2	2517.3	1.4	2519.4	1.5
				G	-	2652.4	(6.6)	2489.3		2498.1	(0.4)
10	5 × 10	R	Unlim.	N	6707.9	7461.8	11.2	6760.9	0.8	6763.1	0.8
				L	6872.9	7557.4	10.0	6926.7	0.8	6928.5	0.8
				G	-	7534.1	(10.6)	6807.1		6799.1	(-0.1)
12	5 × 10	R	All 0	N	7211.7	7861.7	9.0	7298.6	1.2	7299.1	1.2
				L	7428.8	7970.8	7.3	7583.3	2.1	7577.2	2.0
				G	-	7922.7	(6.2)	7456.7		7457.0	(0.0)
16	8 × 6	R	Unlim.	N	4301.5	4638.8	7.8	4367.4	1.5	4366.2	1.5
				L	4443.1	4708.9	6.0	4556.6	2.6	4546.8	2.4
				G	-	4708.1	(5.2)	4477.3		4480.4	(0.1)
18	8 × 6	R	All 0	N	4417.5	4710.1	6.6	4495.5	1.8	4498.2	1.8
				L	4591.4	4900.8	6.7	4730.7	3.0	4725.1	2.9
				G	-	4889.7	(5.6)	4628.4		4648.4	(0.4)
22	10 × 10	R	Unlim.	N	12177.3	13465.1	10.6	12379.7	1.7	12385.7	1.7
				L	-	13634.2	(6.3)	12827.8		12831.8	(0.0)
				G	-	13601.2	(7.4)	12660.4		12614.4	(-0.3)
24	10 × 10	R	All 0	N	12685.8	13861.8	9.3	13078.3	3.1	13101.9	3.3
				L	-	14064.1	(3.4)	13607.8		13620.8	(0.1)
				G	-	13999.5	(4.9)	13350.1		13406.7	(0.4)

(.) : Errors with respect to Heur-N

**Table 19. Comparison of Techniques for Multiple Job Problems for Different Distributions**

Pr. No.	Stns. x Jobs	P.T.	Buffer Caps.	Dis.	NLP CPU (secs.)	Buf. Rule CPU (secs.)	% Err.	Heur-N CPU (secs.)	% Err.	Heur-A CPU (secs.)	% Err.
4	5 x 5	R	Unlim.	N	7.8	0.01	10.7	0.63	0.9	0.35	1.0
				L	33.1	0.01	9.6	2.0	1.1	1.2	1.1
				G	-	0.01	(8.7)	5.5		3.3	(0.2)
6	5 x 5	R	All 0	N	9.5	0.01	10.1	0.71	1.5	0.27	1.7
				L	65.7	0.01	8.2	1.6	1.4	0.97	1.5
				G	-	0.01	(6.6)	5.6		3.3	(0.4)
10	5 x 10	R	Unlim.	N	44.3	0.01	11.2	2.3	0.8	0.43	0.8
				L	144.7	0.01	10.0	4.4	0.8	2.1	0.8
				G	-	0.01	(10.6)	19.9		6.9	(-0.1)
12	5 x 10	R	All 0	N	52.7	0.01	9.0	1.9	1.2	0.65	1.2
				L	404.3	0.01	7.3	6.9	2.1	2.3	2.0
				G	-	0.01	(6.2)	20.4		7.1	(0.0)
16	8 x 6	R	Unlim.	N	32.8	0.01	7.8	1.3	1.5	0.24	1.5
				L	134.6	0.01	6.0	4.7	2.6	1.5	2.4
				G	-	0.01	(5.2)	16.8		6.1	(0.1)
18	8 x 6	R	All 0	N	44.5	0.01	6.6	1.9	1.8	0.28	1.8
				L	315.0	0.01	6.7	4.7	3.0	1.4	2.9
				G	-	0.01	(5.6)	17.0		6.1	(0.4)
22	10 x 10	R	Unlim.	N	242.6	0.01	10.6	4.1	1.7	0.92	1.7
				L	-	0.01	(6.3)	8.9		4.6	(0.1)
				G	-	0.01	(7.4)	44.9		14.5	(-0.3)
24	10 x 10	R	All 0	N	405.8	0.01	9.3	3.6	3.1	1.1	3.3
				L	-	0.01	(3.4)	10.8		4.6	(0.1)
				G	-	0.01	(4.9)	34.5		14.8	(0.4)

(.) : Errors with respect to Heur-N

The results show that the performance of both heuristic procedures is consistent across all the distributions considered. In fact, Heur-A occasionally produces a marginally better overall solution than Heur-N. The error ranges are of similar magnitude for the distributions over all the problems considered.

The computation time requirements are significantly higher for non-normal distributions for all the methods because of the greater amount of calculations required. The gamma distribution requires the most time since its relationships are the most involved. This difference in timings is especially pronounced in the case of optimal solution to the lognormal problems. These increases in computation time increase the preferability of the heuristics in solving the problem.

## **5.7 Summary**

In this chapter the problem of determining optimal part delivery dates in the multiple job assembly line has been addressed. A representation of the system as a precedence network has been used to understand the process and its description. Modifications necessary to incorporate limited buffers between stations have been made in the descriptive model. The objective function has been developed and a computational scheme devised for its evaluation.

Two different solution approaches have been applied. The nonlinear programming approach uses the analytically computed value of the function to determine the optimal solution. This approach has been developed for the normal and lognormal

distributions. A critical path - bottleneck station based heuristic has been implemented with appropriate corrections to incorporate random processing times and limited buffers between stations. The heuristic uses the single job solution techniques developed in Chapter 4 to determine the decisions. The performance of the heuristic has been found to be extremely promising when compared with the optimal solutions, and much better than the safety buffer rule using 1 standard deviation safety time. The solution method gives reasonably good results (within 0.2-4%) compared to the nonlinear programming method, with comparatively negligible computational effort. The heuristic method has been generalized to include all the distributions considered, and retains its effectiveness over them.

## **6.0 Analysis of Results and System Behavior**

### **6.1 Introduction**

In the previous chapters, the methodology for solution of the single and multiple job problem has been developed. In this chapter, this methodology is used to conduct experiments to gain insights as to the behavior of small lot assembly systems with part input at each stage under different operating conditions.

The objective of the experimentation has been to identify any patterns and trends in policy behavior that could be useful in the design and management of such systems. The experiments have been carried out under different scenarios using numerical values. While the output values obtained are obviously dependent upon the numbers chosen, the understanding generated regarding the trends in the behavior of the system is more important. In view of the amount of experimentation involved and the heavy computational requirements of the nonlinear programming solution approach, the results are compared for the solutions determined by the heuristic procedure.

As alluded to above, numerical based observations regarding system behavior are dependent on the numbers chosen. To increase the generality of these observations, the sample problems have been created to consider ranges in costs and vari-

ances. Since the cost values can be scaled up or down without altering the characteristics of the problem, the observations regarding system behavior depend only on the range considered. The effect of varying individual costs is examined in the next section. Changes in magnitudes of variances affect the variability and behavior of the process, as well as other factors like choice of distribution to use for part delivery and operation processing times. Therefore, greater experimentation is required regarding the impact of variances even if the change is uniform. The sensitivity of system behavior to changes in variance, individually and collectively, is also examined in the next section.

The following data is used to conduct the experimentation. The problem size varies between 5 stations  $\times$  5 jobs to 10 stations  $\times$  10 jobs. In creating the problem data, it has been assumed that the subassembly is at least as expensive as the part, and that subassembly costs increase down the line. The part inventory costs vary between 0.5 to 3 units. Subassembly inventory costs at a station are generated in accordance with the costs of the subassembly and the part at the previous station, and some further cost increase due to value addition. Consequently, the values of the subassembly inventory costs increase down the line, and are generated in the range between 1 to 20 units. The finished inventory cost is generated similarly from the subassembly and part costs at the last station. Delivery tardiness costs range from 2 to 3 times the finished inventory costs.

The processing times at assembly stations range between 7 to 10 units, the largest in each problem being 10 units. The minimum among the processing times generated is between 7 and 8 units, and this spread of processing times constitutes a moderate balance between stations. The processing time standard deviations



range from 2 to 4 units, corresponding to coefficients of variation between 0.3 to 0.4. The variabilities of part delivery times and processing times have been kept to be of the same order so that neither dominates the other. Accordingly, part delivery time standard deviations range between 1 to 4 units. Assuming similar coefficients of variation, this implies that the lead times of the parts are of the same order of magnitude as the processing times.

## **6.2 Sensitivity to System Parameters**

In this section the sensitivity of the solutions to changes in system parameters is examined. These comparisons have been done in order to understand the system behavior under changed conditions, and also to get an idea of the kind of errors that can result due to incorrect estimation of the parameters. Moreover, it is necessary to study how the effects of these changes vary across all distributions, and how sensitive the observations regarding system behavior in the later sections may be due to changes in the parameters.

### **6.2.1 Effect of Changes in Costs**

The cost changes experimented with are the finished inventory cost (CF), tardiness cost (CT), makespan cost (CM), and different part inventory costs (CE). The part costs have been varied at three stations, namely, at the bottleneck (station 4), a station before the bottleneck, and a station after the bottleneck, to examine their rel-

ative importance along the line. It is assumed that with a change in part cost, there is a corresponding change in downstream subassembly costs as well. The CF and CT values are relatively larger than the other cost values in the problem due to the structure of the problem of increasing costs along the line. Table 20 shows the results from an indicative set of experiments for each distribution for the sample 8 station  $\times$  6 job problem, along with the percentage change in overall cost arising out of the change.

Increasing the unit cost of finished inventory (CF) to double its original value increases the total cost of the solution considerably, by as much as 45-47% in the sample problem. This is in spite of the decisions after the bottleneck being delayed in order to reduce the impact of the change. This magnitude of increase is due to the high contribution of finished inventory cost to the total cost, as will be seen in Section 6.3. The magnitude of increase is similar across the distributions. Similar trends hold for other magnitudes of changes in CF.

Doubling the unit cost of delivery tardiness causes the prescribed delivery date to be delayed compared to the initial problem to reduce the probability of tardiness. The total cost increase is 3.7% for the normal distribution in the sample problem. This increase is higher by an additional 1-2% for non-normal distributions. Note that even though the unit cost of tardiness is the largest among all unit costs, its effect on the total cost is considerably smaller than the effects of CF and CM, and is of the same order as the effect of individual part costs.

Increasing the makespan cost (CM) to double its original value causes all decisions to be accelerated in order to reduce the overall makespan. The overall cost

**Table 20. Effect of Changes in Selected Parameters on Total Costs at Solutions**

Number of Stations = 8, Number of Jobs = 6  
 Random Delivery & Processing Times, Unlimited Buffers

Parameter Changed	Initial Value	Change	$TC^*$ Normal	$TC^*$ Lognormal	$TC^*$ Gamma
-	-	-	6830	7192	7049
CF	20	+ 20	10035 (46.9)	10427 (45.0)	10378 (47.2)
CT	60	+ 60	7084 (3.7)	7562 (5.1)	7453 (5.7)
CM	10	+ 10	8276 (21.1)	8656 (20.4)	8492 (20.5)
$CE_1$	0.5	+ 1.0	7150 (4.7)	7521 (4.6)	7372 (4.6)
$CE_4$	2.0	+ 1.0	7094 (3.9)	7486 (4.0)	7347 (4.3)
$CE_7$	1.0	+ 1.0	7064 (3.4)	7429 (3.3)	7326 (3.9)
$\sigma D_2$	2.0	+ 2.0	6874 (0.7)	7236 (0.6)	7108 (0.8)
$\sigma D_4$	2.0	+ 2.0	6874 (0.7)	7240 (0.6)	7110 (0.8)
$\sigma D_8$	2.0	+ 2.0	6895 (1.0)	7263 (1.0)	7148 (1.4)
$\sigma P_2$	3.0	+ 2.0	7124 (4.3)	7679 (6.8)	7390 (4.8)
$\sigma P_4$	4.0	+ 2.0	7288 (6.7)	7818 (8.7)	7753 (10.0)
$\sigma P_8$	2.0	+ 2.0	7094 (3.9)	7621 (6.0)	7447 (5.7)

(.): % Cost increase with respect to Initial Problem

increases by 20-21% for the sample problem. This increase is of similar magnitude over the distributions.

The effect of increasing the part inventory costs (CE) by 1 unit, with consequent increases in downstream subassembly inventory costs, depends on where the change occurs on the line. Generally, the increase in CE causes the part to be delayed to minimize its impact on the total cost. Since the bottleneck affects the timings of the entire system, an increase in part cost at that station causes delays in decisions at stations on either side of the bottleneck. The effect on cost is the largest when the change occurs early in the line, as can be seen in the relative values for  $CE_1$ ,  $CE_4$ , and  $CE_7$ . However, the overall impact is moderate, ranging between 3.3-4.7%, and does not change appreciably with distribution.

Costs have been varied individually at stations before and after the bottleneck, to indicate the effect of the change in part inventory cost in different sections of the line. Similar trends will hold if costs are changed at multiple stations instead of just one. Changing all cost values in the problem uniformly would result in simply scaling all the costs. Therefore, there would be no effect on the decisions obtained. Accordingly, uniform changes in costs have not been experimented with.

The following conclusions can be drawn from the experiments described above.

1. Changes in finished inventory cost and makespan cost can have a very significant effect on total cost, while the effects of change in tardiness cost and part inventory costs are relatively moderate.
2. The effects on total cost arising out of cost parameter changes are of similar magnitude across different distributions. This suggests that any comparisons

between different distributions will not be affected substantially if the costs change or there is an error in cost estimation. This increases the generality of the analysis done in later sections.

### **6.2.2 Effect of Changes in Variability**

The effects of changes in the part delivery variances and the processing time variances have also been examined. As in the case of the costs, these changes are made at three stations, namely, at the bottleneck, a station before the bottleneck, and a station after the bottleneck, to examine if the station at which the change occurs has any relationship with the magnitude of the effect. These results are also shown in Table 20.

The effect of increasing the part delivery variance by two units at stations 2, 4, and 6 respectively is basically to accelerate the part concerned. This is done in order to provide for additional protection against the added variability. In the case of the bottleneck (station 4), parts required at neighboring stations are also accelerated marginally because of the consequent changes in operation finish times. The effect on overall cost is greater for changes at stations after the bottleneck, because the change is transmitted more directly to finish times at the end of the line. This effect can be seen to be different from the effect due to change in the CE values, which have greater impact if they occur earlier on the line. The overall cost change, however, is relatively small and ranges between 0.6 to 1.4%. The effect is marginally different for non-normal distributions.

The effect of increasing the processing time standard deviation by 2 units is considerably more significant. This is because the processing time variances are directly added to the operation start time variances, thus increasing system variability. This causes decisions at that station as well as subsequent stations to be delayed. The values of processing time variances have been chosen to reflect coefficients of variation between 0.3 and 0.4. Since the bottleneck station has the largest mean processing time, its variance is accordingly the largest as well. The bottleneck station effectively determines the scheduling of the system, so that a change in variance there causes almost all decisions along the line to be delayed. The highest cost increase also occurs in this case, 6.7% for the normal distribution, and 8.7% and 10% for the lognormal and gamma distributions respectively.

The effect of change in variance is also greater for stations after the bottleneck than for changes before the bottleneck, because the change is transmitted more directly to finish times at the end of the line. The magnitudes of increases in cost as a result of changes in processing time variances are considerably higher than for similar changes in part delivery variances, ranging between 3.9-10% overall. The increases are greater for non-normal distributions because the differences between distributions become more significant at higher variabilities.

Considering the sensitivity of the solution to changes in variance, and also the importance of the choice of distribution when variances are high, the impact of changes in variances has been examined further. These experiments have been done separately for delivery and processing time variances. In these experiments, all variances have been changed by a constant value, the magnitude of the change

ranging between -1.0 to +2.0. The results from these runs are shown in Table 21 and graphed in Figure 19.

These results confirm the observations from changing the individual variances. Increasing all part delivery variances causes all parts to be accelerated, and increases costs by about 4% when all standard deviations are increased by 2 units. Corresponding changes occur when all standard deviations are reduced. The behavior of total policy costs with change in part delivery variances can be seen in Figure 19a.

By contrast, similar changes in processing time variances cause much greater impact on decisions and costs. Increasing processing time variances causes all decisions to be delayed to provide for additional protection against variability. A 2 unit increase in all standard deviations causes a net increase of 27.8% for the normal, and 37.3% and 36.9% for the lognormal and gamma distributions. The impact is considerably higher and grows more rapidly for non-normal distributions, as can be seen in Figure 19b.

Thus, as variances become larger, the choice of distribution becomes more crucial. The impact of buffers and other parameters may also be affected as variability increases. Greater care therefore needs to be exercised in parameter estimation and system design when variabilities are high, especially of processing times.

To summarize, changes in processing time variances have a bigger impact on total costs than that of similar changes in part delivery time variances. At higher variance values, the effect of the change on different distributions is also different,

**Table 21. Effect of Changes in Variances on Total Costs at Solutions**

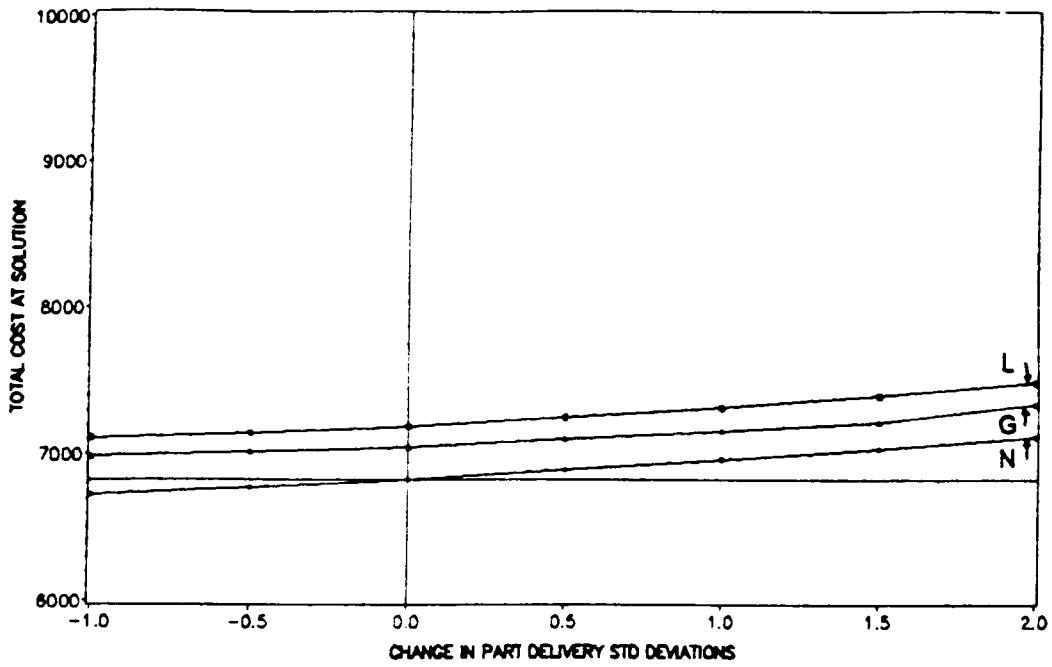
Number of Stations = 8, Number of Jobs = 6  
 Random Delivery & Processing Times, Unlimited Buffers

Parameters Changed	Change	TC Normal	TC Lognormal	TC Gamma
$\sigma D_i, i = 1,..8$	-1.0	6724 (-1.6)	7113 (-1.1)	6986 (-0.9)
$\sigma D_i, i = 1,..8$	-0.5	6772 (-0.8)	7145 (-0.6)	7015 (-0.5)
-	-	6830	7192	7049
$\sigma D_i, i = 1,..8$	+0.5	6894 (0.9)	7250 (0.8)	6994 (0.8)
$\sigma D_i, i = 1,..8$	+1.0	6964 (1.9)	7316 (1.7)	7157 (1.5)
$\sigma D_i, i = 1,..8$	+1.5	7037 (3.0)	7397 (2.8)	7219 (2.4)
$\sigma D_i, i = 1,..8$	+2.0	7115 (4.2)	7485 (4.1)	7341 (4.1)
$\sigma P_i, i = 1,..8$	-1.0	5983 (-12.4)	6135 (-14.7)	6092 (-13.6)
$\sigma P_i, i = 1,..8$	-0.5	6395 (-6.6)	6640 (-6.9)	6578 (-6.7)
-	-	6830	7192	7049
$\sigma P_i, i = 1,..8$	+0.5	7287 (6.7)	7801 (8.5)	7569 (7.4)
$\sigma P_i, i = 1,..8$	+1.0	7755 (13.5)	8454 (17.6)	8223 (16.7)
$\sigma P_i, i = 1,..8$	+1.5	8238 (20.6)	9151 (27.2)	8912 (26.4)
$\sigma P_i, i = 1,..8$	+2.0	8727 (27.8)	9877 (37.3)	9653 (36.9)

(.): % Cost increase with respect to Initial Problem



a) Effect of Changes in All Part Delivery Variances



b) Effect of Changes in All Processing Time Variances

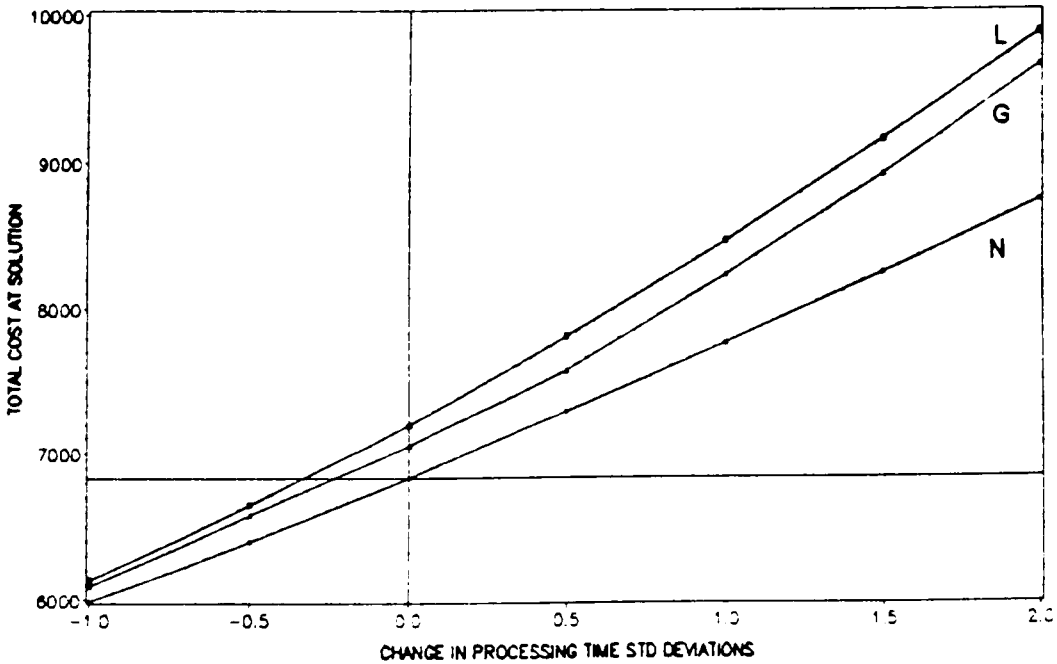


Figure 19. Effect of Changes in Variances on Total Cost in Multiple Job Problem

more in the cases of the lognormal and gamma distributions than for the normal distribution.

### **6.3 Effect of Randomness on Cost Components**

During the course of the development of the prescriptive approach and the analysis so far, the main concern has been the determination of part delivery dates that minimize the overall cost of part and subassembly waiting, finished inventory, makespan and tardiness. For more detailed analysis of system behavior, under different operating conditions, the solutions determined by the methodology are now examined in terms of the individual cost components as well. The cost components considered are as follows.

1. Subassembly inventory cost: cost incurred by incomplete subassemblies waiting for availability of parts and stations.
2. Part inventory cost: cost incurred by parts waiting for availability of subassemblies and stations.
3. Batch earliness cost: cost incurred by completed batch waiting for customer delivery.
4. Batch tardiness cost: cost incurred by completion of batch beyond the customer delivery date.
5. Finished inventory cost: cost incurred by completed assembly jobs waiting for completion of the batch.

6. **Makespan cost:** cost incurred due to the use of the assembly facility for the complete processing of the batch.

The nature of the assembly system as it is modelled certainly causes the occurrence of some costs. Waiting times are incurred by finished assemblies while they wait for the batch to be completed. Subassembly inventory costs are assumed to be increasing down the line, with the highest costs at the finished assembly stage. The solution therefore attempts to delay operations after the bottleneck as much as possible. This reduces the finished inventory cost but causes subassembly waiting costs to be incurred instead. Likewise, makespan cost is incurred even if there is no variability. Therefore some minimum costs would be incurred even if the system were totally deterministic. In the sample problems considered in the experimentation, the total of these minimum costs may be up to 30-60% of the total system cost with randomness, depending on the values of the problem parameters.

Table 22 shows the cost components of the solution for a sample problem with deterministic delivery and processing times. The cost components that are not incurred in a deterministic system are the part inventory cost, and the batch earliness and tardiness costs. This is possible because the timings of the part deliveries and various operations can be determined exactly, thus eliminating these waiting costs.

When random part delivery times are considered, the costs increase significantly, resulting in cost increases of up to 15-30% over the deterministic case in the problems considered. In the example problem shown in Table 22, this increase is about 18%. All the components now have some contribution to the total costs, the largest increases being in the part waiting and finished inventory costs. Though the subas-

**Table 22. Effect of Randomness on System Cost Components at Solutions**

Number of Stations = 8, Number of Jobs = 6

Unlimited Buffers, Normal Distribution

Cost Component	Deterministic Times	Random Dely. Times	Random Dely. & Proc. Times
Subassy. Inventory Cost	705.0	657.0	1385.3
Part Inventory Cost	0.0	239.6	256.1
Batch Earliness Cost	0.0	80.2	415.7
Batch Tardiness Cost	0.0	117.7	341.3
Finished Inventory Cost	2100.0	2414.7	3018.1
Makespan Cost	1185.0	1215.7	1413.0
Total Cost	3990.0	4725.0	6829.7

sembly cost at the final solution is marginally lower, the reduction is considerably smaller than the increases in any of the other components.

Adding randomness in processing times in addition to part delivery times increases solution costs considerably, resulting in overall cost increases of 50-100% over the deterministic system costs in the problems considered. In the sample problem, this increase is about 70%. The effect of random processing times is much larger than that due to random delivery times, and the largest increases occur in subassembly and finished inventory costs. All components, with the exception of part inventory costs, register significant increases.

Based on the cost values presented in Table 22, and the results of other scenarios to be discussed later, the most significant components of the total cost are those relating to finished inventory, subassembly waiting and makespan. Since waiting times are enforced on finished assemblies, and the per unit inventory costs are the highest at this stage, the finished inventory component is the largest contributor in all problems. Also, as will be indicated later, the largest magnitudes of changes, under different scenarios, take place in the finished inventory and subassembly waiting costs.

Decisions derived using deterministic estimates are optimistic compared to those considering randomness, since no safety times are allowed for delivery or processing. In the sample problems used in this chapter, if the deterministic decisions are used when actual delivery times are random, errors of about 20-40% result with respect to optimal policy costs. This error increases to as much as 60-100% when random processing times are introduced. It is therefore important to consider randomness explicitly in making these decisions.

## **6.4 Effect of Buffer Capacities**

Provisions have been made in the descriptive and prescriptive models for consideration of limited buffers between assembly stations. These restrictions may arise due to the cost of providing for buffer positions, as well as due to space and other physical constraints. Though the model cannot prescribe the optimal buffer capacities, some experimentation has been conducted to measure the impact of buffer limitations and guide their size and placement.

There are increases in total costs when buffer capacities are restricted instead of being unlimited. The total increase in cost from unlimited buffers to the buffer capacities at 0, ranges between 2-4% when the sample problems used in this chapter are run with deterministic processing times. When processing times are random, the likelihood of blocking and starving increases. This causes additional cost increase, resulting in net increases of 3-8% in the sample problems due to buffer limitations. The actual magnitudes of these increases depend on the relative imbalance between stations and on the variability of delivery and processing times.

The sample problem used in the previous section has been run under three different buffer configurations. The cost increase in the solution arising out of eliminating buffers completely is 3.2%. Examining the cost components at the solution, shown in Table 23, reveals that the most significant cost increase occurs in the finished inventory costs, while the subassembly cost declines. This behavior occurs because now subassemblies cannot be held back as much as with unlimited buffers, and jobs have to be kept moving through the system to reduce interference. The

consequent increase in finished inventory cost offsets the reduction in subassembly cost. All other cost components also increase, though not as significantly.

The appropriate locations and sizes of these buffers depend on the different means and variabilities, and also on the relative imbalance between stations. However, in small lot assembly problems with unbalanced lines, the role of buffer capacities is less significant than in balanced lines. This is because in unbalanced lines, the faster operations tend to act as buffers themselves. In such lines, the most important buffer positions are the ones immediately preceding and following the relative bottlenecks, to prevent blocking and starving of the critical operations. In addition to the main bottleneck, the relative bottlenecks down the line are also important, since the amount by which jobs can be delayed depends on their durations.

As can be seen in Table 23, provision of just a few buffer positions at the relative bottlenecks results in total costs within about 1% of the costs achieved by the unlimited buffer solution for that problem. This behavior can also be seen in the results for problems 2, 5, 8, 11, 14, 17 and 18 in Table 16 in Chapter 5, and is consistent with the findings in Conway et al. (1988) in their study of the role of buffer capacities in serial production lines. The choice of appropriate locations and capacities of buffers is also governed by other factors, such as the cost of providing buffer positions, and space or other physical constraints. This methodology can be used to enable evaluation of these trade-offs.

Since the subassembly and finished inventory costs are large components of the overall cost, and these are also the values most affected by buffer limitations, changes in their unit costs will affect the cost increase due to buffers. An increase in finished inventory unit cost will cause the cost increase due to buffer limitations to

**Table 23. Effect of Buffers on System Cost Components at Solutions**

Number of Stations = 8, Number of Jobs = 6

Random Delivery & Processing Times, Normal Distribution

Cost Component	Unlimited Buffers	Buffer Caps. 1-0-1-1-0-1-0	All 0 Buffers
Subassy. Inventory Cost	1385.3	1027.2	1017.6
Part Inventory Cost	256.1	264.3	280.1
Batch Earliness Cost	415.7	425.7	434.5
Batch Tardiness Cost	341.3	337.6	345.6
Finished Inventory Cost	3018.1	3426.0	3536.2
Makespan Cost	1413.0	1416.9	1431.7
Total Cost	6829.7	6897.9	7045.8



be even greater. However, part inventory costs are not likely to cause any significant change in the cost increase due to buffers. Changes in variances will affect the cost increase of buffer restrictions, since the probability of blocking and starving increases as variability increases. Therefore, the cost increases are likely to be higher as the variances increase. Again, processing time variances are likely to cause greater effect than part delivery variances as was seen in Section 6.2.

To summarize, buffer restrictions cause increases in cost compared to unlimited buffers. This increase is mainly due to increase in finished inventory which offsets the reduction in subassembly cost. In small lot assembly with unbalanced lines, the role of buffers is not very critical, and provision of just a few buffer spaces results in total costs close to the unlimited buffer solutions.

## **6.5 Effect of Distributions**

The lognormal and gamma distributions consistently yield higher overall costs at the final solutions than the normal distribution. This is a consequence of the skewed nature of the distributions. This behavior becomes more significant when the variances become large, since that is when the differences between the distributions are the most pronounced. In the sample problems considered, the cost increase ranges between 4-10%, depending on the values of the parameters. The decisions and optimal policy costs are generally the most different for the lognormal. This behavior was observed in the solution of the single job problem as well, as discussed in section 4.7.1.

**Table 24. Effect of Distributions on System Cost Components at Solutions**

Number of Stations = 8, Number of Jobs = 6

Random Delivery & Processing Times, Unlimited Buffers

Cost Component	Normal	Lognormal	Gamma
Subassy. Inventory Cost	1385.3	1591.5	1576.1
Part Inventory Cost	256.1	252.4	284.1
Batch Earliness Cost	415.7	331.9	349.2
Batch Tardiness Cost	341.3	469.9	352.1
Finished Inventory Cost	3018.1	3096.4	3067.3
Makespan Cost	1413.0	1449.6	1420.4
Total Cost	6829.7	7192.1	7049.2

In the sample problem shown in Table 24, the cost increase is 5.3% for the lognormal and 3.2% for the gamma. The most significant increases cost occur in the finished inventory and subassembly costs. The makespan values are also larger for non-normal distributions. The batch tardiness cost increases as well, but there is a corresponding decrease in the batch earliness cost. The change in part inventory cost is relatively less significant, and in any case is not a major part of the overall cost.

As seen in Section 6.2, the effect of changes in cost parameters affect distributions similarly. Therefore, the trends in behavior observed here will continue to hold if the cost parameters change. However, the effect of variability was seen to be higher for non-normal distributions. Accordingly, the effects due to distributions observed in this section are likely to be more pronounced as the variances increase.

### **6.5.1 Relationship Between Buffer Capacities and Distributions**

It has also been investigated if the distribution used has any relation with the effect due to buffers, that is, if the cost increases due to buffer limitations are different for the different distributions considered. In order to study this, a set of problems of different sizes have been run in the extreme cases of unlimited and zero buffers for each distribution. The costs at these solutions are shown in Table 25. For each distribution, the costs are shown along with the percentage increase due to buffer limitations.

For the 5 station  $\times$  5 job problem in Table 25, the total cost increase due to buffer limitations using the normal distribution is 3.7%. The corresponding increase when

**Table 25. Effect of Distributions and Buffer Capacities on Total Costs at Solutions**

Random Delivery & Processing Times, Moderately Balanced Line

Stns. X Jobs	Buffers	$TC^*$ Normal	$TC^*$ Lognormal	$TC^*$ Gamma
5 X 5	Unlim.	2341	2420	2390
5 X 5	All 0	2428 (3.7)	2519 (4.1)	2498 (4.5)
5 X 10	Unlim.	6763	6928	6799
5 X 10	All 0	7299 (7.9)	7577 (9.3)	7457 (9.6)
8 X 6	Unlim.	6829	7192	7049
8 X 6	All 0	7046 (3.2)	7495 (4.2)	7327 (3.9)
10 X 10	Unlim.	21767	22298	22263
10 X 10	All 0	22973 (5.5)	23874 (7.0)	23806 (6.9)

(.): % Cost increase with respect to Unlimited Buffer Solution

the lognormal distribution is used is 4.1%. Thus the incremental increase due to the joint effect of buffers and the lognormal distribution is 0.4%. The corresponding value for the gamma distribution for the same problem is 0.8%. Another method of examining this effect is to evaluate the total cost increase from the Normal-Unlimited Buffer problem to the Lognormal-Zero Buffer problem, which is 7.6% in this case. The increase due to buffer restrictions alone for the normal distribution is 3.7%, while the increase due to distribution alone is 3.4%. The residual cost increase of 0.5% ( $7.6 - 3.7 - 3.4$ ) is another measure of the joint effect of buffers and distributions.

Similar analysis of the results for the other problems in Table 25 shows that the cost increases due to buffer limitations are marginally higher for non-normal distributions, ranging between 0.4 to 1.7%. Thus the non-normal distributions may need slightly higher buffer capacities than the normal, to achieve the same percentage increase in overall cost. However, these additional cost increases are relatively smaller than the individual increases due to buffers or distributions by themselves.

The increase in cost due to buffers is also a function of the relative magnitudes of the processing times at the stations. In the problems considered so far, operation times have been moderately balanced. While the bottleneck station processing time has been kept at 10 units, the smallest operation times have been generated between 7 and 8 units. The effect of buffers is likely to be less pronounced as the stations become more unbalanced.

In order to verify this behavior, another set of problems of similar size have been run with a highly unbalanced line, with the processing time values ranging from 2 to 10 units. These results reveal that the impact of buffer limitations is lower when less balanced lines are considered. This behavior results due to the lower likelihood of

interference between operations. The incremental increase due to the joint effect of buffers and distributions is also correspondingly lower. The results for these problems are shown in Table 26.

Although corresponding trends hold as the lines become more balanced, the effect appears to level out as the lines become increasingly balanced. Experimentation showed that the impact of buffers in highly balanced lines, with all processing times ranging between 9.5 and 10 units, were of the same order as in moderately balanced systems. This behavior occurs since the random variability of part delivery damps some of the interference between stations. In any case, highly balanced lines may not be very likely in small lot assembly.

As discussed in Section 6.4, changes in finished inventory and subassembly unit costs are likely to affect the cost increases due to buffers limitations. Also, changes in processing time variabilities affect the cost changes due to distributions as well as buffers. Thus, the magnitude of the joint effect due to buffers and distributions will also be greater. However, the relative magnitude of this joint effect compared to the individual effects of buffer limitations or distributions will follow the same trends as observed in this section.

### **6.5.2 Choice of Distribution**

During the study of the single job problem, it had been found that the errors resulting out of the choice of an incorrect distribution were relatively small. In other words, if decisions are determined using one distribution, when in fact the real process follows another, the errors in total cost with respect to the optimal solution for

**Table 26. Effect of Distributions and Buffer Capacities on Total Costs at Solutions**

Random Delivery & Processing Times, Unbalanced Line

Stns. X Jobs	Buffers	$TC^*$ Normal	$TC^*$ Lognormal	$TC^*$ Gamma
5 X 5	Unlim.	2094	2167	2130
5 X 5	All 0	2135 (2.0)	2215 (2.2)	2170 (1.8)
5 X 10	Unlim.	6269	6416	6315
5 X 10	All 0	6630 (5.8)	6829 (6.4)	6709 (6.2)
8 X 6	Unlim.	5700	6136	5878
8 X 6	All 0	5824 (2.2)	6347 (3.4)	6037 (2.7)
10 X 10	Unlim.	18906	19812	19650
10 X 10	All 0	20124 (6.4)	21112 (6.6)	20877 (6.2)

(.): % Cost increase with respect to Unlimited Buffer Solution

that distribution are small. Tests have been conducted to verify this behavior for the multiple job problem as well.

To measure this error, comparisons have been done using the normal and lognormal distributions, since the results for these two distributions are the most different. The decisions taken using the normal distribution are evaluated using the cost computation routine for the lognormal, and vice versa. The values thus obtained are compared with the actual solution values for that distribution. Table 27 shows the results from these experiments for the same problems as considered earlier.

The errors determined from these experiments, as shown in the table, show that the impact on overall cost due to the use of an incorrect distribution is relatively small, ranging between 0.2 to 2.5% for the problems considered. However, the decisions obtained are found to be quite different (not shown here), so that the actual behavior of the system may be different for different distributions. Also, the estimates of individual cost components and total cost will be different. Use of the normal distribution results in optimistic estimates of system behavior and cost compared to the lognormal. However, the overall cost will not be substantially in error, compared to the solution which would have been obtained had the correct distribution been used. This behavior may be the result of the existence of a relatively large region of indifference in the solution space, as was observed in the solution of the single job problem as well.

Since changes in costs affect different distributions in a similar manner, these observations regarding choice of distribution hold even if there are changes in the cost parameters. Though changes in variances affect the distributions considered differently, the decisions for all distributions react in a similar manner to changes in



**Table 27. Errors Due to Choice of Distribution**

Random Delivery & Processing Times

Stns. X Jobs	Buffers	TC <sup>*</sup> Normal	TC L-EN	TC <sup>*</sup> Lognormal	TC N-EL
5 X 5	Unlim.	2341	2357 (0.7)	2420	2425 (0.2)
5 X 5	All 0	2428	2453 (1.0)	2519	2536 (0.7)
5 X 10	Unlim.	6763	6782 (0.3)	6928	6952 (0.3)
5 X 10	All 0	7299	7420 (1.6)	7577	7677 (1.3)
8 X 6	Unlim.	6829	6870 (0.6)	7192	7212 (0.3)
8 X 6	All 0	7046	7147 (1.4)	7495	7605 (1.5)
10 X 10	Unlim.	21767	22326 (2.6)	22298	22846 (2.5)
10 X 10	All 0	22973	23328 (1.5)	23874	24292 (1.7)

L-EN: Lognormal solution evaluated using normal distribution

N-EL: Normal solution evaluated using lognormal distribution

(.): % Error due to choice of incorrect distribution

parameters. Considering the nature of the solution space, the observations regarding errors in total cost due to choice of distribution are not affected significantly by these changes. This behavior has been found to hold in all the problems and variations considered in this chapter.

## **6.6 Summary**

In this chapter the behavior of small lot assembly systems has been examined to assess the importance of various aspects of system design. The components of the objective function have been examined to determine their relative importance and understand system behavior. The effects of randomness, buffer restrictions, and choice of distribution have been studied. The sensitivity of the the solution to changes in various parameters has been analyzed. The relative importance of various factors and parameters has been compared, using numerical values, to identify trends and patterns in solution behavior that may be useful in the design and operation of such systems.

## **7.0 Conclusion**

### **7.1 Summary and Achievements of Research**

In this research the problem of determining the optimal part delivery dates in stochastic small lot assembly systems has been addressed. The delivery times of parts to stations and the processing times at assembly stations have been modelled as random variables. The objective of the model is to determine the mean due dates for parts in order to minimize the total expected inventory and tardiness cost. The model is based on the descriptive model developed by Wilhelm and Ahmadi-Marandi (1982), and its extensions.

The descriptive model used to represent and describe the behavior of small lot assembly systems has been presented. Three distributions, namely normal, lognormal and gamma, are considered as representations of part delivery and processing times. The model is based on taking the maximum of several random variables. Necessary measures of system performance have been identified and the derivation of the objective function from the descriptive model has been described. The relationships for the moments of the maximum of two random variables for the normal and lognormal distributions have been presented, while similar relations for the gamma distribution have been derived. The properties required for the application of these relationships to the assembly system have been identified, and studies

have been conducted to measure the suitability of different distributions in light of these requirements.

The single job assembly problem is concerned with determining optimal part delivery dates when only one job is produced by the system, or when the entire lot is produced as one job without distinguishing between completions of individual jobs. The nonlinear programming method has been used to determine the optimal solutions to this problem, and it consistently yields solutions far superior compared to the methods used in previous research. This method is also general enough to incorporate non-normal distributions and random processing times. An approximate solution technique based on the optimization of individual stations in isolation, with suitable correction factors, has been developed. The approach has been generalized to include random processing times and the different distributions considered. This method yields results comparable in quality to the optimal solutions with much less computational effort.

In the multiple job problem, a set of identical jobs is assembled in sequence, flowing through the assembly stations in the same order, and delivered to the customer as a batch. A representation of the system as a precedence network is used to understand the process and its description. Modifications necessary to incorporate limited buffers between stations are made in the descriptive model. The nonlinear programming approach has been applied to obtain optimal solutions to the multiple job problem, whereas previous research had been confined to heuristic solutions only. A heuristic procedure, based on solution of a sequence of single station problems, has been developed with modifications necessary to incorporate random

processing times and limited buffers. The heuristic yields solutions close to the optimal solutions with considerably less computational time.

The main achievement of this research has been the incorporation of three major levels of generalization compared to previous research. These are:

1. Use of different distributions as representations of part delivery times in small lot assembly.
2. Consideration of random processing times at assembly stations.
3. Consideration of limited buffers between stations in multiple job assembly systems.

In addition to these generalizations, optimal solutions are now obtained for the single job as well as multiple job problems, which had not been achieved in previous research. To enable quick and easy solution to the problems, approximate/heuristic methods have also been developed, and have been found to yield results comparable in quality to the optimal solutions.

## **7.2 Major Findings of This Research**

During the course of the research, insights have been generated into the problem and optimal policy behavior under different operating conditions. Some of these are summarized below.

1. Multi-stage processes can be analyzed using the methodology developed that considers individual stages in isolation, incorporating appropriate correction factors to capture the effects of variability and relative costs.
2. For the multiple job case, the effect of correlations between finish times of various operations appears to become insignificant, since the single stage solution based approach does not consider these correlations explicitly, but still yields almost optimal solutions.
3. The effect of not using the correct probability distribution for part delivery and operation times becomes significant at higher variabilities. The lognormal and gamma distributions result in higher total costs than the normal distribution. This difference is largely due to increased subassembly and finished assembly waiting costs. The buffer space requirements for these distributions are marginally larger than those required for the normal distribution in order to achieve similar relative cost values, comprising of subassembly and part waiting costs, makespan cost and due date penalty costs.
4. For all practical purposes, the normal distribution can be used to determine part delivery dates. This follows from the fact that when the part delivery dates determined using the normal distribution are used in the total cost expressions for the lognormal or gamma distributions, the value obtained is not very different from the total cost obtained using the delivery dates determined using the lognormal or gamma distribution as the case may be.
5. Variability is the most important factor in the design of the system. The choice of distribution, the effect of buffers, as well as the decisions and total costs are significantly affected by the estimates of variability. The effect of processing time variability is relatively more significant than that due to part delivery time variability.

6. System performance is affected by buffer limitations, and total costs increase due to the increased inventory costs incurred by finished jobs waiting for the completion of the batch. However, in small lot assembly with unbalanced lines, these buffers need not be very large, and provision of a few buffer positions at the bottlenecks enables achieving total costs close to the costs obtained using unlimited buffers. This verifies the results reported in Conway et al. (1988) in their investigation of the role of work-in-process inventory in serial production lines.

### **7.3 Avenues for Future Research**

The avenues for future research in this area pertain to enhancing the modelling scenario considered in this research, generating a better description of the process, as well as to extend the ideas of the approach to applications in other areas. Some of these ideas are outlined below.

1. The model as developed considers one vendor supplied part input at each station. If more than one part is required at a station, there is an accumulation process prior to start of the assembly operation. This feature can be easily incorporated within the same framework by introducing a dummy station for each additional part. Some of the parts or subassemblies may be supplied by branch lines sharing a common facility instead of independent vendors, and may therefore be subject to the same operation start time restrictions as the main assembly. This will necessitate calculation of correlations even for input parts delivery times, while so far they have been assumed to be independent. The descriptive

and prescriptive models have to be modified to consider this feature. Adding this feature will enable application of the methodology to general assembly networks for multi-level product structures.

2. It has been assumed that the order in which operations are to be performed is predetermined. However, it is possible that there may be some flexibility available in determining the sequence of operations along the line. Approaches need to be developed to enable determination of the placement of operations on the line considering their relative durations, variabilities and costs.
3. In the model developed, it has been assumed that the batch to be produced consists of identical jobs. If non-identical jobs are to be considered, each with different processing times, a new method has to be found to identify the bottleneck station and the critical path. Also, if the sequence of production of the jobs through the facility is not specified, job sequencing becomes another decision variable. This will considerably increase the complexity of the prescriptive model, especially if setup costs are included.
4. The part delivery and processing times have been assumed to follow distributions that are unimodal in nature. This enables modelling of random variations in task performance times and small breakdowns at the stations. However, if major breakdowns are possible, then a multi-modal distribution may better describe the variability in performance times. The descriptive and prescriptive models need to be modified to incorporate this type of distribution. The properties necessary for the development of the descriptive model may not be satisfied for such a distribution, necessitating the use of numerical approximations to de-



termine measures of system performance. Also, the objective function may not retain the properties necessary for the use of nonlinear programming methods. It may therefore be even more difficult to be certain of the quality of solutions obtained.

5. In this research, the intent has been to develop a prescriptive model for specifying due dates. The modelling of the queueing processes and transient behavior of the system under different scenarios has not been the concern here. For a more complete description and understanding of the process, approaches need to be developed to describe and study the system from this viewpoint.
6. Scheduling algorithms for PERT networks typically consider early and late completion costs associated with the entire project. However, these are inadequate if storable resources with inventory costs are considered, and if there are costs associated with early and late completion of individual activities. In view of the resemblance in structure between PERT networks and the problem considered in this research, a methodology for scheduling stochastic PERT networks with these features may be developed using some ideas from this research.

## References

1. Adam, N.R., Bertrand, J.W.M., and Surkis, J., 'Priority Assignment Procedures in Multi-level Assembly Job Shops', Times and Finite Buffers, IIE Transactions, (1987), Vol.19, No.3, 317-328.
2. Altioek, T., 'Production Lines with Phase-type Operation and Repair Times and Finite Buffers', International Journal of Production Research, (1985), Vol.23, No.3, 489-498.
3. Altioek, T., and Stidham, 'The Allocation of Interstage Buffer Capacities in Production Lines', IIE Transactions, (1983), Vol.15, No.4, 292-299.
4. Aquilano, N.J., and Smith, D.E., 'A Formal Set of Algorithms for Project Scheduling with Critical Path Method - Material Requirements Planning', Journal of Operations Management, (1980), Vol.1, No.2, pp.57-67.
5. Avriel, M., Nonlinear Programming: Analysis and Methods, (1976), Prentice Hall, Englewood Cliffs, New Jersey.
6. Bigham, P.E., and Mogg, J.M., 'Integer Multiple Production Schedules for Multi-stage Systems with Stochastic Lead Times', International Journal of Production Research, (1981), Vol.19, No.1, pp.97-107.
7. Blumenfeld, D.E., 'Performance Characteristics of Assembly Systems with Fixed and Variable Processing Times', General Motors Research Laboratories Research Publication GMR-5912, (1987).
8. Bury, K.V., Statistical Models in Applied Science, (1975), John Wiley, New York.

9. Buzacott, J.A., 'Automatic Transfer Lines with Buffer Stocks', International Journal of Production Research, (1967), Vol.5, No.1, pp.183-200.
10. Clark, C.E., 'The Greatest of a Finite Set of Random Variables', Operations Research, (1961), Vol.9, No.2, pp.145-162.
11. Conway, R., Maxwell, W., McClain, J.O., and Thomas, L.J., 'The Role of Work-in-process Inventory in Serial Production Lines', Operations Research, (1988), Vol.36, No.2, pp.229-241.
12. Das, S.K., Determination of Optimal Part Delivery Dates in a Stochastic Assembly System, Unpublished M.S. Thesis (1985), Virginia Polytechnic Institute and State University, Blacksburg, VA.
13. Das, S.K., and Sarin, S.C., 'Selection of a Set of Part Delivery Dates in a Multi-Job Stochastic Assembly System', IIE Transactions (1988), Vol.20, No.1, pp.4-11.
14. Davis, E.W., 'Project Scheduling Under Resource Constraints - Historical Review and Categorization of Procedures', AIIE Transactions, (1973), Vol.5, No.4, pp.297-313.
15. Davis, E.W., and Patterson, J.H., 'A Comparison of Heuristic and Optimum Solutions in Resource Constrained Project Scheduling', Management Science, (1975), Vol.21, No.8, pp.944-955.
16. Dodin, B., 'Determining the K Most Critical Paths in PERT Networks', Operations Research, (1984), Vol.32, No.4, pp.860-877.
17. Dodin, B., 'Bounding the Project Completion Time Distribution in PERT Networks', Operations Research, (1985), Vol.33, No.4, pp.862-881.
18. Elmaghraby, S.E., Activity Networks, (1977), John Wiley, New York.
19. Ensore, E.E., and Niebel, B.W., 'Reliable Indirect Standards Achieved Through Computerized Computation of Mean Slot Values of the Gamma Distribution',

Proceedings of the Third National Conference on Computers and Industrial Engineering, (1980), Orlando, U.S.A.

20. Ensore, E.E., Knott, K., and Niebel, B.W., 'A Comparison of Alternate Time Slotting Methods for Indirect Time Standards', Proceedings of the Annual Spring IIE Conference, (1982), New Orleans, U.S.A.
21. Fry, T.D., Oliff, M.D., Minor, E.D., and Leong, G.K., 'The Effects of Product Structure and Sequencing Rule on Assembly Shop Performance', Working Paper, University of South Carolina, 1986.
22. Fry, T.D., Philipoom, P.R., and Markland, R.E., 'Dispatching in a Multistage Job Shop where Machine Capacities are Unbalanced', International Journal of Production Research, (1988), Vol.26, No.7, pp.1193-1223.
23. Fry, T.D., Philipoom, P.R., and Markland, R.E., 'Due Date Assignment in a Multistage Job Shop', Working Paper, University of South Carolina, 1987.
24. Goodwin, J.S., and Goodwin, J.C., 'Operating Policies for Scheduling Assembled Products', Decision Sciences, (1982), Vol.13, pp.585-603.
25. Goodwin, J.S., and Weeks, J.K., 'Evaluating Scheduling Policies in a Multi-level Assembly System', International Journal of Production Research, (1986), Vol.24, No.2, pp.247-257.
26. Grasso, E.T., and Taylor, B.W. III, 'A Simulation-Based Experimental Investigation of Supply/Timing Uncertainty in MRP Systems', International Journal of Production Research, (1984), Vol.22, No.3, pp.253-264.
27. Harrison, J.M., 'Assembly-Like Queues', Journal of Applied Probability (1973), Vol.10, pp.354-367.
28. Hillier F.S., and Boling, R.W., 'Finite Queues in Series with Exponential or Erlang Service Times - A Numerical Approach', Operations Research, (1967), Vol.15, pp.286-33.

29. Himmelblau, D.M., Applied Nonlinear Programming, (1972), McGraw Hill, New York.
30. Ho, Y.C., 'A Survey of the Perturbation Analysis of Discrete Event Dynamic Systems', Annals of Operations Research **3**, Flexible Manufacturing Systems: Operations Research Models and Applications, (1985), Baltzer, p.393.
31. Ho, Y.C., Suri, R., Cao, X.R., Diehl, G.W., Dille, J.W., and Zazanis, M., 'Optimization of Large Multiclass (non-product-form) Queueing Networks Using Perturbation Analysis', Large Scale Systems Journal, (1984), Vol. 7, p.165.
32. Hooke, R., and Jeeves, T.A., 'Direct Search Solution of Numerical and Statistical Problems', Journal of the Association of Computing Machinery, (1961), Vol.8, pp. 212-229.
33. Huang, P.Y., 'A Comparative Study of Priority Dispatching Rules in a Hybrid Assembly/Job Shop', International Journal of Production Research, (1961), Vol.22, No.3, pp. 375-387.
34. Jensen, P.A., and Khan, M.A., 'Scheduling of a Multi-Stage Production System with Setup and Inventory Costs', AIIE Transactions (1972), Vol.4, No.2, pp.126-133.
35. Johnson, N.L., and Kotz, S., Distributions in Statistics - Continuous Univariate Distributions I, (1972a), John Wiley, New York.
36. Johnson, N.L., and Kotz, S., Distributions in Statistics - Continuous Multivariate Distributions, (1972b), John Wiley, New York.
37. Kanet, J.J., 'Towards Understanding Lead Times in MRP Systems', Production and Inventory Management (1982), Vol.23. No.3, pp.1-14.
38. Kanet, J.J., and Christy, D.P., 'Manufacturing Systems with Forbidden Early Departure', International Journal of Production Research, (1984), Vol.22. No.1, pp.41-50.

39. Knott, K., and Sury, R.J., 'A Study of Work Time Distributions in Unpaced Assembly Tasks', IIE Transactions (1987), Vol.19, No.1, pp.50-57.
40. Knott, K., Chandra, J., and Ensore, E.E., 'Time Slotting Based upon an Assumed Lognormal Distribution of Sample Work Task Times', International Journal of Production Research, (1987), Vol.25, No.4, pp.487-512.
41. Koenigsberg, E., 'Production Lines and Internal Storage - A Review', Management Science, (1959), Vol.5, No.2, pp.419-433.
42. Kraemer, S.A., and Love, R.F., 'A Model for Optimizing the Buffer Inventory Storage Time in a Sequential Production System', IIE Transactions, (1970), Vol.2, No.1, pp.64-69.
43. Kraft, C.H., and Eeden, C., A Non-Parametric Introduction to Statistics (1968), Macmillan, New York.
44. Kuester, J.L., and Mize, J.H., Optimization Techniques with FORTRAN (1973), McGraw Hill, New York.
45. Law, A.M., and Kelton, W.D., Simulation Modeling and Analysis, (1982), McGraw Hill, New York.
46. Lee, H.L., and Rosenblatt, M.J., 'Effects of In Process Inventory on an Assembly Line', Toward the Factory of The Future, Proceedings of the Eighth International Conference of Production Research, Stuttgart, (1985), pp. 784-788.
47. Maxwell, W.L., and Mehra, M., 'Multiple-Factor Rules for Sequencing with Assembly Constraints', Naval Research Logistics Quarterly, (1968), Vol.15, pp. 241-254.
48. Melnyk, S.A., and Piper, C.J., 'Leadtime Errors in MRP: The Lot Sizing Effect', International Journal of Production Research, (1985), Vol.22, No.3, pp.485-497.

49. Muth, E.J., 'Stochastic Processes and Their Network Representations Associated with a Production Line Queueing Model', European Journal of Operational Research, (1984), Vol.15, No.1, pp.63-83.
50. Orlicky, J.A., Material Requirements Planning, (1975), McGraw-Hill, New York.
51. Philipoom, P.R., Rees, L.P., Taylor, B.W. III, and Huang, P.Y., 'An Investigation of the Factors Influencing the Number of Kanbans Required in the Implementation of the JIT Technique With Kanbans', International Journal of Production Research, (1987), Vol.25, No.3, pp.457-472.
52. Powell, M.J.D., 'An Efficient Method for Finding the Minimum of a Function of Several Variables Without Using Derivatives', Computer Journal, (1964), Vol. 7, pp. 155-162.
53. Rao, N.P., 'On the Mean Production Rate of a Two-Stage Production System of the Tandem Type', International Journal of Production Research, (1975a), Vol.13, No.2, pp.207-217.
54. Rao, N.P., 'Two-Stage Production Systems with Intermediate Storage', AIIE Transactions, (1975b), Vol.7, No.4, pp.414-421.
55. Rees, L.P., Philipoom, P.R., Taylor, B.W. III, and Huang, P.Y., 'Dynamically Adjusting the Number of Kanbans in a Just-In-Time Production System Using Estimated Values of Leadtime', IIE Transactions, (1987), Vol.19, No.2, pp.199-207.
56. Russell, R.S., and Taylor, B.W. III, 'An Evaluation of Scheduling Policies in a Dual Resource Constrained Assembly Job Shop', IIE Transactions, (1985), Vol.17, No.2, pp.219-232.
57. Russell, R.S., An Evaluation of Scheduling Policies in a Dual Resource Constrained Assembly Job Shop, Ph.D. Dissertation (1983), Virginia Polytechnic Institute and State University, Blacksburg, VA.

58. Saboo, S., and Wilhelm, W.E., 'An Approach for Modeling Small-lot Assembly Networks', IIE Transactions, (1986), Vol.18, No.4, pp.322-334.
59. Sarin, S.C., and Das, S.K., 'Determination of Optimal Part Delivery Dates in a Stochastic Assembly Line', International Journal of Production Research, (1987), Vol.25, No.7, pp.1013-1028.
60. Sculli, D., 'The Completion Time of PERT Networks', Journal of the Operational Research Society, (1983), Vol.34, No.2, pp.155-158.
61. Sculli, D., 'Priority Dispatching Rules in Job Shops with Assembly Operations and Random Delays', Omega, (1980), Vol.8, No.2, pp.227-234.
62. Sculli, D., and Wong, K.L., 'The Maximum and Sum of Two Beta Variables and the Analysis of PERT Networks', Omega, (1985), Vol.13, No.3, pp.233-240.
63. Seidmann, A., Schweitzer, P.J., and Nof, S.Y., 'Observations on The Normality of Batch Production Times in Flexible Manufacturing Cells', International Journal of Production Research (1987), Vol.25, No.1, pp.151-154.
64. Slater, L.J., Generalized Hypergeometric Functions, (1966), Cambridge University Press.
65. Smith, J.M., and Daskalaki, S., 'Buffer Space Allocation in Automated Assembly Production Lines', Operations Research, (1988), Vol.36, No.2, pp.343-358.
66. Smith-Daniels, D.E., and Aquilano, N.J., 'Constrained Resource Project Scheduling Subject to Material Constraints', Journal of Operations Management, (1984), Vol.4, No.4, pp.369-387.
67. Smith-Daniels, D.E., and Smith-Daniels, V.L., 'Optimal Project Scheduling with Materials Ordering', IIE Transactions, (1987), Vol.19, No.2, pp.122-129.
68. St.John, R., 'The Cost of Inflated Planned Lead Times in MRP Systems', Journal of Operations Management, (1985), Vol.5, No.2, pp.119-128.



69. Suri, R., 'An Overview of Evaluative Models for Flexible Manufacturing Systems', Annals of Operations Research 3, Flexible Manufacturing Systems: Operations Research Models and Applications, (1985), Baltzer, pp.13-21.
70. Tippett, L.H.C., 'On the Extreme Individuals and the Range of Samples Taken from a Normal Population', Biometrika, (1925), Vol.17, pp.364-387.
71. Trilling, D.R., 'Job Shop Simulation of Orders that are Networks', Journal of Industrial Engineering, (1966), Vol.17, No.2, pp.59-71.
72. Walker, D.M., and Wysk, R.A., 'A Procedure To Determine The Least Cost Purchased-Part Planned Lead-Time for Items Controlled By an MRP System', Computers and Industrial Engineering (1983), Vol.7, No.4 pp.295-308.
73. Weeks, J.K., 'Optimizing Planned Lead Times and Delivery Dates', Proceedings American Production and Inventory Control Society Annual Conference, (1981), pp.177-188.
74. Whybark, D.C., and Williams, J.G., 'Material Requirements Planning Under Uncertainty', Decision Sciences (1976), Vol.7, No.4, pp.595-600.
75. Wilhelm, W.E., 'A Model to Approximate Transient Performance of the Flowshop', International Journal of Production Research (1986a), Vol.24, No.1, pp.33-50.
76. Wilhelm, W.E., 'An Approach for Modeling Transient Material Flows in Robotized Manufacturing Cells', Material Flow, (1986b), Vol.3, No.1, pp.55-68.
77. Wilhelm, W.E., 'The Application of Lognormal Models of Transient Operations in the Flexible Manufacturing Environment', Journal of Manufacturing Systems (1986c), Vol.5, No.4, pp.253-266.
78. Wilhelm, W.E., 'S<sub>n</sub> Models of The Maximum of Several Dependent Random Variables: Development And Computer Application', Computers and Industrial Engineering (1986d), Vol.10, No.4 pp.335-346.

79. Wilhelm, W.E., 'On the Normality of Operation Times in Small-lot Assembly Systems: A Technical Note', International Journal of Production Research (1987), Vol.25, No.1, pp.145-149.
80. Wilhelm, W.E., and Ahmadi-Marandi, S., 'A Methodology to Describe Operating Characteristics of Assembly Systems', IIE Transactions, (1982), Vol.14, No.3, pp.204-213.
81. Wilhelm, W.E., Saboo, S., and Johnson, R.A., 'An Approach for Designing and Managing Material Flow in Small-Lot Assembly Lines', Journal of Manufacturing Systems, (1986), Vol.5, No.3, pp.147-160.
82. Wilhelm, W.E., and Wang, L., 'Management of Component Accumulation in Small Lot Assembly Systems', Journal of Manufacturing Systems, (1986), Vol.5, No.1, pp.27-39.
83. Winkler, R.L., Roodman, G.M., and Britney, R.R., 'The Determination of Partial Moments', Management Science, (1972), Vol.19, No.3, pp.290-296.
84. Yamashina, H., and Okamura, K., 'Analysis of In-process Buffers for Multi-Stage Transfer Line Systems', International Journal of Production Research (1983), Vol.21, No.2, pp.183-195.
85. Yano, C.A., 'Setting Planned Leadtimes in Serial Production Systems with Tardiness Costs', Management Science, (1987a), Vol.33, No.1, pp.95-106.
86. Yano, C.A., 'Stochastic Leadtimes in Two-Level Assembly Systems' IIE Transactions, (1987b), Vol.19, No.4, pp.371-378.

# Appendix A. Moments and Correlations of the Maximum of Two Lognormals

The lognormal distribution is defined in terms of an associated normal. If the logarithm of a random variable is normally distributed, the variable is said to follow the lognormal distribution. If the associated normal for variable  $j$  has parameters  $\mu_{n_j}$  and  $\sigma_{n_j}$  then the  $i^{\text{th}}$  moment of the lognormal variable  $j$  is given by

$$\mu_j(i) = \exp[i\mu_{n_j} + i^2\sigma_{n_j}^2]$$

For the shifted or three parameter distribution,  $\log(X - \theta)$  is normally distributed, where  $\theta$  represents the shift or location parameter. The moments about zero are altered accordingly, while the central moments remain the same.

If two variables follow the two parameter lognormal distribution with parameters  $\mu_j$  and  $\sigma_j$  respectively and coefficient of correlation  $\rho$ , then the associated normals can be defined by

$$\mu_{n_j} = \ln \left[ \frac{\mu_j^2}{(\mu_j^2 + \sigma_j^2)^{1/2}} \right]$$

$$\sigma_{n_j}^2 = \ln \left[ \frac{\mu_j^2 + \sigma_j^2}{\mu_j^2} \right]$$

The correlation coefficient between the associated normals is given by

$$\rho_{12} = \frac{1}{\sigma_{n1}\sigma_{n2}} \ln\{\rho[(\omega_1 - 1)(\omega_2 - 1)]^{1/2} + 1\}$$

where  $\omega_k = \exp(\sigma_k^2)$ ,  $k = 1, 2$ .

In the three parameter distribution,  $(\mu_j - \theta)$  replaces  $\mu_j$  in the above equations.

The  $i^{\text{th}}$  moments of the maximum of two three-parameter lognormals can be determined using the relations derived by Wilhelm (1986c).

$$M_1(\theta) = M_1 + \theta$$

$$M_2(\theta) = M_2 + 2\theta M_1 + \theta^2$$

$$M_3(\theta) = M_3 + 2\theta M_2 + 2\theta^2 M_1 + \theta^3$$

$$M_4(\theta) = M_4 + 4\theta M_3 + 6\theta^2 M_2 + 4\theta^3 M_1 + \theta^4$$

where  $M_i$  are the moments if  $\theta_1 = \theta_2 = 0$  and are defined by:

$$M_i = M_{i1} + M_{i2}$$

$$M_{i1} = \mu_1(i) \phi\{[\mu_{n1} - \mu_{n2} + i\sigma_{n1}(\sigma_{n1} - \rho_{12})]/a\}$$

$$M_{i2} = \mu_2(i) \phi\{[\mu_{n2} - \mu_{n1} + i\sigma_{n2}(\sigma_{n2} - \rho_{12})]/a\}$$

$$a^2 = \sigma_{n1}^2 + \sigma_{n2}^2 - 2\sigma_{n1}\sigma_{n2}\rho_{12}$$

where  $\phi(z)$  = cumulative standard normal distribution function evaluated at  $z$ .

$M_{i2}$  is obtained by exchanging the subscripts in the formula for  $M_{i1}$ .

The new correlation coefficients  $r[X_3, \max(X_1, X_2)]$  are determined using the product moment  $E[X_3 \cdot \max(X_1, X_2)]$  as defined below.

$$E[X_3 \cdot \max(X_1, X_2)] = [E_1 + E_2] + [\theta E_1(\theta) + \theta E_2(\theta) + \theta_3 M_1(\theta)]$$

where  $E_1$  and  $E_2$  define the product moment if  $\theta_1 = \theta_2 = \theta_3 = 0$  and are:

$$E_1 = \exp[b_1 + (\mu_{n3} + Q^2/2) + (1/2a^2)(D_1^2 + A_1^2/\sigma_{n1}^2 R)] \phi[(\mu_{n1} - \mu_{n2} - D_1)/a]$$

$$E_2 = \Omega[E_1]$$

where the operator  $\Omega[\cdot]$  indicates that  $E_2$  is determined from the equation defining  $E_1$  by exchanging the subscripts, and the other quantities are as defined below.

$$R = 1 - \rho_{12}^2$$

$$C = [(\rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23})/R]^{1/2}$$

$$G_1 = (\rho_{13} - \rho_{12}\rho_{23})/R$$

$$B_{12} = \sigma_{n1} - \rho_{12}\sigma_{n2}$$

$$b_1 = \mu_{n1}$$

$$c_1 = \sigma_{n1}$$

$$Q = (1 - C^2)^{1/2} \sigma_{n3}$$

$$A_1 = [c_1 \sigma_{n2} + \sigma_{n3}(\sigma_{n2} G_1 + \sigma_{n1} G_2)] \sigma_{n2} R$$

$$D_1 = (-B_{12} G_1 + B_{21} G_2) \sigma_{n3} - \sigma_{n1} B_{12}$$

and the following symmetrical relationships hold.

$$G_2 = \Omega(G_1), B_{21} = \Omega(B_{12}), b_2 = \Omega(b_1), c_2 = \Omega(c_1), A_2 = \Omega(A_1), \text{ and } D_2 = \Omega(D_1).$$

$E_1(\theta)$  is determined from the above equation with  $b_1 = c_1 = 0$ , and  $E_2(\theta) = \Omega[E_1(\theta)]$ .

The required correlation can then be determined using the relation

$$r[X_3, \max(X_1, X_2)] = \frac{E[X_3 \cdot \max(X_1, X_2)] - E(X_3)E[\max(X_1, X_2)]}{\sigma X_3 \sigma[\max(X_1, X_2)]}$$

The new correlation coefficient  $r[X_1, \max(X_1, X_2)]$  is a special case and can be determined more easily using the following formula for the product moment  $E[X_1 \cdot \max(X_1, X_2)]$  as defined below.

$$\begin{aligned} E[X_1 \cdot \max(X_1, X_2)] &= \theta_1 \theta_2 \phi[(\mu_{n2} - \mu_{n1})/a] + \theta_1 \exp(\mu_{n2} + \sigma_{n2}^2/2) \phi\{[\mu_{n2} - \mu_{n1} + \sigma_{n2}(\sigma_{n2} - \rho_{12}\sigma_{n1})]/a\} \\ &\quad + \theta_2 \exp(\mu_{n1} + \sigma_{n1}^2/2) \phi\{[\mu_{n2} - \mu_{n1} - \sigma_{n1}(\sigma_{n1} - \rho_{12}\sigma_{n2})]/a\} \\ &\quad + \exp\{(\mu_{n1} + \mu_{n2}) + [(\sigma_{n1}^2 + \sigma_{n2}^2)^2 - 4\sigma_{n1}^2\sigma_{n2}^2\rho_{12}^2]/[2a^2]\} \phi[(\mu_{n2} - \mu_{n1} - \sigma_{n1}^2 + \sigma_{n2}^2)/a] \\ &\quad + \theta_1^2 \phi[(\mu_{n1} - \mu_{n2})/a] + 2\theta_1 \exp(\mu_{n1} + \sigma_{n1}^2/2) \phi\{[\mu_{n1} - \mu_{n2} + \sigma_{n1}(\sigma_{n1} - \rho_{12}\sigma_{n2})]/a\} \\ &\quad + \exp[2(\mu_{n1} + \sigma_{n1}^2)] \phi\{[\mu_{n1} - \mu_{n2} + 2\sigma_{n1}(\sigma_{n1} - \rho_{12}\sigma_{n2})]/a\} \end{aligned}$$

The required correlation can then be determined using the relation

$$r[X_1, \max(X_1, X_2)] = \frac{E[X_1 \cdot \max(X_1, X_2)] - E(X_1)E[\max(X_1, X_2)]}{\sigma X_1 \sigma[\max(X_1, X_2)]}$$

## Appendix B. Moments of the Maximum of Two

### Gamma Random Variables

The density for the two parameter gamma distribution is given by

$$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} \quad (\text{A.1})$$

The distribution function of the maximum is given by

$$Pr(X \leq t) = Pr(X_1 \leq t)Pr(X_2 \leq t) \quad (\text{A.2})$$

The density function is obtained by differentiating the above expression.

$$f_X(t) = f_{X_1}(t)Pr(X_2 \leq t) + f_{X_2}(t)Pr(X_1 \leq t) \quad (\text{A.3})$$

The  $j^{\text{th}}$  moment can be determined using the relation

$$M_j = \int_0^{\infty} x f_1(x) Pr(X_2 \leq x) dx + \int_0^{\infty} x f_2(x) Pr(X_1 \leq x) dx \quad (\text{A.4})$$

Substitution of the expressions for the density and distribution functions and evaluation of the resulting integrals yields relations for the moments required.

## B.1 Erlang Distribution

For the Erlang distribution, the shape parameter  $\alpha$  is an integer. The cumulative distribution function of the Erlang distribution is

$$Pr(X \leq t) = 1 - e^{-t/\beta} \sum_{j=0}^{\alpha-1} \frac{(t/\beta)^j}{j!}$$

Denoting the first integral in equation (A.4) by  $M_{i,1}$  and the second by  $M_{i,2}$ , and substituting the expressions for the density and the cumulative distribution function yields

$$M_{i,1} = \int_0^{\infty} x^i \frac{\beta_1^{-\alpha_1} x^{\alpha_1-1} e^{-x/\beta_1}}{(\alpha_1-1)!} \left[ 1 - e^{-x/\beta_2} \sum_{j=0}^{\alpha_2-1} \frac{(x/\beta_2)^j}{j!} \right] dx$$

The first term is the  $i^{\text{th}}$  moment of  $X_1$ , and is

$$M_{i,1,1} = \beta_1^i \prod_{k=0}^{i-1} (\alpha_1 + k)$$

The second term is

$$M_{i,1,2} = - \int_0^{\infty} \sum_{j=0}^{\alpha_2-1} x^i \frac{\beta_1^{-\alpha_1}}{(\alpha_1-1)! j!} e^{-x\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right)} \frac{x^{\alpha_1+j-1}}{\beta_2^j} dx$$



$$= - \sum_{j=0}^{\alpha_2-1} \frac{\beta_1^{-\alpha_1} \beta_2^{-j}}{(\alpha_1-1)! j!} \left[ \int_0^{\infty} x^j \frac{x^{\alpha_1+j-1}}{(\alpha_1+j-1)!} e^{-x\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right)} \left(\frac{\beta_1 \beta_2}{\beta_1 + \beta_2}\right)^{-\alpha_1-j} dx \right] \frac{(\alpha_1+j-1)!}{\left(\frac{\beta_1 \beta_2}{\beta_1 + \beta_2}\right)^{-\alpha_1-j}}$$

The term in square brackets is the  $j^{\text{th}}$  moment of an Erlang variable with  $\alpha = \alpha_1 + j$  and  $\beta = \frac{\beta_1 \beta_2}{\beta_1 + \beta_2}$ , and is given by  $\prod_{k=0}^{j-1} (\alpha_1 + j + i) \left(\frac{\beta_1 \beta_2}{\beta_1 + \beta_2}\right)^i$ . Substituting this expression yields

$$M_{i,1,2} = - \sum_{j=0}^{\alpha_2-1} \frac{(\alpha_1+j+i)! (\beta_1 + \beta_2)^{-\alpha_1-j}}{(\alpha_1-1)! j! \beta_1^{-j} \beta_2^{-\alpha_1}} \left(\frac{\beta_1 \beta_2}{\beta_1 + \beta_2}\right)^i$$

Adding the two terms together yields the following expression

$$M_{i,1} = \beta_1^i \prod_{k=0}^{i-1} (\alpha_1 + k) - \sum_{j=0}^{\alpha_2-1} \frac{(\alpha_1+j+i)! (\beta_1 + \beta_2)^{-\alpha_1-j}}{(\alpha_1-1)! j! \beta_1^{-j} \beta_2^{-\alpha_1}} \left(\frac{\beta_1 \beta_2}{\beta_1 + \beta_2}\right)^i$$

The corresponding expression for  $M_{i,2}$  is

$$M_{i,2} = \beta_2^i \prod_{k=0}^{i-1} (\alpha_2 + k) - \sum_{j=0}^{\alpha_1-1} \frac{(\alpha_2+j+i)! (\beta_1 + \beta_2)^{-\alpha_2-j}}{(\alpha_1-1)! j! \beta_2^{-j} \beta_1^{-\alpha_2}} \left(\frac{\beta_1 \beta_2}{\beta_1 + \beta_2}\right)^i$$

The  $i^{\text{th}}$  moment can then be obtained by adding the two terms

$$M_i = M_{i,1} + M_{i,2}$$

## B.2 General Gamma Distribution

The infinite series representation for the cumulative distribution function of the general gamma distribution ( Johnson and Kotz, 1972b) is

$$Pr(X \leq t) = e^{-t/\beta} \sum_{j=0}^{\infty} \frac{(t/\beta)^{\alpha+j}}{\Gamma(\alpha+j+1)}$$

Denoting the first integral in equation (A.4) by  $M_{i,1}$  and the second by  $M_{i,2}$ , and substituting the expressions for the density and the cumulative distribution function yields

$$M_{i,1} = \int_0^{\infty} x^i \frac{\beta_1^{-\alpha_1} x^{\alpha_1-1} e^{-x/\beta_1}}{\Gamma(\alpha_1)} e^{-x/\beta_2} \sum_{j=0}^{\infty} \frac{(x/\beta_2)^{\alpha_2+j}}{\Gamma(\alpha_2+j+1)} dx$$

$$= \sum_{j=0}^{\infty} \frac{\beta_1^{-\alpha_1} \beta_2^{-\alpha_2-j}}{\Gamma(\alpha_1) \Gamma(\alpha_2+j+1)} \left[ \int_0^{\infty} x^i \frac{x^{\alpha_1+\alpha_2+j-1} e^{-x\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right)}}{\Gamma(\alpha_1+\alpha_2+j)} \left(\frac{\beta_1\beta_2}{\beta_1+\beta_2}\right)^{-\alpha_1-\alpha_2-j} dx \right]$$

$$\frac{\Gamma(\alpha_1+\alpha_2+j)}{\left(\frac{\beta_1\beta_2}{\beta_1+\beta_2}\right)^{-\alpha_1-\alpha_2-j}}$$

The term in square brackets is the  $i^{\text{th}}$  moment of a gamma variable with  $\alpha = \alpha_1 + \alpha_2 + j$ , and  $\beta = \frac{\beta_1\beta_2}{\beta_1+\beta_2}$ , and is given by  $\prod_{k=0}^{i-1} (\alpha_1 + \alpha_2 + j + k) \left(\frac{\beta_1\beta_2}{\beta_1+\beta_2}\right)^i$ .

Substituting this expression results in

$$M_{i,1} = \sum_{j=0}^{\infty} \frac{\beta_1^{-\alpha_1} \beta_2^{-\alpha_2} \Gamma(\alpha_1 + \alpha_2 + j)}{\Gamma(\alpha_1) \Gamma(\alpha_2 + j + 1)} \prod_{k=0}^{j-1} (\alpha_1 + \alpha_2 + j + k) \left( \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \right)^{\alpha_1 + \alpha_2 + i + j}$$

$$= \frac{\beta_1^{-\alpha_1} \beta_2^{-\alpha_2}}{\Gamma(\alpha_1)} \left( \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \right)^{\alpha_1 + \alpha_2 + i} \sum_{j=0}^{\infty} \frac{\Gamma(\alpha_1 + \alpha_2 + j + i)}{\Gamma(\alpha_2 + j + 1)} \left( \frac{\beta_1}{\beta_1 + \beta_2} \right)^j$$

$$= \frac{\beta_1^{-\alpha_1} \beta_2^{-\alpha_2}}{\Gamma(\alpha_1)} \left( \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \right)^{\alpha_1 + \alpha_2 + i} \cdot$$

$$\left[ \frac{\Gamma(\alpha_1 + \alpha_2 + i)}{\Gamma(\alpha_2 + 1)} + \frac{\Gamma(\alpha_1 + \alpha_2 + i + 1)}{\Gamma(\alpha_2 + 2)} \left( \frac{\beta_1}{\beta_1 + \beta_2} \right) + \frac{\Gamma(\alpha_1 + \alpha_2 + i + 2)}{\Gamma(\alpha_2 + 3)} \left( \frac{\beta_1}{\beta_1 + \beta_2} \right)^2 + \dots \right]$$

$$= \frac{\beta_1^{-\alpha_1} \beta_2^{-\alpha_2} \Gamma(\alpha_1 + \alpha_2 + i)}{\Gamma(\alpha_1) \Gamma(\alpha_2 + 1)} \left( \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \right)^{\alpha_1 + \alpha_2 + i} \cdot$$

$$\left[ 1 + \frac{(\alpha_1 + \alpha_2 + i)}{(\alpha_2 + 1)} \left( \frac{\beta_1}{\beta_1 + \beta_2} \right) + \frac{(\alpha_1 + \alpha_2 + i + 1)}{(\alpha_2 + 2)} \left( \frac{\beta_1}{\beta_1 + \beta_2} \right)^2 + \dots \right]$$

$$= \frac{\beta_1^{-\alpha_1} \beta_2^{-\alpha_2} \Gamma(\alpha_1 + \alpha_2 + i)}{\Gamma(\alpha_1) \Gamma(\alpha_2 + 1)} \left( \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \right)^{\alpha_1 + \alpha_2 + i} F\left(1, \alpha_1 + \alpha_2 + i, \alpha_2 + 1, \frac{\beta_1}{\beta_1 + \beta_2}\right)$$

where  $F(a, b, c, z)$  is the generalized hypergeometric series given by

$$F(a, b, c, z) = 1 + \frac{ab}{c} z + \frac{a(a+1)b(b+1)}{c(c+1)} \frac{z^2}{2!} \dots$$

The relation for  $M_{i,1}$  is simplified using the following transformation from Slater (1966).

$$F(\alpha, \beta, \gamma, z) = (1-z)^{\gamma-\alpha-\beta} F(\gamma-\alpha, \gamma-\beta, \gamma, z)$$

$$M_{i,1} = \frac{\beta_1^{-\alpha_1} \beta_2^{-\alpha_2} \Gamma(\alpha_1 + \alpha_2 + i)}{\Gamma(\alpha_1) \alpha_2 \Gamma(\alpha_2)} \left( \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \right)^{\alpha_1 + \alpha_2 + i} \left( \frac{\beta_2}{\beta_1 + \beta_2} \right)^{-\alpha_1 - i} \cdot$$

$$F\left(\alpha_2, -\alpha_1 - i + 1, \alpha_2 + 1, \frac{\beta_1}{\beta_1 + \beta_2}\right)$$

$$= \frac{\Gamma(\alpha_1 + \alpha_2 + i)}{\Gamma(\alpha_1) \alpha_2 \Gamma(\alpha_2)} \frac{\beta_1^{\alpha_2 + i}}{(\beta_1 + \beta_2)^{\alpha_2}} F\left(\alpha_2, -\alpha_1 - i + 1, \alpha_2 + 1, \frac{\beta_1}{\beta_1 + \beta_2}\right)$$

For integer values of  $\alpha_1$ , the hypergeometric series in the above formula reduces to a finite terminating series and can be calculated exactly (Rao, 1975a). For non-integer values, the series converges and can be evaluated by summation to the accuracy desired since the argument  $z = \frac{\beta_1}{\beta_1 + \beta_2}$  falls within the circle of convergence  $|z| < 1$  (Slater, 1966).

The corresponding term for  $M_{i,2}$  is obtained by exchanging the subscripts in the formula for  $M_{i,1}$ , and is

$$M_{i,2} = \frac{\Gamma(\alpha_1 + \alpha_2 + i)}{\Gamma(\alpha_2) \alpha_1 \Gamma(\alpha_1)} \frac{\beta_2^{\alpha_1 + i}}{(\beta_1 + \beta_2)^{\alpha_1}} F\left(\alpha_1, -\alpha_2 - i + 1, \alpha_1 + 1, \frac{\beta_2}{\beta_1 + \beta_2}\right)$$

The  $i^{\text{th}}$  moment can then be obtained by adding the two terms

$$M_i = M_{i,1} + M_{i,2}$$

# Appendix C. Correction Equations for Different Distributions

## Distributions

The approximate solution technique uses the following equations to correct the decisions from the independent solution at station  $i$ , where  $i = 1, \dots, N-1$ .

$$DS_i = \sigma D_i - \sigma F_{i-1}$$

$$CP_i = \frac{CE_i}{CS_i + CE_i}$$

$$ACR_i = \max\left(\frac{CS_{i+k} + CE_{i+k}}{CS_i + CE_i}\right)^{1/k}, k = 1, \dots, N - k$$

$$VF_k = \frac{\sigma S_k}{\sigma F_k} = \frac{\sigma S_k}{\sqrt{\sigma S_k^2 + \sigma P_k^2}}$$

$$NVF_k = \prod_{l=i}^k VF_l = \prod_{l=i}^k \frac{\sigma S_l}{\sqrt{\sigma S_l^2 + \sigma P_l^2}}$$

$$\Delta_i^V = (AV_i \cdot VF_i + \sum_{k=i+1}^{N-1} BV_i \cdot NVF_k)$$

$$\Delta_i^P = (AP_i \cdot VF_i + \sum_{k=i+1}^{N-1} BP_i \cdot NVF_k)$$

$$\mu D_i^{Cl} = \mu D_i^I + \Delta_i^V + \Delta_i^P$$

The equations for the intercepts ( $AV_i$  and  $AP_i$ ) and the slopes ( $BV_i$  and  $BP_i$ ) of the correction factors for the distributions considered are as follows.

### **Normal Distribution**

$$AV_i = - (-0.01461ACR_i^2 + 0.31393ACR_i + 0.14990)DS_i$$

$$BV_i = - (-0.00927ACR_i^2 + 0.14457ACR_i - 0.09764)DS_i$$

$$AP_i = (-0.01707ACR_i^2 + 0.38048ACR_i + 0.15472)(0.5 - CP_i)\sigma D_i$$

$$BP_i = (-0.00965ACR_i^2 + 0.13324ACR_i - 0.03230)(0.5 - CP_i)\sigma D_i$$

### **Lognormal Distribution**

$$AV_i = - (-0.03050ACR_i^2 + 0.45698ACR_i + 0.06696)DS_i$$

$$BV_i = - (-0.01112ACR_i^2 + 0.15775ACR_i - 0.11561)DS_i$$

$$AP_i = (-0.03620ACR_i^2 + 0.59250ACR_i - 0.27085)(0.5 - CP_i)\sigma D_i$$

$$BP_i = (-0.00266ACR_i^2 + 0.07320ACR_i + 0.03409)(0.5 - CP_i)\sigma D_i$$

### **Gamma Distribution**

$$AV_i = - (-0.02914ACR_i^2 + 0.49023ACR_i - 0.12720)DS_i$$

$$BV_i = - (-0.00132ACR_i^2 + 0.06537ACR_i + 0.03989)DS_i$$

$$AP_i = (-0.03620ACR_i^2 + 0.59250ACR_i - 0.32085)(0.5 - CP_i)\sigma D_i$$

$$BP_i = (-0.00250ACR_i^2 + 0.07088ACR_i + 0.04238)(0.5 - CP_i)\sigma D_i$$

**The vita has been removed from  
the scanned document**