A MODEL FOR MULTI-PLANT COORDINATION : IMPLICATIONS FOR PRODUCTION PLANNING

A Thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Rohit Bhatnagar

Faculty of Management

McGill University

October 6, 1994

ISBN 0-612-00078-8

Short Thesis Title (62 characters including spaces)

Multi-Plant Coordination: Implications for Production Planning

Dedicated to my parents and to Mamta

for their love and support

Abstract

Firms in several d... ete parts manufacturing industries, e.g., electronics equipment, computers,telecommunications equipment etc. operate in a multi-plant environment where products are processed successively at several plants. Prior studies have ignored the interaction between different plants in a multi-plant scenario. The objective of this dissertation is to study the impact of coordination on the cost performance of a two-plant firm.

We propose a model that jointly determines production and inventory decisions so that the total cost of holding inventory and overtime, at the two plants is minimized. Our model captures the interaction between the two plants and is preferable to the uncoordinated or the sequential approach which ignores this interaction. We consider the case with limited capacity and explicitly model setup times. Strategies based on Lagrangian relaxation and Lagrangian decomposition methodologies are proposed to solve the model.

Two main findings emerge from this research. First, our results indicate that coordination could lead to improved cost performance and enhanced profits for firms. Two parameters, the setup time to processing time ratio and the capacity utilization at the two plants played a significant role in determining the cost improvements. Managerial implications relating to implementation of the coordinated model are discussed. The second important finding of this research is that Lagrangian decomposition consistently outperforms Lagrangian relaxation in terms of achieving better deviation from the optimal solution, for this problem. A Linear Programming based technique for further enhancing the convergence between the upper and lower bounds is presented.

In the quest for improved performance, multi-plant coordination represents an important strategy for firms. The contribution of the current research is in modelling some of the salient issues of this problem and exploring promising methodological directions.

Preface

Les entreprises de plusieurs industries manufacturières de pièces physiques, e.g. équipement électronique, ordinateurs, équipement de télécommunication, etc... ont une infrastructure multi-usines, i.e. où les produits sont fabriqués successivement à plusieurs usines. Des aspects importants du problème de la planification de la production, relativement aux entreprises multi-usines et dans le contexte d'une entreprise à deux usines, fonts l'objet de cette dissertation.

Nous proposons un modèle qui détermine les décisions de production et d'inventaires de sorte que le cot total de storage et d'heures supplémentaires soit minimisé aux deux usines. Notre modèle retient bien l'interaction entre les deux usines et il est préférable à comparer aux approches séquentielles et non-coordinnées, lesquelles ignorent cette interaction. Nous analysons le cas réaliste et modélisons explicitement les temps d'installation, ce qui est une représentation plus précise du type de système de prodution étudié que celle de modèles qui utilisent un coût d'installation constant comme proxy pour le temps d'installation. Nous proposons différentes stratégies, basées sur la relaxation de Lagrange et sur la décomposition de Lagrange, comme solutions pour résoudre le modèle.

Deux principales découvertes resortent de cette recherche. En premier lieu, nos résultats indiquent que la coordination pourrait mener à une meilleure performance au niveau du cot ainsi qu'à des profits supplémentaires pour les entreprises. L'analyse de variance a révélé que la valeur de deux paramètres, le ratio du temps d'installation au temps de fabrication et la capacité d'utilization aux deux usines, jouent un rôle significatif dans la détermination des améliorations de coûts. Nous élaborons sur l'importance de cette recherche au niveau du management. Le deuxième résultat important est que la décompostion de Lagrange surpasse constamment la relaxation de Lagrange en obtenant une meilleure déviation de la solution optimale pour le problème de coordination dans une structure multi-usines. Une technique d'algèbre linéaire pour accentuer la convergence entre les limites (supérieures et inférieures) est enfin proposée.

Dans la recherche d'une performance améliorée, la coordination de la structure multi-usine est une stratégie importante pour les entreprises. La contribution de cette recherche est au niveau de la modélisation de certains points (du problème) les plus à propos tout en explorant différentes méthodologies prometteuses.

Acknowledgements

The river meanders around one last mountainous slope, and suddenly the vast plains are in sight. The steep mountain slopes that gave momentum, the banks that gave direction will soon become a remembrance. But the flow that has been generated will remain, and will guide the course of the river as it confronts new challenges. As I conclude this thesis, I find myself at a turning point too. In moving on to my rendezvous with the new challenges of an academic life, I take this opportunity to acknowledge the lofty mountains that gave me momentum and the banks that shaped my direction, as I meandered through the Ph.D program at McGill.

My foremost thanks are due to my dissertation committee comprising Professors Richard Loulou (advisor), Pankaj Chandra and Jean-Louis Goffin.

Professor Loulou's timely insights inspired me to explore several new research directions in my thesis. His comments were always very challenging - at the same time he was very supportive when things did not take anticipated directions. Interacting with him was a very enriching experience - whether it was through one of the several courses I took with him, or it was the research project on shop floor synchronization at Northern Telecom. He made the doctoral program an enjoyable and memorable experience. I could not have asked for a better advisor.

Pankaj's contribution to this thesis is unique. Our close association predates my Ph.D years at McGill considerably. I thank him for being very supportive during these last five years at McGill, especially at times when I only had his word about the light at the end of the tunnel. I learnt a lot from our joint work on variability in assembly systems and shop floor synchronization at Northern Telecom. I thank Pankaj and Deepa for being very gracious hosts to me and Mamta during our first days at Montreal and on several occasions thereafter.

Professor Goffin taught me my first and last courses at McGill. I especially thank him for providing several important methodological inputs into my thesis.

I also acknowledge the help of Professors S.K. Goyal and Rafaella Delpilar for serving on my dissertation committee during the second and third phases of the Ph.D program. I thank Professors David Saunders, Emine Sarigollu and Mohan Gopalakrishnan for being on my final defense committee and for adding valuable perspectives to some of my interpretations and analysis.

I thank Alain Meer for helping me with the french translation of my abstract. I greatly appreciate the kindness of the Faculty of Arts Computer Lab administration in allowing me the liberal use of their computing facilities.

I thank the trustees of the Max Binz Fellowship and Northern Telecom for their generous financial support during my Ph.D. program.

Throughout these years, I was lucky to have a group of very bright and supportive colleagues. I especially thank France Gaudreau, Jean-Guy Simonato, Henri Fouda among my batch mates and Atiqur Rahman, Sudheer Gupta, Mirjana Vajic and Faisal Bari among my younger colleagues, for their friendship and the good times I shared with them. I wish you all good luck in your future endeavours.

My parents have been a source of inspiration for every major undertaking in my life. I thank them for their vision in giving me the education that they themselves did not have the opportunity to receive. I also thank my other family members for their love and support.

My last and most notable acknowledgement is to my wife, Mamta, without whose support this thesis would not have taken its final shape. The space here is too small to fully list her complete contributions. I wish her luck and hope she fulfils her dreams.

Contents

Bibliography

Appendices

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Issues

The erosion of market shares in several important industries has forced North American firms to take a new look at their manufacturing practices. Studies of successful Japanese and American companies underline, among others, the need to reduce inventory levels and manufacturing lead times, to achieve better coordination between the different autonomous units that comprise the organization and to effectively adapt to rapidly changing manufacturing requirements as product life cycles shorten (Dertouzos et al., 1990). Reduced inventory (and lead time) results not only in better cost performance but also exposes the inter-dependence of individual units providing impetus for better quality management, coordination between units and flexibility to switch to new products quickly as consumer requirements change. Some authors have advocated surplus production capacity as a strategy for inventory minimization. This option facilitates the scheduling of production as close as possible to actual requirement of items, reducing the need to hold inventory, but it entails higher capital investment. Moreover, in microelectronics industries like computers, telecommunications equipment etc., which are characterized by increasingly compressed product life cycles, demand for individual items may fluctuate rapidly from period to period. Providing excessive capacity may therefore be uneconomical. Given fixed capacity, an important decision problem for the firm is to satisfy demand while ensuring optimal use of its limited productive resources such as investment in inventory, payment for regular and overtime labour etc. While regular labour is often a fixed, committed charge in the short to medium term, costs incurred for carrying inventory and for scheduling overtime depend on managerial decisions which must be made prudently.

Another reality that managers must confront in today's business environment is an increasingly complex and globalized **supply chain**. A supply chain refers to a network of facilities that procures raw materials, transforms them into intermediate and finished goods and delivers the finished products to customers through a distribution system (Lee and Billington, 1992). To remain globally competitive, firms must source raw materials from the most effective supplier (who may be considerably far-flung), and supply a disbursed customer base. Pressure to provide good service to customers invariably results in an increased accumulation of inventories in the supply chain. For example, it has been estimated that Hewlett-Packard Company has more than 3 billion dollars invested in worldwide inventories (Billington, 1994).

In the context of supply chain management, a strategic imperative for the firm is to minimize inventory throughout the supply chain while providing good service to customers. A vital requirement for effective supply chain management is coordination between the individual units comprising the organization. From an inventory point of view, coordination is desirable because it captures the complex inter-relationships between inventory at different sites, in the firm's decision making process. The benefits that flow from improved communication between different units when coordinated decisions are implemented are also of considerable value to the firm. This advantage is increasingly seen as being critical to the firm's market responsiveness. Planning for coordination between different entities in a supply chain is therefore likely to be a significant management concern in future. We use an example of a supply chain to describe the two main types of entities, manufacturing and distribution, to develop the pattern of interaction between these entities and to enumerate the benefits of coordination. Figure 1.1 is a representation of a supply chain which is adapted from Lee and Billington (1992). These authors have described the supply chain for a computer manufacturer where work-in-process undergoes value addition at several manufacturing plants till the finished computer systems are obtained. The finished computer systems are sent to distribution centres from where they are distributed to customers.

All manufacturing plants and distribution centres carry inventory stockpiles and receive inputs from external suppliers. The manufacturing plant at the first level is concerned with semiconductor manufacturing. At this plant, silicon discs sliced from a silicon crystal are imprinted with the dense patterns of an integrated circuit. The main technical processes at this plant include wafer fabrication, wafer probing, device assembly and device testing. For details of the manufacturing processes, see Cooper et al., (1992) and Bassok and Akella (1991). The **chips**, as the tested devices are called are sent to the next level plant which is responsible for printed circuit board (PCB) manufacturing. Several different chips are mounted on plastic boards to produce PCBs. The main processes here include surface mounting and through hole insertion of components, in-circuit testing, functional testing and the burn-in test for weeding out infant mortality. Details of PCB manufacturing can be found in Bhatnagar et al., (1993) which describes a field study carried out at a telecom equipment plant of Northern Telecom Fiberworld Division. After testing, PCBs are sent to the third level plant where several PCBs and other electronic devices are assembled to produce finished computer systems. The completed computer systems are sent to the distribution centres from where they are distributed to the final customers. In a company like Hewlett Packard, manufacturing plants may be as far apart as United States and

Asia for catering to a customer base spread over USA, Europe and Asia (Billington, 1994). This trend towards procuring, manufacturing and selling globally has increased considerably in the last decade or so and it appears unlikely to diminish in future.

Spurred in a large measure by this trend, the study of supply chain management has received notable research attention recently. Work to date encompasses a broad range of issues. There has been much debate about the effectiveness of the different organisational structures found in North America and Japan. This stream of research has focused on determining empirically if either of these organisational structures is inherently more flexible and adaptable for today's uncertain and rapidly changing business environment (OECD, 1991). The enterprises are typically organized in a modular structure consisting of a network of interlinked facilities called **keiretsu**, a grouping of industrial firms linked by equity cross ownership, interlocking management, and long term buying-selling relationships. Each firm in a keiretsu specialises in the process associated with a part of the value addition chain. Such an organization has resulted in substantial benefits to each profit making centre within the keiretsu and is believed to be one of the main sources of Japan's overwhelming industrial competitiveness (OECD, 1992). In contrast to the Japanese system of networking and subcontracting within the group structure, North American firms have extensively relied on a strategy of horizontal and vertical integration. Sometimes vertical integration may be necessary due to strategic reasons. Examples in the computer industry are International Business Machines (IBM) in North America and Acer Computers in Taiwan which manufacture their requirements of semiconductor chips in-house. Acer Computers was forced to establish its own wafer fabrication facilities in collaboration with Texas Instruments of USA after its sales of computers were affected by the U.S.-Japan Semiconductor Agreement of 1986 which resulted in a severe shortage and steep price increases of dynamic random access memory chips (DRAMs) (Far

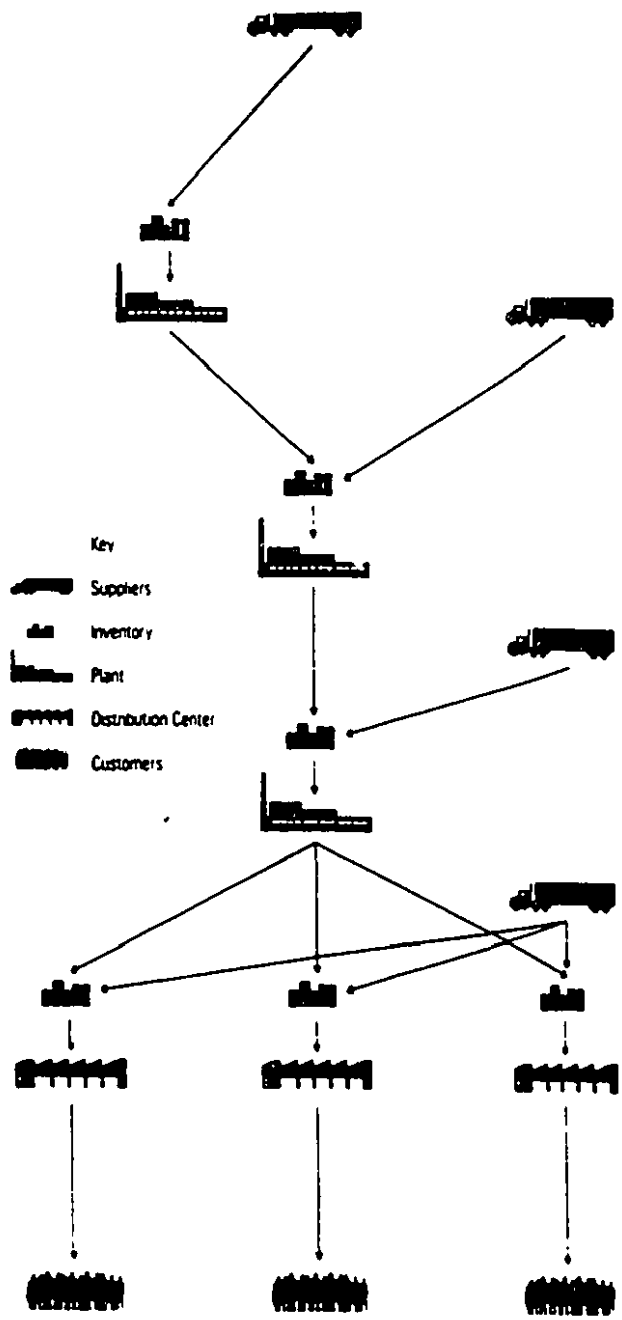

Figure 1.1: Representation of Supply Chain

Source: "Managing Supply Chain Inventory: Pitfalls and Opportunities," Lee H.L., and Billington C., Sloan Management Review, Spring 1992.

Eastern Economic Review, 1989).

Unfortunately, despite vertical integration, firms in North America are characterized by lack of intra-firm cooperation. The focus on narrowly defined short term performance measures to judge managerial performance often results in extremely compartmentalized organizations where managers responsible for successive stages of the value addition chain have little motivation to coordinate their efforts. For example, the problems arising from the lack of interaction between product design and production departments is well documented in Dertouzos et al., (1990). In a similar manner decisions relating to production planning in different plants within the same firm are often addressed in a sequential manner. The downstream plant typically takes its production planning decisions independently and communicates these decisions to the upstream plant which uses this information as input. This method ignores interaction between the plants and results in sub-optimal planning for the organization as a whole, thereby increasing the overall cost of operations.

Whether firms adopt a keiretsu type of structure or choose to integrate vertically, it appears inevitable that firms in North America will have to improve their coordination patterns both within the organization and with their supply network if they are to be competitive in the business environment of the future. A modified form of the Japanese keiretsu may well be the likely future structure of firms in North America, with different firms specializing in different parts of the value addition process. This view is supported by Drucker (1990) who observes that the manufacturing firms of the future will probably be organized as a **flotilla**, consisting of interlinked modules centred around closely related production processes. The trend towards networking will be accelerated by the availability of on-line information exchange systems via efficient telecommunications such as Electronic Data Interchange (EDI). To propose a decision making framework for such an organization, a range of issues need to be deliberated at both strategic and operational levels. See Table 1.1 for a summary of the most salient of these issues. These issues relate to the two main types of entities, manufacturing and distribution that exist in a supply chain. Our discussion of these issues is from a fairly broad perspective with the objective of bringing out both strategic and operational imperatives for firms.

In a broad sense, manufacturing entities encounter decisions relating to two main activities, material procurement and production or the transformation of the procured items into finished goods. The issues of material procurement and supplier relations have been in the spotlight as these are believed to be a major source of the competitiveness of Japanese companies. An excellent treatment of these issues is contained in a recent book which makes a cross-regional comparison of automobile companies in the United States, Europe and Japan (Womack et al., 1990). Although this book deals with the automobile industry, most issues are pertinent to other industries as well. For examples in the electronics industry, see OECD 1991, 1992 publications. Some features that characterize the Japanese procurement system are, tiered structure, long term relationship between the supplier and the buyer, close association of the supplier from early stages of the design process, rational joint cost analysis for ensuring reasonable mutual profits and a continuous decline in cost of product over the lifetime of the product (see Table 1.2). Toyota Motor Company, for example sources out complete parts to about 300 first tier suppliers who in turn sub-contract the required components to a second level of suppliers. General Motors, by contrast deals with almost 2500 suppliers, by directly procuring all components and doing the assembly at its own plants. One of the most distinguishing aspects of the relationship between Japanese firms and their suppliers is a pricing system that works on a **market price minus** rather than a **supplier cost plus** system. Starting from a target market price, techniques like value engineering and value analysis (see Womack et al., 1990 p.148 for details) are used to reduce costs so that reasonable profits can be assured for both the firm and its suppliers. The core of this system is a rational framework for analyzing costs, establishing prices and sharing profits which replaces the relative bargaining power of the two sides.

Establishing such a framework requires a long term vision that motivates the supplier to share proprietary information for the benefit of both the parties. Japanese suppliers are willing to share process information with the buyer firms, not because they are inherently more cooperative, but because cooperation has continuously ensured higher profits in the past. A worthwhile measure towards improved supplier buyer relations maybe a long term policy that rewards the supplier for coordination efforts by an investment in process improvement out of cash benefits of coordination. Also this policy needs to be integrated with the operational level supply policy by setting out suitable performance measures that reinforce coordination.

For production activities, strategic issues that are critical are enumerated in Table 1.1. Choice of appropriate technology is critical, given the realities of compressed product life cycles which in the computer industry may be as short as six months to one year. The firm must consider the trade-off between the costs and benefits of dedicated versus flexible technology. The issue of production capacity that the firm chooses to have at different facilities has been alluded to before. The traditional trade-off in capacity planning is between the economies of adding large capacities versus the cost of installing capacity before it is needed (Luss, 1982). In the context of a supply chain, an additional concern is to allocate limited resources to add capacity judiciously across all entities. Capacity added at one entity in isolation in a supply chain may fail to give adequate improvements unless capacity is sufficient at other entities. The capacity decision needs to be integrated with good product and process design to ensure ease of manufacturing,

Activity		Issues	Benefits of Coordination
Manufacturing			
(a)	Material Procurement	*Strategic*	
		1. Structure of Supplier Network	1. Increased Competitiveness 2. Reduced Costs
		Operational	
		1. Efficiency of Ordering Policy	1. Greater Reliability 2. Reduced Costs
(b)	Production	*Strategic*	
		1. Technology Choice 2. Capacity Expansion 3. Facility Location 4. Product/Process Design	1. Increased Competitiveness 2. Improved Capital Utilization
		Operational	
		1. WIP Management 2. Process Improvement 3. Overtime/Expediting Subcontracting	1. Reduced Costs 2. Improved Service
Distribution			
		Strategic	
		1. Warehouse Location	1. Reduced Costs
		Operational	
		1. Inventory 2. Transportation 3. Expediting	1. Reduced Costs 2. Faster Response

Table 1.1: Issues for Supply Chain Management

Japanese	North American
1. Tiered Structure - Sourcing of complete parts - 300 suppliers at Toyota	1. Untiered Structure - 2500 suppliers at General Motors
2. Long term relationship	2. Short term contracts
3. Suppliers closely associated with design	3. Suppliers brought in late into design process
4. Rational joint cost analysis	4. Intense cost pressure from buyer
5. Merging learning curves of buyer and supplier	5. Little exchange of process information

Table 1.2: Features of Material Procurement

reduced costs, good quality and flexibility to meet future changes. From the point of view of supply chain management, the focus of product and process design must be on good performance on these measures for the entire set of entities. Interaction between processes of different manufacturing entities must therefore be captured in the decision making process.

At the operational level, the trade-off relates to the optimal balancing between inventory, setup, overtime, backorder and expediting costs. As mentioned earlier, these decisions are typically done sequentially for different manufacturing entities in the supply chain and the interaction between processes at different entities is ignored. Sequential planning results in sub-optimality and hence higher costs for the firm as a whole. Coordination between different entities results not only in reduced costs but also fosters better communication between managers at different entities.

Distribution entities serve as an interface between manufacturing plants and customers. As outlined in Table 1.1, at a strategic level, the important decision making issue relates to warehouse location, while at the operational level, decisions must be taken regarding inventory, transportation and expediting. From the point of view of supply chain management, coordination between manufacturing and distribution assumes significance. Most companies manage manufacturing and distribution planning independently, without any coordination between the two functions. The rationale for coordination between these functions arises from the conflicting impact of their individual objectives. Manufacturing would like to plan large production lots to increase efficiency, but this increases inventory and limits flexibility. Distribution on the other hand, would like to consolidate batches of different items together to reduce transportation costs necessitating additional setups or increased inventory. Frequent shipments (espoused in just-in-time manufacturing systems) reduce inventory but result in increased transportation costs. Planning either function in isolation from the other may therefore worsen the performance of the other function as well as the overall performance of the firm. Bendiner (1993) has argued that production and distribution planning are an interrelated set of activities and must be managed and controlled as part of an integrated system. The author recommends instituting performance measures and organizational changes that reinforce integration. Lee et al., (1993) present another aspect of coordination between manufacturing and distribution entities in a supply chain by using the concept of **design for customization**. The authors show that by performing customization of products at distribution centers instead of at the manufacturing plant, companies can benefit substantially in terms of inventory costs and flexibility. This represents an integration of the product/process design process across both manufacturing and distribution entities.

These are some of the salient issues that are important for the study of supply chain management. It is evident from the previous discussion that these issues represent a rich and varied set of research questions that are likely to be pertinent in the years to come. Analysis of these issues poses an important challenge to researchers in both academic and industrial settings. Treating all the above issues explicitly in a single model will apparently make the model very difficult to solve. A good strategy in such a scenario is to first consider subsets of the above issues within sub-models and to later build linkages between these sub-models. One distinction which we have already made in Table 1.1 serves as a intuitive guideline for differentiating sub-models, i.e., strategic versus operational issues. A set of manufacturing entities and one manufacturing plus one distribution entity are alternative ways of differentiating sub-models. These represent different approaches to coordination using distinctive sets of issues. In this dissertation we focus on the operational issues relating to coordination between manufacturing entities and hence the term multi-plant coordination. We use a real life multi-plant manufacturing situation at IBM as the motivating example for this research. However the constructs defined herein apply in general to discrete parts manufacturing industries.

The multi-plant structure is found in discrete part manufacturing industries like computers, telecommunication equipment etc., (Bassok and Akella, 1991). In a multi-plant scenario, an important managerial issue is the determination of production quantities at each plant such that total costs at all plants are minimized. This problem is typically addressed sequentially or in an **uncoordinated** manner. For example, in a two-plant firm, the problem is first solved for the downstream plant and its production plan is determined. This defines the demand vector for the upstream plant. Using this demand as input, the problem for the upstream plant is then solved independently. The uncoordinated approach is inadequate because it does not consider the interactions that exist between the two plants and can result in sub-optimal production plans.

In contrast, we propose a **coordinated** approach which jointly determines production quantities at both plants. The objective of this research is to compute the cost benefits of coordinated production planning, in a multi-plant manufacturing system. The two-plant case is considered although the extension to the more general n-plant scenario is not difficult. The products are successively processed at the two plants. See Figure 1.2 for a representation of the problem.The upstream plant, A, produces semiconductor devices called **chips** which are then mounted on ceramic substrate called **modules** at Plant B. Each module requires several chips and a chip could be used in several modules. We consider a single bottleneck workcenter at each plant. If a well defined bottleneck workcenter does not exist, the workcenter could represent what Karmarkar et al., (1992) call an **approximate composite model** of an entire facility. Given demand forecasts over a planning horizon, the problem is to determine production/inventory levels at each plant so that the overall costs are minimized. While the optimal solution to this problem may not be optimal for any of the plants considered individually, it represents the best result for the joint problem. This approach is distinct from the uncoordinated approach where an optimal solution is found for the downstream plant considered individually but the upstream plant must find the best solution for the demand defined by the production plan of the downstream plant. We consider the capacitated case, wherein the total of setup and processing times must not exceed the specified capacity.

An important aspect of the kind of manufacturing systems under study is that the cost in terms of actual cash outflow associated with setting up a machine is often negligible. This is because no materials are consumed in setting up the machine. The cost of a setup is essentially the time consumed which reduces the productive capacity of the facility. Most previous efforts in literature have dealt with setup time indirectly by means of a pre-determined setup cost. Theoretically, the setup cost should reflect the shadow price of the capacitated resource. Since the shadow price depends on resource usage, the use of a pre-determined setup cost is an inaccurate representation of the production system. In this paper we explicitly model setup times, which more accurately reflects the true situation in the class of manufacturing systems under study.

Since we formulate our models as mixed-integer programs, the methodological underpinnings of this research are built around the theory relating to the optimization of mixed-integer problems. The specific methodology that we use is Lagrangian relaxation. The main justification for this choice is that Lagrangian relaxation has often provided good solutions to difficult combinatorial problems. We also present an alternative solution strategy based on Lagrangian decomposition (called variable splitting by some authors) which is a generalized version of Lagrangian relaxation. Details of these alternative methodologies are described in later chapters. Our results indicate that the coordinated approach to production planning in multi-plant firms leads to improved cost performance, reducing total costs by more than 10% in several cases. Exploration of a variety of experimental factors at different levels demonstrates the robustness of our results, i.e., cost improvement is shown to be not dependent on a specific mix of input factors. The gap between the lower and upper bounds (which represents the maximum deviation from optimality) turns out to be moderate and does not increase with increase in problem size. This is reasonable, given that the problem we are considering is extremely difficult. In general, Lagrangian decomposition outperforms Lagrangian relaxation as a solution methodology, yielding higher cost improvements and lower gaps between upper and lower bounds.

An important outcome of this research is that it establishes the importance of multi- plant coordination and provides motivation for further research efforts. We believe that the issues outlined in this dissertation represent an important set of research questions. Given the emphasis on supply chain coordination, this effort relating to multi-plant coordination is timely. A general direction of research is outlined in the research framework presented in Chapter 2. We believe that this framework could serve as a guide for future efforts in this area

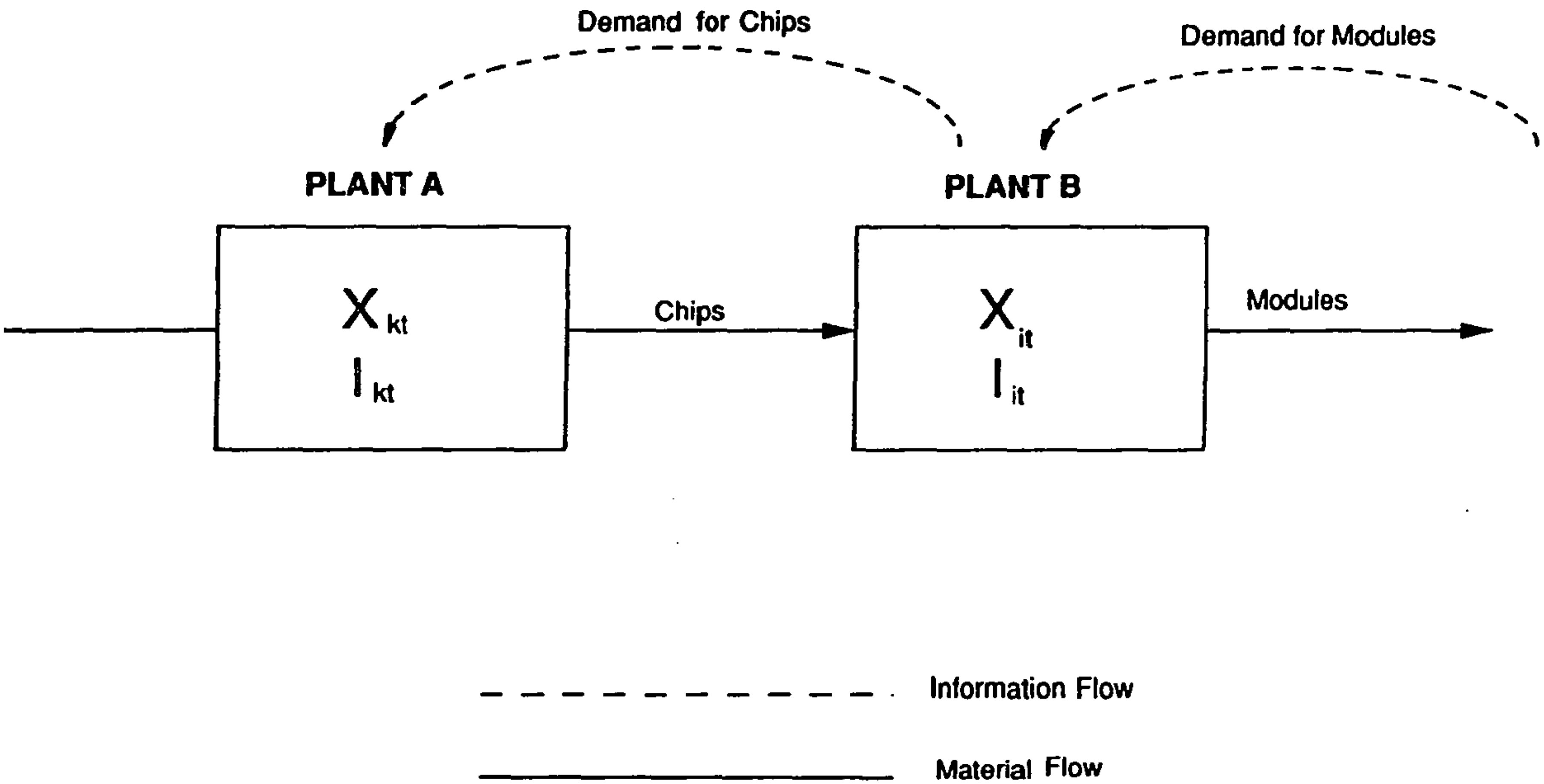

Figure 1.2: Representation of Multi-Plant Coordination Problem

which in turn will help clarify this framework. The alternative methodologies that have been used, help to focus on their relative advantages and disadvantages. The technique used for analyzing the gap between bounds also helps to generate better feasible solutions. In conclusion, this dissertation provides meaningful answers to several aspects of the multi-plant coordination problem. Specifically, we successfully model the critical issues explicit treatment of setup times and inter-plant interaction.

1.2 Summary

Globalization of manufacturing activities has led to intensified competition forcing companies to improve their manufacturing practices. Pressure has mounted on companies to effectively manage their supply chain by reducing inventory at each entity, while providing good service to customers. Vitally important for achieving this objective is coordination, between functions (e.g., manufacturing and distribution) and within functions (e.g., multiple manufacturing plants). A host of important issues at both strategic and operational levels need to be deliberated in order to implement such coordination. Considering all issues in a single model however can quickly overwhelm the decision maker and make the model computationally intractable. Sub-models focusing on subsets of related issues are the answer to this problem. This dissertation is an effort at defining one such sub-model. Specifically, the focus here is on coordination between multiple plants in a multi-plant manufacturing firm. Two important aspects of this work set it apart from previous related research efforts. We model setup times explicitly and also consider the interaction between processes at different plants. Simultaneous treatment of these issues in a single model has not been attempted till now and hence this dissertation bridges a gap in our knowledge of multi-plant coordination. Alternative solution algorithms based on Lagrangian relaxation and Lagrangian decomposition are presented. In general, Lagrangian decomposition appears to be the superior approach providing sharper bounds between the upper and lower bounds. Experimentation on random data provides evidence for the superior performance of our model. Analysis of variance of cost improvement obtained for a variety of experimental factors at different levels establish the robustness of our results.

1.3 Organization

The rest of this dissertation is organized as follows. In Chapter 2 we review a broad segment of literature relevant for studying multi-plant coordination issues with the objective of positioning our research within an overall framework. In Chapter 3 we formally state the assumptions and describe the uncoordinated and the coordinated models for multi-plant operations. Chapter 4 describes our research methodology where we present alternative strategies for relaxing constraints for efficient solution of the coordinated model. An algorithm for obtaining an approximate solution to the coordinated problem is taken up. In Chapter 5 we discuss results and show that the coordinated model provides superior cost performance. Important managerial implications of the research are discussed. Finally in Chapter 6 we present conclusions and outline directions for future work.

Chapter 2

Literature Review

For studying coordination in manufacturing systems we distinguish between two distinct streams in literature. At the most general level, coordination has been studied in terms of integrating decisions in different functions, e.g., facility location, inventory planning and production planning, distribution, marketing etc. We refer to research efforts at this level of functional integration as **general coordination**. At another level, the problem of coordination has been addressed by linking decisions within the same function at different echelons in the organization. A large vertically integrated firm has a hierarchy of production plants making semi-finished products for assembly into final products. Production decisions at these plants must be coordinated, if the firm is to achieve the performance measure targets it has set for itself. In order to be effective, such coordination must take into consideration, the effects of, uncertainty of final demand, uncertainties in production process at each plant, capacity constraints at each plant and the interaction between the plants. We refer to this second level of coordination as **multi-plant coordination**. Each plant here refers to a manufacturing facility that is centred around related production processes. There is considerable overlap and interaction between the areas of general coordination and multi-plant coordination as defined above. However, there is currently no well defined framework in literature that explains such interaction. Moreover, there is no unified body of literature that deals comprehensively with either type of coordination. In this chapter, we review previous research relating to each of these streams. The objective of this exercise is to provide an underpinning for our work in terms of the current state of research. A review will also help us identify different solution methodologies and their relative advantages and disadvantages.

2.1 General Coordination

General coordination can be seen as an attempt by firms to integrate decisions pertaining to different functions, e.g., production and distribution. Research efforts in this area have been directed towards coordinating the operations of firms with a multi-echelon production-distribution structure and measuring the effect of such coordination in terms of the impact on operational performance measures like total cost (setup and inventory holding), overall lead time, average service level etc. We have classified general coordination research into three categories each representing attempts to coordinate different operations of the firm. These categories represent respectively, integration of decision making pertaining to (i) supply and production planning (ii) production and distribution planning (iii) inventory and distribution planning. Table 2.1 presents a classification of the research in general coordination and later we describe the major issues in each category.

Insert Table 2.1 here.

The importance of a coordinated relationship between the supplier and the buyer has been emphasized in literature. As Goyal and Gupta (1989) note, coordination between the supplier and buyer can be mutually beneficial to both.

Supply and Production Planning	Production and Distribution Planning	Inventory and Distribution Planning
Monahan(1984), Rosenblatt and Lee(1985), Lee and Rosenblatt (1986),Banerjee (1986a,1986b), Goyal (1988), Goyal and Gupta (1989),Zoller(1990)	King and Love(1980), Williams (1981), Blumenfeld et al., (1987), Ishi, Takahashi and Muramatsu(1988) Pyke and Cohen(1990a,b), Chandra and Fisher (1992)	Federgruen and Zipkin(1984),Burns et al., (1985),Dror and Ball (1987), Chandra(1990), Anily and Federgruen(1990)

Table 2.1: General Coordination Issues

Studies on coordination between supplier and buyer have focused on determining the order quantity which is jointly optimal for both. Using such an order quantity may lead to increase in overall profits which could be shared in some equitable manner between the two parties. For a given annual supply, vendors are often interested in procuring large individual orders, to benefit from reduced costs of order processing, manufacturing and distribution. However a buyer would want to purchase his optimal order quantity each time since any deviation from this quantity would increase his total cost of inventory holding and ordering. Monahan (1984) proposed a quantity discount model in which he showed that the vendor could increase his profits by enticing the buyer to purchase a greater quantity in return for a discount on purchase price. He showed that by offering a discount to offset increased holding costs, the vendor could motivate the buyer to increase his order quantity by a factor $K = \sqrt{\frac{S_2}{S_1} + 1}$, where S_2 is the vendor's fixed order processing and manufacturing setup cost and S_1 is the buyer's fixed order processing cost. The author assumes a lot for lot policy i.e., the vendor's lot size (for production/procurement) is identical to the order quantity of the buyer. Also, the results suggest that since the buyer is compensated only exactly (and no more) for his increased holding cost (due to higher order quantity) he would be indifferent towards increasing his order quantity. Bannerjee (1986a, 1986b) generalized Monahan's model by incorporating the effect of the vendor's inventory cost which had been ignored by Monahan (1984) in calculating the discount that the vendor must offer to the buyer. Also see Joglekar (1988) for related work.

The assumption of identical lot size for vendor and buyer was relaxed by Rosenblatt and Lee (1985). They showed that the optimal ordering quantity of the vendor is an integer multiple of the buyer's order quantity and proposed a linear discount schedule to determine the optimal pricing policy which the seller could offer to the buyer. Given such a discount schedule, the authors showed that the buyer would optimize his total cost by revising his economic order quantity. This could be exploited by the supplier to increase his profits. This result is stronger than that of Monahan(1984) because there is motivation for both the supplier and the buyer to increase the order quantity. The interested reader is referred to Goyal and Gupta(1989) for a comprehensive review of literature pertaining to buyer-vendor coordination.

It is evident from the above discussion that there are some limitations of the research on supplier-buyer coordination. First, most studies assume that the vendor faces a constant, deterministic demand. Second, the treatment of the production process at the vendor is a gross simplification of the actual situation - single machine, single product and uncapacitated situations. Third, with the emphasis on just-in-time manufacturing, larger purchase quantities in order to get discounts would be difficult to justify as a matter of policy. However it must be noted that just-in-time purchasing can be successful only when demand is stable over time as noted by Karmarkar (1989). In situations where demand is dynamic (which is very often the case in real life) the research direction outlined by Goyal and Gupta (1989) is likely to be useful.

The second type of general coordination research treated in literature is at the level of integrating production planning and distribution planning. Given customer demands over the planning horizon, the decision problem for a production manager is to determine optimal production/inventory levels for all products so that the total cost of setup and inventory holding is minimized over the entire horizon. The distribution manager, on the other hand, must determine schedules for distribution of products to customers so that the total transportation cost is minimized. The decisions relating to production and distribution can be made independently, if there is a sufficiently large inventory buffer to decouple the two. However this would lead to increased holding costs and longer lead times of products through the supply chain. Pressure to reduce inventory and lead times in the supply chain has forced companies to explore the issue of closer coordination between production and distribution. Recently there have been some noteworthy attempts in this area. King and Love (1980) describe the implementation of a coordinated production-distribution system at Kelly Springfield, a major tire manufacturer with four factories and nine major distribution centres located throughout the United States. The authors present a case study describing a coordinated system for the manufacturing plants and distribution centres. The implementation of this system resulted in substantial improvements in overall lead times, customer service and average inventory levels. The annual costs were also reduced by almost $ 8 million. Williams (1981) considered the problem of joint scheduling of production and distribution in a complex network. The performance of a dynamic programming based algorithm was compared to several existing heuristics. The objective of the problem was to minimize average production and distribution cost per period. However the author assumed a constant demand rate limiting the applicability of his work. Pyke and Cohen (1990) have presented a model of a single product integrated production-distribution system with stochastic demand. The authors obtain steady state distributions for key measures of system performance. This work was later extended to cover the multi-product case (Pyke and Cohen, 1990b). However these results do not hold in the case where demand is dynamic. Blumenfeld et al., (1987) considered the problem of synchronized scheduling of production and distribution for a parts producer supplying parts to a final assembly shop. The scenario that these authors considered assumed fixed transportation costs per shipment and one destination per part type. The authors reported the successful implementation of this research at the Delco electronics division of General Motors, that resulted in a 26% reduction in logistics costs. Ishii, Takahashi and Muramatsu (1988) have described a model for minimizing the inventory of **dead stock** or unsold inventory left over at the end of the product life cycle. This issue is likely to become more important product life cycles compress further.

Cohen and Lee (1988) presented the first comprehensive model for coordinating decisions in a supply chain. They proposed a framework for the evaluation of a supply chain in terms of performance attributes like cost, manufacturing lead time etc. Their analytical model seeks to address the following issues :

1. How can production and distribution control policies be coordinated to achieve synergies in performance ?

2. How do service level requirements for material input, work-in-process and finished goods availability affect costs, lead times and flexibility ?

The authors subdivided the problem into four sub-models each corresponding to major physical activities in the supply chain viz., material control, production, finished goods stockpile and distribution. Each sub-model had a cost associated with it comprising setup cost (wherever applicable), holding cost and shortage cost. The objective of the problem was to minimize the total cost over all the sub-models. The sub-models were linked together by means of **local** service targets which were the fill rates that had to be satisfied for each sub-model. An overall optimization model which minimizes the cost over all the sub-models involves a constrained, nonlinear optimization problem and is intractable. Instead the authors used a hierarchical heuristic which decomposes the problem into sub-problems corresponding to each of the sub-models described above. Each sub-problem was optimized separately in a given sequence. The output of a sub-model solution was used as the input data for other sub-problems. This methodology yields an upper bound on the objective of minimum overall cost. The research represents an innovative attempt at integrating several subsystems in a supply chain. Cohen and Lee (1989) consider resource deployment decisions in a global manufacturing and distribution network. Their work addresses issues that are specifically relevant for firms that source material globally. The objective used is to maximize global after tax profits. The model considers variable and fixed costs for procurement, production, distribution, transportation as well as the tariffs, duties and transfer pricing. The authors have assumed a standard fixed transportation cost for transporting items from one place to another. When considering production-distribution systems for bulky items (like petrochemicals) or for items requiring considerable transportation through a distribution network (like automobiles), a more detailed treatment of transportation cost is necessary. One approach is to consider detailed vehicle routing rather than use a fixed transportation cost.

The aspect of integration of vehicle routing and production planning has been analyzed recently by Chandra and Fisher (1992). The authors have developed an integrated model to coordinate production scheduling at a manufacturing plant with the distribution policies to serve a set of geographically disbursed customers. They consider a plant (with a finished goods stockpile) which supplies finished goods to a set of retailers located over a large geographical area. The following trade-offs have to be considered in order to coordinate production and distribution decisions. (i) Large batches to meet production's objective of few setups pushes up the inventory of finished goods at the warehouse. (ii) Consolidating loads of different items to reduce transportation costs requires additional setups or inventory requirements. (iii) Frequent shipments may result in higher transportation costs and increased setup costs although inventory levels may be reduced. The proposed model coordinates the capacitated lot sizing problem (at the manufacturing plant) and the vehicle routing problem (for minimum cost distribution of finished goods to customers). The authors report a reduction in total operating cost, for a range of problem parameters, of 3% to 20% compared to an **uncoordinated** approach where production and distribution decisions are made independently. They also suggest that the benefits of coordination increases as the length of the planning horizon, the number of products & retail outlets, and vehicle capacity increases. It is also found beneficial to coordinate these functional activities where production capacity at the plant is less binding, and distribution costs increase relative to production costs.

A number of authors have also addressed the third type of problem in general coordination i.e., coordinating inventory planning with distribution planning. This problem considers the scenario where a number of customers have to be supplied from one or more warehouse(s). The decision problem is one of deciding the replenishment policy at the warehouse and the distribution schedule for each customer so that the total cost of inventory and distribution is minimized. The trade-off is one of reduction in inventory costs versus an increase in the transportation costs. For example, shipping in smaller quantities and with higher frequency would reduce the inventory level at the warehouse but would entail a higher transportation cost. Federgruen and Zipkin (1984) consider a one warehouse, multiple retailer systems and a single planning period but allow for random demands at the retailers. The authors show that the coordinated model results in substantial cost savings. Bell et al., (1983) developed a computerized multi-period coordinated inventory control/distribution scheduling model. Dror and Ball (1987) and Chandra (1990) have reported heuristic solution methods for coordinated multi period models. Chandra (1990) considers the case where the customers face dynamic demand. The minimization problem is treated over a finite planning horizon of discrete time periods and heuristic solutions are provided. The author compares the results of a coordinated model (warehouse ordering policy and distribution schedules to retailers determined jointly) with a base case where the two decisions are taken independently. The results show that significant savings may be achieved with the coordinated model. This is primarily due to the fact that replenishment at the warehouse occur as close as possible to the transportation. The author allows for products to be shipped to customers before due date (but not later). The findings indicate that the coordinated policy results in cost savings even if the plant incurs the holding cost for goods that were shipped to the customer before due date. This direction of research is important in cases where transportation is a substantial part of the overall cost for the operations of firms.

Burns et al., (1985) developed an infinite horizon coordinated model for the above problem. Anily and Federgruen (1990) modeled this problem in a scenario where one warehouse supplies several geographically dispersed customers. The authors assumed that the customers face a constant demand although the rate could vary from one customer to another. The model is specific to the scenario wherein all customers are divided into various regions. Each time one of the customers in a given region receives a delivery, this delivery is made by a vehicle that visits all other outlets in the region as well. Heuristics are presented for computing the upper and lower bounds on the system wide costs and these are shown to be asymptotically tight as the number of customers increases. Since the proposed heuristics attempt to link two very difficult problems, the above models may be particularly useful in addressing the multi-plant coordination problem which would require a similar linkage. An important extension of the above research could be to study the impact of errors in demand forecast and the role of safety stock at different locations.

In this section we have briefly surveyed models where authors have analyzed the impact of functional coordination on performance measures like cost and lead time. Most of the above models consider multi-item, single stage manufacturing systems and then try to improve the system performance by coordinating different functions. However when the product being manufactured is complex, as for example in the case of computers, telecommunication equipment etc., the processing is often divided between a number of plants. Such an organization of the production process among different specialized plants was called **focused factories** by Skinner (1974). In such cases it is important to consider the problem of coordinating the production plans of the different manufacturing plants. This is the problem of multi-plant coordination and is the focus of the next section..

2.2 Multi-Plant Coordination

As mentioned earlier, the multi-plant coordination problem seeks to link together the production plans of several manufacturing plants which are part of a vertically integrated firm i.e., output from one plant becomes an input into another plant. The objective of such coordination is to achieve near optimal results on performance measures like total cost, manufacturing lead time etc., for the entire organisation. Coordination efforts must model the impact that production planning at one plant has on production planning at another plant. Such models must also take into consideration uncertainties associated with both the demand and the production processes. Published work in this area is scant and to our knowledge not many researchers have addressed the above problem. A possible approach to this problem can be seen in the work of Cohen and Lee (1988). In their work the authors model a serial multi-stage, batch production process. A product is allowed to be processed on more than one line. For each batch of a product processed at a workstation, the authors approximate the total production lead time by the weighted sum of setup times, processing times, material delay times and the waiting times at the workstations. The workstation is treated as a M/G/1 queue and this enables an estimation of the waiting time at the workstation in the spirit of Karmarkar, Kekre and Kekre (1983) and Zipkin(1986). The authors make the approximation that the departure process from one workstation to another is poisson. They also do not consider any capacity limitations on the production line. Nevertheless this study is noteworthy for being the first to describe a comprehensive coordination model.

Beek, Bremer and Putten (1985) have addressed the issue of flexibility and design in multi-level assembly systems. Flexibility can be achieved by cutting down setup costs which reduces the batch size and the assembly lead time. Design issues relate to physical structuring of the assembly network. An industrial application of the model at Philips Industries in Eindhoven is described. The model is useful in comparing different assembly structures for assembling complex products. The batch size calculation takes into consideration the inventory holding costs, setup costs and assembly lead times. This research direction is important as the model coordinates operations at several facilities with the objective of reducing lead time. The authors assume constant demand and future research is needed to extend the findings to other demand scenarios.

Kumar et al., (1990) have considered a variant of the above production planning problem in a supplier-buyer scenario with uncertain but bounded demand conditions. They assume a supply contract wherein the quantity to be supplied in each period is specified in the contract (for N periods) in terms of an upper bound (U) and a lower bound (L). At the beginning of each period the buyer specifies the actual quantity he will purchase and this quantity is contractually obliged to be between U and L every period. In certain cases, describing the demand by only U and L rather than approximating a distribution for the demand amounts to (possibly) ignoring available information. However the authors justify this on the basis that it makes the model much more easy to solve since exact closed form solutions are known to exist for only very simple problems (single period) under the assumption of stationarity of demand. They also assert that in certain industries with short product life cycles, it may be difficult to gather sufficient data to deduce the demand distribution with a high degree of confidence and the demand patterns may show nonstationarity. The value of available information is traded off against the ease of solving the problem. Given this type of requirement specification (in terms of upper and lower bounds) for each future period, the plant manager at the supplier's plant must decide how much to produce in each period so that his total costs are minimized. The model charges holding costs against positive inventories and penalty costs against backorders. A state variable whose value can range between 0 and 1 determines the actual demand for product i in period t. The decision variables are the production quantities of each product in every period. Each decision is evaluated at each set of realized demands determined by the state variable. The objective is to choose the minimum of the maximum costs of all decisions (Minmax problem). The problem considered is a multi-period, multi-product, single stage one with capacity constraints. This research is also dependent on the type of supply contract specified by the authors and therefore the scope of this research is limited to the industries (e.g., semiconductor industry) where such contracts are applicable. Bassok and Akella (1991) have solved the integrated production planning problem for a two plant firm where the downstream plant faces stochastic demand and the yield from the upstream plant is random. The objective is to simultaneously determine the optimal production level , and the raw material order quantity (from the upstream plant) so that total expected cost is minimized subject to capacity and release level constraints. However given the complexity of the problem, these authors make several strong assumptions. It is assumed that all products require only a single raw material which is manufactured at the upstream plant. Also, the time horizon considered for the problem is a single period.

The above discussion brings out one critical issue that needs to be addressed when we attempt to coordinate the operation of multiple plants i. e., the question of lotsizing. Lotsizing in a multi-plant scenario is complicated by the fact that lotsize of a product at one plant will affect the lotsize of all the components that go into this product. One alternative to the above problem is to completely isolate the plants from each other by means of intermediate inventory. However the increased cost of inventory and the increased lead time for products through the supply chain makes this a poor choice. Another approach (already referred to earlier) is the sequential determination of production plans for the plants. Beginning with the plant that supplies the finished goods warehouse, the production plan is prepared and this defines the requirements for the previous plant. This procedure is continued till the production plan for all plants is prepared. Such a procedure ignores the interaction between various plants and will yield sub-optimal production plans. A lotsizing model in a multi-plant scenario must correctly account for the interdependence between different plants. Research in this area can draw upon the substantial literature in lotsizing. Another issue is the actual implementation of the lotsizing algorithms. Most lotsizing algorithms are implemented on a **rolling horizon** basis. This method of resolving the model at the beginning of each time period causes disruption of previously planned production activities which is known as **nervousness**. In order to be effective, multi-plant coordination must consider the impact of nervousness on cost and lead time. Also, consideration of stochastic demand would make it necessary for managers to define safety stock levels necessary to maintain required customer service levels. Effective multi-plant coordination must be able to integrate the issues of lotsizing, nervousness and safety stock into a coherent framework. Attempts to address the multi-plant coordination problem can draw on the existing research on the above questions. We now discuss the major research efforts in the areas of nervousness, lotsizing and safety stock and classify these efforts in order to identify important issues. This is important for integrating the relevant issues so that a framework of multi-plant coordination may be proposed.

2.2.1 Nervousness Issues

An important issue that arises in coordinating the multi-plant structure is the impact of nervousness of demand on total cost and lead time. Nervousness arises due to two reasons. The first reason for nervousness is the **horizon effect**. Schedules are developed on a rolling basis wherein a sequence of production decisions is determined by successive solution of the finite-horizon, multi-period model. The decision for the current period is implemented and as the period elapses, demand for a new period is appended to the horizon and the model is re-solved with the additional information. This may lead to changes in the production plan in a later period which disrupts the schedules made earlier. This may also lead to increase in total cost of operation. The second reason that gives rise to nervousness is the **new forecast** effect. As new and more accurate data regarding requirements in future periods becomes available, it is incorporated in the model to get a new production plan. Both types of nervousness can be disruptive for manufacturing systems. For example, if a schedule revision specifies a setup in a period where no production was planned or considerably alters the production quantity, this will disrupt plans concerning personnel scheduling and machine loading. In a multi-plant structure such disruption can propagate to all plants necessitating frequent revision of production plans. It is therefore important in a multi-plant scenario to choose a strategy that maintains a balance between the disruptive effect of nervousness and the need to respond to new and more accurate information. Research efforts in nervousness pertaining to both horizon effect and new forecast effect are outlined in Table 2.2.

Baker (1977) has studied the efficiency of optimizing a finite-horizon, multi-period model for a single stage production system and implementing those decisions on a rolling basis. Finite horizons are used in production planning because of the limited availability of future demand data and the uncertainty associated with these data. The author focuses on the finiteness of the future information. The motivating question in this study was to find how good were optimal, static decision models for the system when implemented on a rolling basis.Given reliable but limited demand data for future demand, the Uncapacitated Dynamic Lot Sizing (DLS) Model (Wagner and Whitin, 1958) was used to evaluate the cost of implementing rolling schedules. The length of the rolling forecast horizon was varied and the schedule was rolled successively over 48 time periods. The solution of the DLS model for the entire 48 periods (assuming that

	Horizon Effect	New Forecast Effect
Single Stage	Baker(1977), Baker and Peterson(1979), Sridharan et al., (1987)	Carlson et al.,(1979), Blackburn and Millen(1980) Carlson,Beckman and Kropp (1982) De Bodt and Van Wassenhove (1983),Kropp and Carlson (1984), Ho(1989), Sridharan and Berry(1990)
Multi Stage	Chand(1983), Blackburn et al., (1986)	

Table 2.2: Nervousness Issues

this much information is available at the outset) was treated as the benchmark or the optimal solution for making comparisons. The results showed that rolling schedules achieved costs within 10% of optimality. The choice of the most appropriate length of forecast horizon was dependent on whether or not the demand pattern was seasonal. Without seasonality, the best forecast period was found to be the **natural cycle**. The natural cycle is the replenishment interval for the EOQ Model. However, when demand showed a seasonal pattern, the use of multiples of seasonal cycle as forecast horizon was found to be effective. The results of this study imply that the number of periods used as forecast horizon is a crucial parameter if rolling schedules are to be utilized effectively. The choice of a proper length of forecast horizon may be dependent on the demand pattern (whether seasonal or not). Once a good forecast horizon is chosen, the use of rolling horizons in a multi-period dynamic demand model can lead to efficient performance. However the systems considered are very simple - uncapacitated, single stage with no forecast errors.

Baker (1977) also does not consider any specific cost of disruption of schedules. Disruption due to nervousness is clearly undesirable in production systems and has an associated cost. At the same time it may be economically infeasible to ensure perfect stability of plans. In this context, it becomes important to define what is a reasonable level of nervousness. Carlson et al., (1979) address the issue of determining the amount of nervousness that can be deemed **economically tolerable** in a manufacturing system. They suggest that practising managers have a tendency to tolerate non-optimality more than unstable schedules. However it may be a better strategy to strike a balance between the cost of dealing with nervousness (i.e., cost of schedule changes) and the cost of a non-optimal solution (resulting from ensuring stability of plans). Nervousness imposes two kinds of costs : (i) cost of lot size changes for periods in which setups are already scheduled (ii) cost of scheduling setups in periods in which they were not previously scheduled (**new setups**). The authors analyze the effects of only new setups and assume that the changes involved in the first category are far less costly to implement. This view is similar to the suggestions made by Mather (1977). The cost of scheduling a new setup depends critically on the period for which it is scheduled. New setups for the first several periods in the horizon may be impossible to effect due to unbreakable commitments and can be considered to have an infinite cost. On the other hand a new setup near the end of a long scheduling horizon may have a relatively low cost. The authors express the cost function of implementing a schedule change in period k, (v_k), as follows:

$$v_k = \begin{cases} \infty & \text{when k = 1,2,....p} \\ f\{k\} & \text{when k = p+1,p+2,....,r} \\ 0 & \text{when k = r+1,r+2,....,N} \end{cases}$$

where N is the length of the forecast horizon. In the first p periods, no schedule changes are allowed. The authors suggest that a reasonable value for p may be the minimum achievable lead time offset for the item being currently produced. This implies that items have already been released for final assembly and no changes can be made in the schedule. Similarly r might be set equal to the cumulative lead time of all components and raw materials required to produce the item in question. This implies that the product has not entered even the first stage of manufacture and so changes can be made at very low costs. The objective of the model is to minimize the total cost function which consists of setup, holding and schedule change costs.

Minimize $C = \sum_{k=1}^{N} h_k.I_{k+1} + \sum_{k=1}^{N} s_k.\delta(x_k) + \sum_{k=1}^{N} v_k\delta(\delta(x_k) - \delta(\hat{x_k}))$

subject to

$I_{k+1} \; = \; \sum_{j=1}^{k} x_j - \sum_{j=1}^{k} d_j$

where $\delta(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$

$\hat{x_k}$ = Production Lot Size in period k in the existing schedule;

x_k = Production Lot in new schedule (to be determined);

I_k = Beginning inventory in period k

d_k = Amount demanded in period k;

h_k = holding cost per unit of inventory carried into period $k+1$;

s_k = setup cost;

v_k = schedule change cost;

The algorithm is applied to analyze both the horizon effect and the new forecast effect for small problems (6 periods). The authors also perform sensitivity analysis to test the changes in the value of the optimal solution as a result of changes in the values of schedule change costs. This is important for the decision maker because he can be confident that the most imminent scheduling decision is optimal for a range of schedule change costs.

The methodology outlined in the last paper allows the manager to strike a balance between the cost of making schedule changes and the savings that such changes bring about. The result is neither complete dependence on repeated use of the scheduling algorithm whenever new information is obtained nor a striving for stability at all costs. Determining realistic values of p and r as well as the schedule change cost function f{k} is important from the point of view of multi-plant coordination. The important issue here, not considered by the authors, is the relationship of these parameters to capacity. Disruption of planned production needs to be analyzed in terms of the additional demands it makes on the available capacity and the cost of providing additional capacity. Extreme instability of schedules can then be avoided by either freezing the production schedule over a period of time or by limiting the change in schedule to a specified limit. This is common in companies like Toyota (Monden, 1983) which have successfully implemented just-in-time production systems.

The approach suggested above is used by Sridharan et al., (1987) who analyze the effect of freezing a part of the master production schedule (MPS) in order to ensure stability in operations within the context of an MRP system. The authors consider uncapacitated lot sizing decisions using the Wagner-Whitin algorithm. Uncertainty in demand forecast is assumed to be negligible and safety stock is set to zero. Hence only nervousness due to horizon effect (i.e., addition of new periods at the end of the horizon) is considered. The objective of this research is to examine the impact of three important design factors in terms of cost and the stability of the MPS when lotsizing decisions are implemented on a rolling basis. The factors considered are (i) method used to freeze the MPS, (ii) proportion of the MPS that is frozen and (iii) length of the planning horizon for the MPS. Cost is represented by the percentage increase in total cost over the optimal cost. The schedule instability represents the average change in quantity per order over the simulation run and incorporates changes in both quantity and timing of the MPS orders. Two methods can be used to freeze the MPS - specifying the number of periods (for which the MPS is frozen, e.g., 1,2.......N periods in future) or specifying the number of future orders (as determined in the current period) to be frozen. In the first case the schedule rolls over to the next period and the model is re-solved. In the second case the schedule rolls over to the period immediately succeeding the last frozen order. The second factor i.e., proportion of the planning horizon to be frozen, is the ratio of the **freeze interval** (number of periods frozen, K) and the length of the planning horizon (number of periods forming the planning horizon, N). The third factor is the length of the planning horizon (N) and this was expressed as a multiple (K) of the natural cycle (T). The analysis of results shows that cost error becomes significant only when the frozen portion of the planning horizon exceeds 50% of the total planning horizon and that the cost of freezing increases rapidly if more than 80% planning horizon is frozen. The cost error for freezing schedules for the period based model exceeds that for the order based procedure when the proportion of planning horizon frozen is greater than 0.5. This research demonstrates the important effect of freezing the MPS on both the cost performance as well as schedule stability. However the authors consider very simple single stage, single product, uncapacitated systems, which limits the generalizability of the findings.

The above work has been extended to the stochastic demand case in Sridharan and Berry (1990). Given stochastic demand and a required service level, the authors seek to determine the impact of design parameters for MPS freezing on cost and schedule instability under rolling planning. Cost and schedule instability are used in the same context as before. The design parameters considered are (i) the MPS lotsizing method (ii) planning horizon length (iii) frequency of re-planning production schedule (iv) proportion of MPS that is frozen and (v) type of planning information used to freeze the MPS (number of periods versus number of orders). Freezing a portion of the MPS provides a means of stabilizing plant and vendor schedules against nervousness due to demand uncertainty. However freezing the MPS introduces a lag in responding to the changes induced by the uncertainty and this may lead to shortages or excessive inventories at the MPS level. The results indicate that using order based MPS freezing leads to more beneficial results as against period based MPS freezing. Longer freeze intervals produce higher cost errors and reduce schedule instability using either type of freezing. The cost errors become larger as the amount of demand variability is increased. An interesting result is that in order to meet a given customer service level goal, an increased level of safety stock is required under frequent re-planning, producing an increase in the total production and inventory costs and hence the cost error. Another finding pertains to the length of the planning horizon. Longer planning horizons lead to increased cost error and this effect becomes more pronounced as the variability increases. These findings suggest that although lead time considerations may warrant the use of long planning horizons, a reduction in the planning horizon length can lead to more stable schedules and a lower MPS lotsizing cost error when demand uncertainty exists. The results of the study provide an important comparison of MPS freezing techniques under deterministic and stochastic demand conditions. However the study pertains to a simplified single stage, uncapacitated case. An important direction of future research could be to extend the above results to more complex operating conditions.

De Bodt and Van Wassenhove(1983) analyze the impact of forecast errors on total system cost (setup plus holding) under conditions of demand uncertainty. The authors consider the effectiveness of single level lotsizing techniques in a rolling schedule environment with forecast errors. The performance of two well known heuristics (Silver Meal Heuristic (SM) and Least Unit Cost Heuristic (LUC)) is analyzed under a constant demand pattern with normally distributed random errors. The presence of forecast errors leads to more frequent ordering as compared to the case when the demand is known with certainty. The authors show that the cost increase due to forecast errors results in an additional cost of continuously carrying half a period's demand. The major drawback of this study seems to be in its restrictive consideration of single level, uncapacitated systems which limits its applicability. The results are also dependent on the assumption of level demand with normally distributed errors. Nevertheless it is among the first efforts to consider the effect of uncertainty in demand on lot sizing process within the MRP context.

Blackburn et al., (1986) examine the effectiveness of alternative strategies for dealing with the problem of nervousness. The authors compare the relative effect of alternative strategies like freezing the master production schedule, use of safety stock, lot for lot policy, forecasting of demand beyond the planning horizon etc,. The authors suggest an alternative strategy called the **change cost procedure** which is based on the work of Carlson et al., (1979). Each time new information is available, the model is re-solved after modifying the setup cost for each period, depending on whether the item is scheduled in that period or not. The objective is to encourage setups in periods where they are scheduled previously and vice versa. This approach ensures that the schedule will change only when the joint consideration of the setup, carrying and the schedule change costs indicates that it is beneficial to do so.

The above review of literature relating to nervousness issues in production systems underlines some of the basic trends. Current research has focused on the determination of the optimal planning horizon length as the production schedule and its cost has been found to be sensitive to the planning horizon considered. The cost of disruption due to schedule changes has also been addressed although these costs are hard to establish. Since production planning is implemented in multi-plant firms on a rolling basis, nervousness issues need to be addressed in such a scenario. Most authors have focused on single stage, uncapacitated systems. This is unrealistic in real life situations. More realistic systems need to be considered in future research where the propagation of nervousness through several plants can be correctly accounted for. There is also a need to study nervousness in terms of capacity requirements as well as the cost of providing additional capacity. For discrete parts manufacturing firms, setup time needs to be used in the models because in these firms there is no setup cost for items in terms of actual cash outflow but the only cost is in terms of the time consumed. Another direction of research could be to consider bounds on forecast revision. As pointed by Kumar et al., (1990), the supplier-buyer relationship in the semi-conductor industry is coordinated by a contract which specifies the upper and lower bounds on demand. This mechanism is also used in firms in computer industry, operating in a multi-plant scenario. Given such bounds on forecast revision, algorithms need to be developed for guiding managers in the choice of lotsize, so that required customer service is achieved.

2.2.2 Lotsizing Issues

The multi-plant structure is a complex multi-stage manufacturing system. Each plant itself represents a multi-stage system in which the flow of products may be serial, parallel, assembly or general (Billington et al., 1983). Lotsizing is important when the operations of multiple plants is considered under tight capacity constraints. The problem is complicated by the interdependence of different plants. Two distinct issues need to be addressed. First, each individual plant needs to be represented by a simpler but an equivalent system which captures the salient features of the original plant, especially capacity usage. Second, a suitable lotsizing technique needs to be developed which can be applied to the simplified system. These issues can draw on the considerable literature on lotsizing. Lotsizing literature can be broadly classified into two main streams : models which assume that time required for setups is negligible, and models that explicitly consider setup time. We first discuss models that do not consider setup time. A classification of some of the major problem types which have been tackled till now, is presented in Table 2.3.

Insert Table 2.3 here.

The dynamic programming based solution procedure for the uncapacitated, single item dynamic demand situation, proposed by Wagner and Whitin (1958) has served as an important paradigm for lot sizing analyses. Approximate solutions to the uncapacitated, single item, single stage model have been suggested by De Matteis (1971) and Silver and Meal (1973). The major advantage of these approaches is that they are computationally much more efficient than the exact solutions. Zangwill (1969) extended the basic model to include backlogging of demand. However none of the above models takes into consideration the finite processing capacity of the manufacturing facility. The inclusion of this constraint considerably complicates the analysis. Florian and Klein (1971) devised a dynamic programming based shortest path algorithm for the case with constant capacity in every period with concave production and storage costs. The authors showed that the optimal solution to the above problem consists of independent sub-plans wherein the inventory level is non-zero in every period except the last. In the sub-plans the production level, if positive, is at capacity except for at most one period. Love (1973) developed an optimal schedule for the concave cost model with constraints on production and inventory in each period. Using network flow concepts the author showed that for arbitrary bounds on production and inventory there is an optimal schedule such that if for any two periods production does not equal zero or its upper or lower bound, then the inventory level in some intermediate period equals zero or its lower or upper bound. An algorithm for searching for such schedules is provided. Swoveland (1975) developed a shortest path procedure for this problem with piecewise concave production and holding costs.

	Single Stage	Multi- Stage
A. Single Item		
Exact Approaches	Uncapacitated Wagner and Whitin(1958), Zangwill(1969)	Uncapacitated Zangwill(1969),Love(1972) Afentakis et al.,(1984)
	Capacitated Florian and Klein(1971), Jagannathan and Rao(1973), Love(1973), Swoveland (1975), Baker et al., (1978) Lambrecht and VanderEecken (1978b),Chung and Lin(1988)	Capacitated Lambrecht and VanderEecken(1978a)
Approximations	Uncapacitated De Matteis(1971), Silver and Meal(1973)	Uncapacitated Berry(1972), More(1974), New(1974), Coleman and Mcknew(1991)
	Capacitated Bitran and Matsuo(1986)	Capacitated
B. Multi-Item		
Exact Approaches	Uncapacitated	Uncapacitated Zangwill(1966),Crowston and Wagner(1973),Crowston et al.,(1973),Steinberg and Napier(1980)
	Capacitated Barany et al.,(1984),Van Roy and Wolsey(1987),Pochet and Wolsey(1991)	Capacitated Bassok and Akella(1991)

Table 2.3: Lotsizing Issues

	Single Stage	Multi-Stage
Approximations	Uncapacitated	Uncapacitated
		Blackburn and Millen(1982)
	Capacitated	Capacitated
	Manne(1958),Dzielinski and Gomory(1965), Lasdon and Terjung(1971),Lambrecht and Vanderveken(1979),Dixon Silver(1981),Thizy and Van Wassenhove(1985),Karmarkar and Schrage(1985),Eppen and Martin(1987), Trigeiro et al.,(1989),Lozano et al., (1991),Diaby et al. (1992a,b)	Billington, McClain and Thomas(1983), Zahorik, Thomas and Trigeiro(1984)

Table 2.3 Lotsizing Issues (contd.)

Jagannathan and Rao (1973) consider the above production planning problem for a generalized cost function with bounds on backlogging, inventory and production capacity. Baker et al., (1978) present a tree search capacity constrained dynamic demand problem. Lambrecht and VanderEecken (1978b) also present a model for the capacity constrained lotsizing problem with different production cost and holding/shortage cost functions than those used by Baker et al., (1978). Barany et al., (1984) solved the multi-item capacitated lotsizing problem to optimality by adding strong valid inequalities. Also see Van Roy and Wolsey (1987), Pochet and Wolsey (1991) and Magnanti and Vachani (1990) for related work.

The simplest version of the multi-item, single-level capacitated dynamic lotsizing problem (MISLCLSP) consists of scheduling N items over a horizon of T periods such that demands are fulfilled without backlogging. The objective is to minimize the sum of setup costs and inventory holding costs over the horizon subject to constraints on total capacity in each period. These algorithms are solved on a rolling basis (Baker, 1977). MISLCLSP is NP-hard since the single item version of this problem is known to be NP-hard (Florian et al., 1980). Most of the optimal approaches discussed above for capacitated, single stage lotsizing models have been tested only on small to medium sized problems and analysis of these algorithms indicates that running times would increase substantially for real life problems. Therefore efficient heuristics have been found to be necessary for solving larger problems. Heuristics for solving MISLCLSP can be divided into **common sense heuristics** and mathematical programming based heuristics (Maes and Van Wassenhove, 1988). Common sense heuristics comprise three main steps : (i) A batching step in which available capacity is first allocated to fulfil demand for that period (for all items) and then, balance capacity, if any is used to produce a batch of items for a future period. The decision to produce an item for a future demand is based on a priority index which trades off the savings in setup cost against the increase in holding cost. (ii) A feasibility routine which ensures that a feasible solution is obtained when the algorithm stops. (iii) An improvement step to perturb the solution slightly in order to make additional savings. Examples of such heuristics can be found in Eisenhut (1975), Lambrecht and Vanderveken (1979), Dixon and Silver (1981), Maes and Van Wassenhove (1986a,b), Dogramaci et al., (1981), Karni and Roll (1982) etc. As Maes and Van Wassenhove (1988) note, the above heuristics have important shortcomings. While they yield reasonable cost performance on the average, in certain specific situations (e.g., seasonal demand, very tight constraints etc.) they can lead to substantially poor results compared to the mathematical programming based heuristics. These heuristics cannot be applied in situations where there are several constrained resources. They are also less flexible in the sense that they are designed for specific problem instances and the heuristics will not be applicable to changed problem characteristics (e.g., addition of constraints on product life, aggregate inventory levels etc.).

The second type of heuristics that have been proposed for solving MISLCLSP are the mathematical programming based approaches. These heuristics relax the capacity constraint and solve the resulting N independent sub-problems (one for each item) efficiently by Wagner-Whitin algorithm. This solution is likely to violate the capacity constraint in the original problem. Different strategies have been used to perturb the solution so that it is feasible in the original problem. For periods where capacity is violated, Newson (1975a, 1975b) forced the production of one item at a time to zero (by assigning it infinite setup cost) and then found the best way of scheduling this item in the rest of the periods using Wagner-Whitin algorithm. The item with the least cost for shifting was moved to the period shown by the Wagner-Whitin schedule. This procedure was repeated till all infeasibilities were removed. The problem with this procedure is that it may fail to find a feasible solution. Thizy and Van Wassenhove (1985) suggested a Lagrangian Relaxation based procedure for solving the above problem. Lagrangian Relaxation of the capacity constraints yielded a lower bound to MISLCLSP. Fixing the setups given by this solution, the authors solved a transportation problem to determine a primal feasible solution which was an upper bound to the above problem. The lagrangian multipliers were updated using subgradient optimization (Held Wolfe and Crowder, 1974) and the above procedure was repeated for a pre-determined number of iterations. Computational results have been presented for small problems. A shortcoming of the above approach is that the solution time for the transportation sub-problem increases rapidly with increase in the number of periods and therefore it will be difficult to use for real life problems.

Eppen and Martin (1987) used a variable redefinition approach for solving MISLCLSP. They developed an alternative equivalent formulation to the classical formulation of MISLCLSP which had the property that its LP Relaxation had a value equal to the Lagrangian dual with respect to the capacity constraints. Bitran and Matsuo (1986) have studied approximate formulations for the above problem. They proposed heuristics for two alternative forms of the above problem and showed that under mild conditions of forecast error these forms of the problem are equivalent to the original problem. The heuristic procedure proposed by the authors is shown to be pseudo polynomial.

All the above models consider setup times indirectly by using a constant setup cost. As noted earlier, an important aspect of discrete manufacturing systems is that the cost associated with setting up a machine is often negligible. The cost of a setup is essentially the time consumed which reduces the productive capacity of the facility. Since the cost of the resource consumed depends on the actual level of resource usage, constant setup cost is an inaccurate representation of the production system.

The earliest work on the lotsizing problem with explicit treatment of setup times was by Manne (1958) who used a linear programming based approach. Manne showed that the linear programming solution to the problem provides a good solution whenever the number of items is large compared to the number of periods in the planning horizon. His work was extended by Dzielinski and Gomory(1965) using approximations based on Dantzig-Wolfe decomposition(1960) and by Lasdon and Terjung(1971) using generalized upper bounding procedure. Billington et al.,(1983) suggested a Lagrangian heuristic based on the idea of **product compression.** The authors suggest that in most production systems there are only a few **constrained workcenters** i.e., workcenters where capacity is likely to be a binding constraint so as to cause scheduling difficulty. Lotsizing is critical only for the constrained workcenters and the other workcenters can be scheduled on a lot for lot basis. The proposed heuristic can handle setup times. However little experimentation is reported. Trigeiro et al.,(1989) addressed the issue of single stage capacitated lotsizing with explicit treatment of setup times. They provided interesting insights into the difficulty of the problem. They showed that when setup times are considered, even the problem of determining whether a feasible solution exists, is NP complete. The authors considered nonstationary costs, demands and setup times. Lagrangian relaxation of the capacity constraint led to decomposition into a set of uncapacitated problems solvable by dynamic programming. The Lagrangian dual costs were updated by subgradient optimization and a heuristic routine was used to generate a feasible solution at each iteration of the algorithm. The procedure was terminated when the duality gap between the Lagrangian solution (lower bound) and the feasible solution (upper bound) were within specified limits or after a specific number of iterations. This study represents an important departure in that it considers production systems and demand patterns often found in real life. While it has been established that the problem is very difficult to solve, other models (as for example those that use setup costs as a proxy for setup time) solve a less realistic version of the problem. Several important extensions need to be established for the above problem. The relative importance of the parameters needs to be established to identify which ones have maximum impact on cost. Such studies would suggest a direction for the future improvement of production systems. The work of Trigeiro et al., (1989) considers single stage systems. Unfortunately most real life manufacturing systems are multi-stage. Lozano et al., (1991) have also addressed the lotsizing problem with setup times using the primal-dual method. The algorithm was shown to have monotone and finite convergence properties. The authors showed that this method yielded better results than the subgradient method although it required greater CPU times. Lagrangian relaxation based approaches for large systems have been reported by Diaby et al., (1992a, 1992b). Also see Lozano et al., (1991) for related work.

Models for multi-stage systems have been proposed both within the MRP framework and as general production models. Zangwill (1969) and Love (1972) have developed efficient dynamic programming based algorithms for uncapacitated serial systems. Love (1972) shows that the optimal policy is nested for concave production and storage costs if storage costs are non-decreasing in order of facility and production costs are non-increasing in time. Nested policy implies that if a facility orders an item in a particular period, all downstream facilities also order in that period. Crowston et al., (1973) consider the issue of lotsizing in multi-stage assembly systems. The authors propose a dynamic programming algorithm for such a system with constant, continuous final product demand and infinite planning horizon. They show that under the assumption of time invariant lot sizes, the optimal lot size at each facility is an integer multiple of the lot size at the successor facility. Steinberg and Napier (1980) develop an optimal procedure for the multi-period, multi-product, multi-level lotsizing problem by modelling the system as a constrained generalized network problem with fixed charge arcs and side constraints. The resulting minimum cost flow problem yields optimal lotsizing decisions at all levels. Blackburn and Millen (1982) address the issue of lotsizing in multi-stage material requirement planning (MRP) systems. The authors sequentially apply a single stage heuristic to each stage with a set of modified setup and holding costs to account for the inter-dependencies among stages. The modification of the setup cost at each stage reflects the incremental cost across all stages of a setup at this stage. Similarly, the modified holding cost reflects the incremental cost of holding an extra unit at this stage, at all other stages. The intention here is similar to transfer price mechanism i.e., to achieve coordination of decisions at different stages of the process without a central planning unit dictating production schedules.The advantage of this kind of an approach is that it retains the simplicity of a single stage lotsizing algorithm and gives reasonably accurate results for products which do not have many levels in the bill of materials. However deviation from optimality increases as the number of levels in the product structure increase. Also the authors consider uncapacitated production systems at all stages. The results of the heuristic are compared to optimal results obtained by solving the multi-stage lot sizing problem to optimality. The results show that the deviation from optimal results increases as the number of levels in the product structure increases. The advantage of this approach is that it retains the simplicity of single stage lotsizing algorithms and gives reasonably accurate results for product structures that do not have many levels. Capacity constrained extensions of the above model for general product structures are likely to be useful, given the intractability of optimal algorithms for multi-stage lot sizing. Billington et al., (1983) have introduced the idea of **product structure compression** which has the objective of reducing the size of the problem while retaining the salient features of the problem in terms of demand, cost, lead times and capacity requirements. The authors suggest that in most production systems there are only a few **constrained facilities**, i.e., workcenters where capacity is likely to be a binding constraint so as to cause scheduling difficulty. Lotsizing is critical only for the constrained workcenters and other workcenters can often be scheduled on a lot for lot basis. Karmarkar et al., (1992) concur with this view when they talk about **approximate composite model** to represent a manufacturing system.

The above discussion on lotsizing touches on one area in this multi-issue domain. From the point of view of multi-plant coordination, there are two promising directions for research. The first area to focus on is to work on approximate representations of single plants in the spirit of Billington et al., (1983) such that the number of workcenters representing a plant are reduced considerably. This representation must be able to capture the salient features of the original system. The objective here is to reduce a complex manufacturing system into its most critical constrained facilities. Once this is achieved, it may be easier to use this approximate representation of a plant in a multi-stage lotsizing algorithm where each stage is an individual plant. The second area is the development of robust heuristics which capture the interaction between the plants. Consideration of setup times in the spirit Trigeiro, McClain and Thomas (1989) is an important criterion in discrete parts manufacturing systems. Clearly this is a very difficult problem to solve optimally. However good heuristics would help quantify the benefits of coordination as compared to the current practice of optimizing the objective plant by plant which ignores the inter-linkages between plants. One more direction needs to be investigated in this context. Usually, firms establish operational performance targets for measures like lead time at a higher strategic level of decision making, taking into consideration, the firm's priorities, competitive environment and industry norms. A two level procedure may be envisaged in such a situation where operational performance measures targets have been defined. First an optimization based heuristic is solved to determine lotsizes and these are plugged into a detailed simulation to check whether the targets are achieved.

2.2.3 Safety Stock Issues

In the previous subsection we discussed models for lotsizing in multi-stage production systems. Most of these models have assumed deterministic demand. However this is rarely true in real situations. The use of safety stock is widely prevalent in industry to counter variability that may be present in various forms, e.g., variability in demand forecast, variability in processing time or yield, variability in vendor replenishment time and quantity etc. Products also have to compete for limited processing time and resources at each stage. Consequently the manufacturing system does not have full flexibility to reschedule to combat the above forms of variability. Setting of safety stock in a multi-plant scenario is further complicated on account of the inter-dependencies which exist between plants. An important issue that arises in this context is the determination of safety stock at each plant if the firm has to achieve a pre-specified customer service level. As we pointed out in our discussion of nervousness, firms with a multi-plant structure often consider bounds on forecast revision for each plant, to limit schedule nervousness. This maybe treated as a limit on the flexibility of the plants to adjust their level of production to changing forecasts. If however the firm wishes to provide a better service level, use of safety stock may be necessary. In this subsection we briefly review research on safety stock from the point of view of multi-plant coordination. The interested reader is referred to Graves (1988) for a more complete review of safety stock in manufacturing systems. A classification of the work on safety stock is presented in Table 2.4.

Clark and Scarf (1960) presented an optimal inventory policy for a serial system with stochastic demand. The authors assumed a linear processing cost and a linear inventory holding cost. No ordering costs are considered. The objective used was to minimize the expected discounted costs. The optimal policy is computed by solving a series of one stage inventory problems.Beginning with the last stage the optimal policy is computed under the assumption that sufficient input is available from the previous stage. From this optimal policy for the last stage, the authors then determine the costs imputed on the downstream stage by a stockout at the upstream stage. This analysis is successively carried on to the

	Uncapacitated	Capacitated
Exact Analyses	Clark and Scarf(1960), Schmidt and Nahmias(1985)	
Approximations Without Lotsizing (i.e., lot for lot)	Simpson(1958),Hansmann (1959),Miller(1979), Wijngaard and Wortmann(1985) Graves(1988)	
Approximations With Lotsizing	Clark and Scarf(1962), Lambrecht et al., (1984), Carlson and Yano(1986),Yano and Carlson(1985,1987)	Lambrecht et al.,(1985)

Table 2.4: Safety Stock Issues

upstream stages.The model has been extended to the case of an assembly system (two components with differing lead times being assembled into a single end item) by Schmidt and Nahmias (1985). However this modest change in structure makes the analysis considerably difficult. It would therefore appear that extension of the above models to general structures would be difficult. Given the complexity of exact analysis, the focus should largely be on good heuristics.

Approximation models for setting safety stock in manufacturing systems fall into two categories. The first category is where lotsizing is not considered i.e., lot for lot policy is followed with each stage ordering in every period. For this lotsizing policy, Simpson (1958) argues that planning must be done for a maximum reasonable demand which has been pre-specified. Each stage must be able to always fulfil the request of the downstream stage under such demand conditions and safety stock must be planned likewise. The inherent assumption here is that when an extraordinary demand situation arises, the system will take extraordinary actions (like expediting etc.). Hence the manager only needs to plan safety stock for satisfying the maximum reasonable demand. This idea is in consonance with the idea of a bound on forecast revision that came up in our discussion on nervousness. The authors show that the optimal policy is an **all or nothing** policy i.e., either there is no inventory between two stages or there is sufficient inventory to completely decouple the two stages. Hansmann (1959) considers a similar problem except that he assumes that there can be a delay in supplying the demand of the downstream stage. The poorer the service provided by the upstream stage, the longer will be the replenishment lead time for the downstream stage and more excess inventory will be needed. In this case the optimal policy turns out not to be an all or nothing policy. Miller (1979) introduced the concept of hedging which consists of inflating the master production schedule to reflect the uncertainty in the end item demand.

The second category of approximation methods for setting safety stocks consists of those models that consider lotsizing. Clark and Scarf (1962) extend their previous work to allow for a fixed ordering cost at each stage. Their model computes a (s, S) policy for each stage with successive stages being linked by a penalty cost of a stockout. Lambrecht et al., (1984, 1985) extend the above analysis to assembly systems and to the case where downstream lotsizes are greater than the upstream lotsizes. The authors suggest that protection against uncertainty may take the form of either safety stocks or safety time. Safety time is the time between the production batch or procurement lot becoming available and the time when it is needed to produce some subsequent assembly or finished product. The authors use a Markov Decision Process analysis to provide insight into near optimal policies for the above systems.

Carlson and Yano (1986) address the issue of determining cost effective safety stock levels for each item in the product structure under stochastic demand. A single product is considered which is assumed to have a multi-level assembly structure. The objective is to minimize average total setup and holding costs per period subject to achieving a specified customer service level. The problem is a general nonlinear (non-convex) stochastic integer optimization problem and is computationally intractable. The authors used a heuristic approach and proposed upper and lower bounds on the optimal solution. The results indicate that safety stocks can be utilized beneficially at production stages where setup and disruption costs are high. This has important implications for the multi-plant scenario where we need to determine the amount of safety stock that must be carried at each plant. However, this study ignores capacity constraints. Consideration of capacitated systems is specially relevant for multi-plant coordination. Research needs to be done to establish how capacities of different plants affect the need and level of safety stock for achieving a given customer service level.

Graves (1988) addresses the issue of planned lead time for each stage which serves as the target figure for this stage. The author suggests that the greater is the planned lead time, smoother is the aggregate output and hence lower is the required production system flexibility. However with higher planned lead time, the inter-stage and intra-stage inventory also increases. Therefore the trade-off examined is between inventory cost and production flexibility. Should we eliminate the need for production system flexibility by having long lead times and as a consequence carry greater amount of work in process inventory. Or alternatively should the production system be designed to be flexible in terms of varying load from period to period. The current emphasis on reduced inventory and lead time clearly makes the first alternative unattractive to implement. The issue of capacity is once again ignored. Research needs to be done to link planned lead times, setup times and capacity of equipment.

The above discussion brings up some of the important issues relating to safety stock that need to be addressed for multi-plant coordination. First there is need for establishing the concept of a maximum reasonable demand. This is the maximum level of demand for which safety stock planning needs to be done. This concept needs to be integrated with the idea of limited production flexibility as used by Graves (1988) i.e., the limited ability of a plant to adjust its capacity to changing forecasts. This limit could conform to the forecast revision bounds that have been discussed in the subsection on nervousness. This limit on forecast bound represents the flexibility of the system to adjust its output to match changed forecasts. Safety stock needs to be planned only for demand beyond this bound. Research is needed to clarify the above issues. There is also need to establish inter-linkage between plant capacity and the need for safety stock.

2.3 Summary of Literature Review

In this chapter, we have reviewed research relating to three issues that are critical for multi-plant coordination. Some trends are evident from the above review.

Research on nervousness has largely focused on single stage, uncapacitated systems. This represents an important gap in our knowledge about the phenomenon of nervousness in production systems. In multi-plant firms, there is limited flexibility to respond to changed forecasts due to finite capacity and the dependence of the downstream plants for components or subassemblies on the upstream plants. These issues are important for multi-plant coordination and need to be addressed in future research.

Research on lotsizing is extensive and multi-faceted. However, some important aspects of production systems are relatively less well represented. The first aspect is the explicit treatment of setup times which makes the capacitated lotsizing problem much more difficult to solve. This issue need to be addressed, especially for production systems with tight capacity constraints where setup time may have a significant impact. The second point relates to the interdependence of the production planning process at different plants on one another. This issue assumes greater significance when forecasts are inaccurate and an upstream plant may fail to provide the required inputs if the revised forecasts are substantially larger than the previous requirements.

Research on safety stock has addressed several important issues. However, many question remain to be answered. The concept of a **maximum reasonable demand** needs to be clarified for planning for safety stocks. Work is also needed for the case where the demand process is not fully known, as is common in the industries that we are concerned with in this research.

2.4 Framework for study of Multi-Plant Coordination

Before articulating the specifics of a model for multi-plant coordination, it would be useful to first establish a broad framework which will serve as a guideline for research. This has two important purposes. First, this will aid us in the modelling process for the current research. Second, it will serve as a guideline for related future efforts on multi-plant coordination. Evidently, the study of multi-plant coordination embraces several varied aspects and it would not be possible to address all concerns comprehensively in the current research alone. We hope, however, that this framework will motivate other similar efforts in this area which will in turn help to elucidate the proposed framework.

As has been previously argued, decision problems relating to lotsizing, nervousness and safety stock determination are critical to studying multi-plant coordination and these need to be integrated in an appropriate unified model. However, a large problem incorporating all these issues could easily become intractable. We therefore propose the following hierarchical approach. The determination of lotsizes is of primary importance and must be addressed first. The important aspects that this problem must be able to capture are explicit treatment of setup times and interdependence between different plants. Each plant must be modeled by minimal number of workcenters. If a well defined bottleneck workcenter exists, it becomes the natural choice for modelling the entire plant. If there is no well defined bottleneck, the workcenter could represent what Karmarkar et al., (1992) call an **approximate composite model** which approaches the performance of the entire facility. For this purpose constrained facilities need to be identified in the spirit of Billington et al., (1983). We need to distinguish here between the **macro** and **micro** levels of modelling. At the macro level, the approximate workcenters (representing the entire plant) need to be considered.

The results from this model may then be plugged into a detailed micro model to test if the performance targets are being met. The micro model would be simulation based and would include most of the shop-floor details which have been ignored in the macro model.

Once the above situation is successfully modeled, the issues of implementation of the above scheme on a rolling basis as well as the determination of appropriate safety stock need to be operationalized. For this there would be need to define (as in Graves , 1988) the concept of a maximum reasonable demand which is the maximum level of demand for which safety stock planning needs to be done. This concept will need to be integrated with the idea of limited production flexibility or the limited ability of a plant to adjust its capacity to changing forecasts. The limits could conform to the forecast revision bounds that have been discussed in the section on nervousness. This limit on forecast bound represents the flexibility of the system to adjust its output to match changed forecasts. Safety stock would need to be planned only for demand beyond this bound.

In this dissertation, we focus primarily on the macro level model relating to losizing issues in order to address the twin concerns of setup time and interdependence between different plants. In the next chapter, we will address the above issues when we formally define our model for multi-plant coordination.

Chapter 3

Model for multi-plant coordination

Figure 1.2 is a representation of the multi-plant coordination problem which, as earlier stated, is motivated by a real life situation at IBM. Our model is a combination of the manufacturing system at IBM and the framework outlined in the previous chapter. The basic details of the IBM organization are as follows. The upstream plant, A (located in Burlington, Vermont), produces semiconductor devices called **chips**. The chips are then transported to Plant B (located in Bromont, Quebec) where they are mounted on ceramic substrate called **modules**. In terms of the plant structure, we consider the two plant case as depicted in Figure 1.2 as this adequately demonstrates the phenomenon of interaction between plants. We shall show subsequently that the extension to the more general n plant case is straightforward. Within each plant, we consider the case of a single bottleneck workcenter. This choice is guided by the fact that a model with a single workcenter is an important building block for the more realistic model with multi workcenters at each plant. Later in this dissertation, we will propose ways in which this more complex modelling can be approached.

The product structure adopted, mimics reality and allows for **chip commonality** i.e., a chip is used in several different modules. Similarly, a module may require several different chips. In keeping with our framework described earlier, we focus our attention on the deterministic case, where the demand process for modules can be perfectly predicted over the forecast horizon and shortages are prohibited. However the demand is allowed to be dynamic in time which is an important aspect of the business environment in the computer industry.

At IBM (as is the case in most discrete parts manufacturing industries), productive capacity is expended on two distinct activities. Setting up a machine requires a setup time which is independent of the batch of items being processed while a standard time per unit is required for processing of items. The interested readers can find details of the chip making process in Cooper et al., (1992) and of module fabrication in Chandra and Gupta (1993). Due to the impact on problem complexity, a vast majority of previous research efforts have ignored setup times by using a constant setup cost as a surrogate. In contrast, we follow the lead of Trigeiro et al., (1989) and others and explicitly model setup times as this is a more accurate representation of the problem. Setup costs are not considered because the productive capacity lost in setting up the machine is the sole cost implication of setups in discrete parts manufacturing industries. If required however, setup costs can be included in our model with no impact on the problem complexity. Capacity is assumed to be limited, i.e., the total requirements for setup and processing times for all items must be within the specified regular capacity in each period. However in case of need, limited overtime resources can be used. For example, IBM uses five week days as regular capacity, while the two weekend days are treated as overtime. Two types of costs are incurred for use of overtime resources. A fixed cost is incurred in each period in which overtime usage is positive. This reflects the fixed portion of costs relating to scheduling of overtime work. A variable cost proportional to the amount of overtime usage is also incurred. Holding costs are incurred on end of period inventory while production costs using regular resources are assumed to be constant in all periods and are ignored.

The cost of transportation of chips between Plant A and B is captured implicitly in our model due to the uniqueness of the problem. The main mode of transfer of chips between Burlington and Bromont is by trucks which operate every weekday. Transportation cost incurred during weekdays is largely fixed and can be ignored. If however, there is work over the weekend (i.e., overtime), extra transportation has to be scheduled for transferring chips without delay to Bromont. Once the decision to schedule overtime is taken, the additional transportation costs are largely fixed. The implication of this for our model is that the fixed overtime costs at Plant A need only be inflated by the fixed transportation costs between Burlington and Bromont to capture the effect of transportation during overtime. This is easily accommodated in the structure of our overtime costs.

Given demand forecasts over a planning horizon, the objective of the multi-plant coordination problem is to determine production/inventory levels at each plant so that the overall costs of inventory holding and overtime (including transportation) are minimized. Our focus on this set of costs was corroborated in our discussions with managers at IBM.

We now proceed to define the notation which will be used throughout, in the rest of this dissertation.

Notation

$$i = \begin{cases} \text{index for chips} & = 1, \ldots, \text{m} \\ \text{index for modules} & = \text{m+1}, \ldots, \text{m+n} \end{cases}$$

t = index for time periods = 1-L, ...,T

p = index for plants = 1 ... Plant A

= 2 ... Plant B

Decision Variables

X_{it} = Quantity of product i processed in period t

I_{it} = Inventory for product i in period t

O_{pt} = Amount of Overtime used at Plant p in period t

$$Y_{it} = \begin{cases} 1 & \text{if product } i \text{ is produced in period } t \\ 0 & \text{otherwise} \end{cases}$$

$$z_{pt} = \begin{cases} 1 & \text{if overtime is used at Plant } p \text{ in period } t \\ 0 & \text{otherwise} \end{cases}$$

Parameters

d_{it} = Quantity of module i required in period t

u_{ik} = number of units of chip k required for producing each unit of module i

h_i = holding cost per unit of product i per period

s_i = setup time for product i

b_i = processing time per unit of product i

Cap_{pt} = available regular capacity at Plant p in period t

v_{pt} = variable cost of overtime at Plant p in period t

f_{pt} = fixed cost of overtime at Plant p in period t

OT_{pt} = maximum allowable overtime at Plant p in period t

M_{i2t} = $Min\ ([(Cap_{2t} + OT_{2t} - s_i)/b_i],\ \sum_{t'=t}^{T} d_{it'})$

M_{i1t} = $Min\ ([(Cap_{1t} + OT_{1t} - s_i)/b_i],\ \sum_{t'=t}^{T} \sum_{k=m+1}^{m+n} u_{ki}.d_{it'})$

3.1 Uncoordinated Model

As mentioned earlier, the production planning problem in a multi-plant scenario is typically addressed sequentially or in an **uncoordinated** manner. This is a direct consequence of the way in which managerial roles are defined in the North American industry. Managers focus on their immediate responsibilities and since this is the basis of their performance appraisal, there is little motivation to communicate with the interlinking parts of the organization, either upstream or downstream. Operationally, this translates into the following scenario. The downstream plant first solves its own production planning problem and determines the production plan for modules. This defines the demand vector for chips which is passed onto the upstream plant. Using this demand as input, the problem for the upstream plant, A, is then solved independently. Formally the problem for Plant B may be stated as follows :

Plant B

Minimize

$$\sum_{i=m+1}^{m+n} \sum_{t=1}^{T} h_i.I_{it} + \sum_{t=1}^{T} v_{2t}.O_{2t} + \sum_{t=1}^{T} f_{2t}.z_{2t}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n \tag{3.1}$$

$$\sum_{i=m+1}^{m+n} (s_i.Y_{it} + b_i.X_{it}) - O_{2t} \leq Cap_{2t} \qquad \forall t = 1,\ldots,T \tag{3.2}$$

$$O_{2t} \leq OT_{2t}.z_{2t} \qquad \forall t = 1,\ldots,T \tag{3.3}$$

$$X_{it} \leq M_{i2t}.Y_{it} \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n \tag{3.4}$$

$$z_{2t}, Y_{it} \in (0,1), \quad X_{it}, I_{it} \geq 0, \quad I_{i0} = 0 \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n \tag{3.5}$$

Constraints (3.1) are the inventory balance constraints which ensure that demand for modules is satisfied either through production in the current period or by using inventory carried over from a previous period. The capacity constraints (3.2) impose the restriction that the total of processing and setup time for all modules scheduled in a period must be within the regular and overtime capacity available at Plant B in that period. Constraints (3.3) limit the amount of overtime resources that can be used and ensure that appropriate fixed costs of overtime are incurred in each period in which overtime usage is positive at Plant B. Constraints (3.4) are the setup enforcement constraints which ensure that if a module is produced in a given period, setup time will also be incurred for this module in that period. Constraints (3.5) impose non negativity on the production and inventory variables, force the setup and fixed overtime variables at Plant B to be binary, and the starting inventory of modules in period 1 to be zero for the appropriate indices. The objective is to minimize the total costs at Plant B, comprising inventory holding costs and the fixed and variable costs of scheduling overtime over the entire horizon.

The solution of the above problem for Plant B defines the chip requirement $\sum_{i=m+1}^{m+n} u_{ik}.X_{it}$ for every chip k in each period from Plant A which uses this as an input to determine its own production plan. A constant lead time of L periods is assumed between the two plants. The planning horizon for Plant A is therefore $1-L$ to $T-L$. The problem for Plant A may be stated as follows :

Plant A

Minimize

$$\sum_{i=1}^{m}\sum_{t=1-L}^{T-L} h_i.I_{it} + \sum_{t=1-L}^{T-L} v_{1t}.O_{1t} + \sum_{t=1-L}^{T-L} f_{1t}.z_{1t}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = \sum_{i'=m+1}^{m+n} u_{i'k}.X_{i't} \qquad \forall t = 1-L,\ldots,T-L, \quad i = 1,\ldots,m \qquad (3.6)$$

$$\sum_{i=1}^{m}(s_i.Y_{it} + b_i.X_{it}) - O_{1t} \leq Cap_{1t} \qquad \forall t = 1-L,\ldots,T-L \tag{3.7}$$

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1-L,\ldots,T-L \tag{3.8}$$

$$X_{it} \leq M_{i1t}.Y_{it} \qquad \forall t = 1-L,\ldots,T-L, \quad i = 1,\ldots,m \tag{3.9}$$

$$z_{it}, Y_{it} \in (0,1), \quad X_{it}, I_{it} \geq 0, \quad I_{i0} = 0 \qquad \forall t = 1-L,\ldots,T-L, \quad i = 1,\ldots,m \tag{3.10}$$

Constraint (3.6) is the inventory balance constraint at Plant A, which ensures that demand for chips is satisfied either through production in the current period or by using inventory carried over from a previous period. Constraint (3.7) imposes the restriction that the total of processing and setup time for all chips must lie within available regular and overtime capacity at Plant A in each period. Constraint (3.8) ensures that in each period with positive overtime usage, fixed overtime costs are incurred. Constraint (3.10) is the setup enforcement constraint which ensures that if a chip is produced, a setup must also be incurred in that period. Constraint (3.10) imposes non negativity on the production and inventory variables and forces the setup variable and the fixed overtime variable at Plant A to be binary, and the starting inventory of chips in period 1 to be zero. The planning horizon for Plant A is offset by L time periods as compared to Plant B as the constant lead time for supplying all components required by Plant B is assumed to be L time periods.

3.2 Coordinated Model

The uncoordinated approach is inadequate for production planning in a multi-plant scenario because it ignores the interactions that exist between the processes at the two plants. This approach can result in sub-optimality for the firm as a whole. For optimal performance, the objective must be to determine the best mix of inventory of modules and overtime usage at the two plants in each period so that overall cost is minimized. In the uncoordinated approach, Plant B drives the model and this *may* impose a *bad* demand vector on Plant A to which it cannot adapt, resulting in excessive inventory/overtime costs and consequently higher overall costs. In contrast, the coordinated approach jointly determines production quantities at both the plants by minimizing the total costs at the two plants. This objective captures the important goal of integrating the processes at the two plants and has been much emphasized in operations management literature. The coordinated problem may be stated as follows:

Problem C

Minimize $C_{opt} =$

$$\overbrace{\sum_{i=m+1}^{m+n}\sum_{t=1}^{T} h_i.I_{it} + \sum_{t=1}^{T} f_{2t}.z_{2t} + \sum_{t=1}^{T} v_{2t}.O_{2t}}^{PlantB} + \underbrace{\sum_{i=1}^{m}\sum_{t=1-L}^{T-L} h_i.I_{it} + \sum_{t=1-L}^{T-L} f_{1t}.z_{1t} + \sum_{t=1-L}^{T-L} v_{1t}.O_{1t}}_{PlantA}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n \qquad (3.11)$$

$$\sum_{i=m+1}^{m+n} (s_i.Y_{it} + b_i.X_{it}) - O_{2t} \leq Cap_{2t} \qquad \forall t = 1,\ldots,T \qquad (3.12)$$

$$O_{2t} \leq OT_{2t}.z_{2t} \qquad \forall t = 1,\ldots,T \qquad (3.13)$$

$$X_{it} \leq M_{i2t}.Y_{it} \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n \qquad (3.14)$$

$$I_{it-1} + X_{it} - I_{it} = P_{it} \qquad \forall t = 1-L,\ldots,T-L, \quad i = 1,\ldots,m \qquad (3.15)$$

$$\sum_{i=m+1}^{m+n} u_{ik}.X_{it} = P_{kt-L} \qquad \forall t = 1,\ldots T, k = 1,\ldots m \qquad (3.16)$$

$$\sum_{i=1}^{m}(s_i.Y_{it} + b_i.X_{it}) - O_{1t} \leq Cap_{1t} \qquad \forall t = 1-L,\ldots,T-L \quad (3.17)$$

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1-L,\ldots,T-L \quad (3.18)$$

$$X_{it} \leq M_{i1t}.Y_{it} \qquad \forall t = 1-L,\ldots,T-L, \quad (3.19)$$

$$i = 1,\ldots,m$$

$$z_{pt}, Y_{it} \in (0,1), \quad X_{it}, I_{it} \geq 0, \quad I_{i0} = 0 \qquad \forall t,\ i,\ p \quad (3.20)$$

The formulation of Problem C is *a priori* superior to the uncoordinated approach described before, because it accounts for the interaction between the two plants. However, this formulation is computationally intractable because of two sets of complicating constraints : the capacity constraints (3.12) and (3.17) and constraints (3.16) which contain the coupling terms $\sum_{i'=m+1}^{m+n} u'_{ik}.X_{it}$. The capacity constraints tie the products together at each plant while the coupling constraints tie the chips and the set of modules for which they are used (according to the bill of materials). For the given formulation of C, direct application of branch and bound will be computationally so expensive as to be impractical for real life problems. Evidently the problem needs to be restructured so that alternative strategies other than direct application of branch and bound method can be implemented to solve the problem efficiently. Lagrangian relaxation is one methodology which has often provided the best existing algorithm for some of the most difficult combinatorial optimization problems. Lagrangian relaxation transforms the original problem by relaxing the complicating constraints and adding a penalty term equal to the product of the complicating constraints' violations and Lagrange multipliers. In the next chapter we deliberate the significance and operational details of Lagrangian relaxation as a solution methodology for our model.

Chapter 4

Research Methodology

As stated in the previous chapter, **Lagrangian relaxation** has often provided the best existing algorithm for several complex combinatorial optimization problems (Geoffrion (1974) and Fisher (1981) for a comprehensive review and detailed treatment of the theory of Lagrangian relaxation). The overall objective of this chapter is to deliberate the significance of Lagrangian relaxation as a solution methodology and to specify the operational details for a Lagrangian relaxation based solution algorithm for our model. We first present a brief review of the theory of Lagrangian relaxation. Next, we examine the advantages and disadvantages of different strategies of relaxation of constraints and present algorithms based on these strategies.

4.1 Background

Using constructs from Fisher (1981), we reproduce some of the major building blocks necessary for the successful application of Lagrangian relaxation methodology. The original problem C can be stated in a generic form as follows:

Minimize

$$C_{primal} = cx$$

subject to

$$\begin{aligned} Ax &= b \\ Dx &= e \\ x &\geq 0 \quad \text{and integral} \end{aligned}$$

The objective function cx in the above formulation is a generic representation of all cost terms in Problem C (see page 66) and is defined over the cost coefficients and the decision variable set x. The constraints in the above formulation may be partitioned into two sets, one of which is relatively easy to solve by special algorithms while the other set comprises complicating constraints (for Problem C this is the set of the coupling and the capacity constraints). As noted above, Lagrangian relaxation transforms the original problem by relaxing the complicating constraints and adding a penalty term incorporating the product of the complicating constraints' violations and Lagrange multipliers, λ. The transformed problem is called the Lagrangian problem and can be represented as follows: **Minimize**

$$C_{Lagrangian}(\lambda) = cx + \lambda(b - Ax)$$

subject to

$$\begin{aligned} Dx &= e \\ x &\geq 0 \quad \text{and integral} \end{aligned}$$

For any λ, the Lagrangian problem is relatively easy to solve due to the special structure of the constraint set Dx=e and is a lower bound on the optimal value of the primal problem (Fisher, 1981). The best choice of λ would be the optimal solution to the Lagrangian dual problem.

$$C_{Ldual} = Maximize_{\lambda} \quad C_{Lagrangian}(\lambda)$$

Alternative procedures may be used for selecting promising values of the Lagrange multipliers, λ. Most schemes for determining appropriate values of λ have as their objective, finding optimal or near optimal solutions to the above Lagrangian dual problem. Ideally we would like to relax constraints in such a manner that the Lagrangian problem is easy to solve and the optimal solution to the Lagrangian dual is close to the primal optimal. Then, by suitably perturbing these near optimal solutions, good feasible solutions to the primal problem can be obtained. The true optimal solution value is therefore bounded between the value of the best known Lagrangian problem (lower bound) and the value of the best known feasible solution (upper bound). The smaller the gap between the lower and the upper bounds, the more certainty we have that the feasible solution is close to the optimal. Several authors have observed (Fisher (1981), Thizy (1991), Chen and Thizy (1991) and Millar and Yang (1993) for specific examples) that the choice of the constraints to relax has a critical impact on the quality of the bounds and the computational requirements for solving the Lagrangian problem. The specific constraints that we choose to relax will therefore exert considerable influence on the size of the gap we can obtain between the upper and the lower bounds. We now analyze some of these alternative relaxations of Problem C and this analysis will be used later to evolve an efficient algorithm for the multi-plant coordination problem.

4.2 Choice between Competing Relaxations

The choice of the constraint to be relaxed is dependent on the trade-off between the ease of solving the relaxed problems, and the quality of bounds obtained. Relaxing fewer or different constraints may result in sharper bounds but may increase the computational requirements so heavily as to make such a relaxation impractical for large problems. In the formulation of Problem C, there are two sets of complicating constraints: the capacity constraints (3.12) and (3.17) and the coupling constraints (3.16). We analyze the relative advantages and disadvantages of three competing relaxations which may be derived from Problem C as follows:

(1) Relax only coupling constraints.

(2) Relax both capacity and coupling constraints.

(3) Do not relax any constraints but reformulate the problem by creating copies of variables. As described later, this technique called **Lagrangian decomposition** (also referred to as **variable splitting** by some authors) creates independent sub-problems by using one copy in two different subsets of constraints and then relaxing the condition that the two copies should be identical.

We now describe these relaxations/reformulations and our preliminary computational experience with them. We present the Lagrangian problems for each of the above cases, discuss the structure and the advantages/disadvantages of the relaxed problems and methods for optimizing the Lagrangian dual.

4.2.1 Relaxing Coupling Constraints

For Problem C we note that if the coupling constraints (3.16) are relaxed, the problem decomposes into two independent single plant problems for Plant A and Plant B respectively. However, each of these sub-problems are capacitated lotsizing problems and we cannot expect to get optimal solutions efficiently, for large problem instances. For real life problems therefore, this relaxation appears *a priori* impractical from an implementation point of view. Nevertheless, we explored the impact of relaxing the coupling constraints on the quality of bounds for small problems. Starting with Problem C, the coupling constraints were dualized using Lagrange multipliers and adding the appropriate penalty term to the objective function. The procedure followed is detailed below.

a. Dualize constraints (3.16) with dual variables π_{kt} and add the penalty function $\pi_{kt}.(P_{kt-L} - \sum_{i=m+1}^{m+n} u_{ik}.X_{it})$ for all $k = 1, \ldots$,m and $t = 1, \ldots$ T to the objective function.

The above yields a Lagrangian problem which decomposes into two independent capacitated single-plant problems. These problems $C_{RCoupleB}$ and $C_{RCoupleA}$ are shown below:

Problem $C_{RCoupleB}$

Minimize

$$\sum_{i=m+1}^{m+n} \sum_{t=1}^{T} h_i.I_{it} + \sum_{t=1}^{T} v_{2t}.O_{2t} + \sum_{t=1}^{T} f_{2t}.z_{2t} - \sum_{i=m+1}^{m+n} \sum_{t=1}^{T} \sum_{k=1}^{m} \pi_{kt}.u_{ik}.X_{it}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \quad \forall t = 1, \ldots, T, \quad i = m+1, \ldots, m+n \qquad (4.1)$$

$$\sum_{i=m+1}^{m+n} (s_i.Y_{it} + b_i.X_{it}) - O_{2t} \leq Cap_{2t} \quad \forall t = 1, \ldots, T \qquad (4.2)$$

$$O_{2t} \leq OT_{2t}.z_{2t} \quad \forall t = 1, \ldots, T \qquad (4.3)$$

$$X_{it} \leq M_{i2t}.Y_{it} \quad \forall t = 1, \ldots, T, \quad i = m+1, \ldots, m+n \qquad (4.4)$$

$$z_{2t}, Y_{it} \in (0,1), \quad X_{it}, I_{it} \geq 0, \quad I_{i0} = 0 \quad \forall t = 1, \ldots, T, \quad i = m+1, \ldots, m+n \qquad (4.5)$$

Problem $C_{RCoupleA}$

Minimize

$$\sum_{i=1}^{m} \sum_{t=1-L}^{T-L} h_i.I_{it} + \sum_{t=1-L}^{T-L} v_{1t}.O_{1t} + \sum_{t=1-L}^{T-L} f_{1t}.z_{1t} + \sum_{i=1}^{m} \sum_{t=1-L}^{T-L} \pi_{kt}.P_{kt}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = P_{it} \quad \forall t = 1-L, \ldots, T-L,$$

$$i = 1, \ldots, m \qquad (4.6)$$

$$\sum_{i=1}^{m}(s_i.Y_{it} + b_i.X_{it}) - O_{1t} \leq Cap_{1t} \qquad \forall t = 1-L, \ldots, T-L \qquad (4.7)$$

$$\sum_{t'=1-L}^{t} P_{kt'} \geq \sum_{i=m+1}^{m+n} \sum_{t'=1}^{t+L} u_{ik}.d_{it'} \qquad \forall t = 1-L, \ldots, T-L,$$

$$k = 1, \ldots, m \qquad (4.8)$$

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1-L, \ldots, T-L \qquad (4.9)$$

$$X_{it} \leq M_{i1t}.Y_{it} \qquad \forall t = 1-L, \ldots, T-L, \qquad (4.10)$$

$$i = 1, \ldots, m$$

$$z_{it}, Y_{it} \in (0,1), \quad X_{it}, I_{it} \geq 0, \quad I_{i0} = 0 \qquad \forall t = 1-L, \ldots, T-L,$$

$$i = 1, \ldots, m \qquad (4.11)$$

We note that a new constraint (4.8) which was not present in C has been added in the formulation of Problem $C_{RCoupleA}$. By ensuring that the cumulative production of chips in each period is at least equal to the cumulative demand for chips up to this period (i.e., there are no backorders for chips), this constraint strengthens the formulation for Problem $C_{RCoupleA}$. This constraint is redundant in Problem C because constraints (3.15) and (3.16) automatically ensure the above requirement. Solving problems $C_{RCoupleB}$ and $C_{RCoupleA}$ to optimality (using branch and bound) yields a lower bound on C. An upper bound was generated using the sequential approach i.e., solve $C_{RCoupleB}$, define requirements for Plant A, solve $C_{RCoupleA}$. The procedure used for optimizing the Lagrangian dual was **subgradient optimization**, the operational details of which are described later. Obviously, we could only select small problems for this exercise because capacitated single plant problems were required to be solved to optimality by a branch and bound procedure. For the few problems that were solved, the average gap between the upper and lower bounds was in the vicinity of 5%. This gap could be further reduced by continuing subgradient optimization for a larger number of iterations (we terminated the procedure after only 10 iterations because of the difficulty of solving capacitated lotsizing problems by direct application of branch and bound). In order to get a better feel for the quality of bounds in the cases considered, the original problem was solved to optimality. In all cases there was a tight fit between the upper bound and the true optimal value (less than 1% deviation). While we have run very few experiments for this relaxation, in general it appears that the gap between the upper and lower bound for this relaxation will be quite small. As well, the upper bound appears to be quite close to the optimal. However, from a computational point of view, the above relaxation is clearly impractical because of the heavy requirements for solving capacitated lotsizing problems optimally by branch and bound. The bounds obtained are useful as a bench mark to compare the bounds from the other relaxations that are considered later.

4.2.2 Relaxing Capacity and Coupling Constraints

The second relaxation we consider is obtained from Problem C by the Lagrangian relaxation of the capacity constraints at both the plants (Constraints (3.12 and (3.17)) and the coupling constraints between the two plants (3.16). The problem then decomposes into independent problems for Plant A and Plant B respectively. Moreover since there are no complicating constraints tying up the products, the problem for each plant decomposes into independent problems for each individual item. We first present Problem C in an equivalent form, C_{LRelax} by creating copies of the overtime variables O_{pt}. As explained later, creating copies in this manner gives a better structure to the relaxed problems by separating the overtime problems and facilitates their efficient solution. The equivalent problem C_{LRelax} is presented below.

Problem C_{LRelax}

Minimize $C_{LRelax} =$

$$\overbrace{\sum_{i=m+1}^{m+n}\sum_{t=1}^{T} h_i.I_{it} + \sum_{t=1}^{T} f_{2t}.z_{2t} + \sum_{t=1}^{T} v_{2t}.O_{2t}}^{PlantB} + \underbrace{\sum_{i=1}^{m}\sum_{t=1-L}^{T-L} h_i.I_{it} + \sum_{t=1-L}^{T-L} f_{1t}.z_{1t} + \sum_{t=1-L}^{T-L} v_{1t}.O_{1t}}_{PlantA}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \qquad \forall t = 1, \ldots, T, \quad i = m+1, \ldots, m+n \tag{4.12}$$

$$\sum_{i=m+1}^{m+n} (s_i.Y_{it} + b_i.X_{it}) - O'_{2t} \leq Cap_{2t} \qquad \forall t = 1, \ldots, T \tag{4.13}$$

$$X_{it} \leq M_{i2t}.Y_{it} \qquad \forall t = 1, \ldots, T, \quad i = m+1, \ldots, m+n \tag{4.14}$$

$$O_{2t} \leq OT_{2t}.z_{2t} \qquad \forall t = 1, \ldots, T \tag{4.15}$$

$$I_{it-1} + X_{it} - I_{it} - P_{it} = 0 \qquad \forall t = 1-L, \ldots, T-L, \quad i = 1, \ldots, m \tag{4.16}$$

$$\sum_{i=1}^{m}(s_i.Y_{it} + b_i.X_{it}) - O'_{1t} \leq Cap_{1t} \qquad \forall t = 1-L, \ldots, T-L \tag{4.17}$$

$$X_{it} \leq M_{i1t}.Y_{it} \qquad \forall i, t = 1-L, \ldots, T-L, \quad i = 1, \ldots, m \tag{4.18}$$

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1-L, \ldots, T-L \tag{4.19}$$

$$\sum_{i=m+1}^{m+n} u_{ik}.X_{it} = P_{kt-L} \qquad \forall t = 1, \ldots T, k = 1, \ldots m \tag{4.20}$$

$$O_{2t} = O'_{2t} \qquad \forall t = 1, \ldots T \tag{4.21}$$

$$O_{1t} = O'_{1t} \qquad \forall t = 1-L, \ldots T-L \tag{4.22}$$

$$z_{pt}, Y_{it} \in (0,1), \qquad X_{it}, I_{it} \geq 0, \;\; I_{i0} = 0 \qquad \forall t, i, p \tag{4.23}$$

It can be easily seen that C and C_{LRelax} are equivalent. In the formulation, C_{LRelax}, constraints (4.12), (4.13), (4.14) and (4.15) represent the problem for Plant B while constraints (4.16), (4.17), (4.18) and (4.19) represent the problem for Plant A. Constraints (4.20) are the coupling constraints between the two plants.

Constraints (4.21) and (4.22) create copies of overtime variables at either plant. The addition of these constraints facilitate the decomposition of the Lagrangian problem into easily solvable, independent overtime sub-problems. The relaxation procedure is detailed below.

a. Dualize constraints (4.13) with dual variables γ_{2t} and add the penalty function $\gamma_{2t}.(Cap_{2t} - \sum_{i=m+1}^{m+n}(s_i.Y_{it} + b_i.X_{it}) + O'_{2t})$ for all t = 1, ...,T to the objective function.

b. Dualize constraints (4.17) with dual variables γ_{1t} and add the penalty function $\gamma_{1t}.(Cap_{1t} - \sum_{i=1}^{m}(s_i.Y_{it} + b_i.X_{it}) + O'_{1t})$ for all t = 1-L, ...,T-L to the objective function.

c. Dualize constraints (4.20) with dual variables π_{kt} and add the penalty function $\pi_{kt}.(P_{kt-1} - \sum_{i=m+1}^{m+n} u_{ik}.X_{it})$ for all $k = 1, \ldots$,m and $t = 1, \ldots$ T to the objective function.

d. Dualize constraints (4.21) with dual variables δ_{2t} and add a penalty function $\delta_{2t}(O'_{2t}$ - $O_{2t})$ for all t = 1, ...,T to the objective function

e. Dualize constraints (4.22) with dual variables δ_{1t} and add a penalty function $\delta_{1t}(O'_{1t}$ - $O_{1t})$ for all t = 1-L, ...,T-L to the objective function

Problem C_{LRelax} with the penalty terms added in the objective function and the constraint set defined by the special structure constraints namely, (4.12), (4.14), (4.15), (4.16), (4.16), (4.18), and (4.19) represents the Lagrangian problem for the current relaxation. Separating subsets of constraints we get two independent problems for each plant. We call these decomposed problems $B_{LRelax1}$, $B_{LRelax2}$, $A_{LRelax1}$ and $A_{LRelax2}$ respectively. These problems can be represented as follows:

Decomposed Problems for C_{LRelax}

Problem $B_{LRelax1}$

Minimize $C_{BLRelax1} =$

$$\sum_{i=m+1}^{m+n}\sum_{t=1}^{T} h_i.I_{it} - \sum_{i=m+1}^{m+n}\sum_{t=1}^{T}\gamma_{2t}.b_i.X_{it} - \sum_{i=m+1}^{m+n}\sum_{t=1}^{T}\sum_{k=1}^{m}\pi_{kt}.u_{ik}.X_{it} - \sum_{i=m+1}^{m+n}\sum_{t=1}^{T}\gamma_{2t}.s_i.Y_{it} + \sum_{t=1}^{T}(\delta_{2t}+\gamma_{2t}).O'_{2}$$

subject to

$$\begin{aligned}
I_{it-1} + X_{it} - I_{it} &= d_{it} \qquad \forall t = 1,\ldots,T, \qquad i = m+1,\ldots,m+n \\
X_{it} &\le M_{i2t}.Y_{it} \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n \\
O'_{2t} &\le OT_{2t} \qquad \forall t = 1,\ldots,T
\end{aligned}$$

<u>Problem $B_{LRelax2}$</u>

Minimize $C_{BLRelax2} =$

$$\sum_{t=1}^{T} v_{2t}.O_{2t} + \sum_{t=1}^{T} f_{2t}.z_{2t} - \sum_{t=1}^{T}\delta_{2t}.O_{2t}$$

subject to

$$O_{2t} \le OT_{2t}.z_{2t} \qquad \forall t = 1,\ldots,T$$

<u>Problem $A_{LRelax1}$</u>

Minimize $C_{ALRelax1} =$

$$\sum_{i=1}^{m}\sum_{t=1-L}^{T-L} h_i.I_{it} - \sum_{i=1}^{m}\sum_{t=1-L}^{T-L}\gamma_{1t}.b_i.X_{it} + \sum_{i=1}^{m}\sum_{t=1-L}^{T-L}\pi_{kt}.P_{kt} - \sum_{i=1}^{m}\sum_{t=1-L}^{T-L}\gamma_{1t}.s_i.Y_{it} + \sum_{t=1-L}^{T-L}(\delta_{1t}+\gamma_{1t}).O'_{1t}$$

subject to

$$\begin{aligned}
I_{it-1} + X_{it} - I_{it} - P_{it} &= 0 \qquad \forall t = 1-L,\ldots,T-L, \qquad i = 1,\ldots,m \\
X_{it} &\le M_{i1t}.Y_{it} \qquad \forall t = 1-L,\ldots,T-L, \qquad i = 1,\ldots,m \\
\sum_{t'=1-L}^{t} P_{kt'} &\ge \sum_{i=m+1}^{m+n}\sum_{t'=1}^{t+L} u_{ik}.d_{it'} \qquad \forall t = 1-L,\ldots,T-L, \\
& \qquad k = 1,\ldots,m \\
O'_{1t} &\le OT_{1t} \qquad \forall t = 1-L,\ldots,T-L
\end{aligned}$$

Problem $A_{LRelax2}$

Minimize $C_{ALRelax2} =$

$$\sum_{t=1-L}^{T-L} v_{1t}.O_{1t} + \sum_{t=1-L}^{T-L} f_{1t}.z_{1t} - \sum_{t=1-L}^{T-L} \delta_{1t}.O_{1t}$$

subject to

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1-L, \ldots, T-L$$

Problems $B_{LRelax1}$ and $A_{LRelax1}$ decompose into uncapacitated single item lotsizing problems (one for each item) which can be efficiently solved by the dynamic programming based Wagner-Whitin (1958) algorithm. For these problems, O'_{pt} is fixed at its upper bound if the appropriate coefficient $(\delta_{pt} + \gamma_{pt})$ is negative and takes the value 0 otherwise. We note that an additional constraint has been added in Problem $A_{LRelax1}$. The justification for adding this constraint is the same that for the previous relaxation (see page 73). Problems $B_{LRelax2}$ and $A_{LRelax2}$ represent the overtime problems which can be easily solved by the following easy inspection procedure.

If $\quad (v_{pt} - \delta_{pt}).OT_{pt} + f_{pt} < 0$

then

$$z_{pt} = 1, O_{pt} = OT_{pt}$$

else

$$z_{pt} = 0, O_{pt} = 0$$

Optimal solutions to the above problems provide a lower bound on C. Using the solution of the Lagrangian problem as a starting point, a complex heuristic was used to generate feasible solutions (upper bound). Subgradient optimization was used for the optimization of the Lagrangian dual and this procedure was terminated after reaching a specified number of iterations. The details of the subgradient optimization procedure and the heuristic are explained in the next chapter. The gap between upper and lower bounds appears to be much larger than was the case when only the coupling constraints were relaxed. This large gap (between the upper and the lower bounds) may be attributable to either an inherent duality gap between the optimal solution to the primal problem and the optimal solution to the Lagrangian dual (for the given relaxation) or a large deviation of the best known feasible solution from the true optimal or an incomplete computation of the best lower bound. This is an unfortunate ambiguity and we address this issue in Chapter 5. As earlier we also compared the best feasible solution to the true optimal value of the primal problem (obtained by direct application of branch and bound). The best feasible solution on the average turned out to be 1 % more expensive than the true optimal solution. While the large gap between the best upper and lower bounds is a cause for concern, we observe that the upper bound deviates from the true optimal by a much smaller quantity. Comparing the results for the two relaxations (see Figure 4.1), it appears that a large portion of the gap arises due to the relaxation of capacity constraints at either plant, rather than the relaxation of the coupling constraints. A scheme which explicitly incorporates the capacity constraints rather than dualize them could result in a lower duality gap. We discuss one such approach in the next section.

4.2.3 Variable Splitting/Lagrangian Decomposition

Sometimes the complicating constraint may not be the best choice for relaxation. Millar and Yang (1993) show that good results may be obtained for the multi-item, single stage, capacitated lotsizing problem (without setup times) by using a reformulation approach which avoids the most obvious choice of relaxing the capacity constraint (earlier used by Thizy and Van Wassenhove, 1985). The authors created two copies of X_{it}, the production quantity variables, used one copy in a different subset of constraints, and then relaxed the condition that the two copies should be identical. This separated the problem into two different sub-problems each of which could be efficiently solved by exploiting the special underlying structure. The gap between the upper and lower bounds were shown to be consistently better than those obtained by using traditional Lagrangian relaxation of the capacity constraints. This technique was pioneered by Glover and Mulvey (1980) who gave it the name **variable splitting**. Later Guignard (1984) established several important properties and used the term Lagrangian decomposition. Guignard and Kim (1987) analytically showed that Lagrangian decomposition is superior to traditional Lagrangian relaxation where all but one specially structured constraint set are dualized.

Guignard and Kim (1987) provide an interesting interpretation of what the Lagrangian decomposition does from a primal viewpoint. They show that optimizing the Lagrangian decomposition dual is equivalent to optimizing the primal objective function on the intersection of the convex hulls of the constraint sets. They also prove that if any of the constraint subsets possess the integrality property (Geoffrion, 1974) then Lagrangian decomposition will provide as good (but no better) a lower bound as the strongest of the Lagrangian relaxation bounds. If none of the subsets has the integrality property, Lagrangian decomposition is likely to provide better bounds. Considering the Lagrangian relaxation of capacity constraints in Problem C, we observe that the constraint subsets do not have the integrality property. Lagrangian decomposition therefore promises lower bounds which are potentially better than those obtained by Lagrangian relaxation of the capacity constraints.

With the above background, we now explore a new formulation equivalent to Problem C and its Lagrangian decomposition on lines similar to Millar and Yang (1993) which may lead to tighter bounds than those obtained by Lagrangian relaxation of capacity and coupling constraints. The goal here is to explicitly include the capacity constraints in one of the decomposed problems rather than relaxing them and adding the corresponding penalty term to the objective function. This formulation $C_{Ldecomp}$ (equivalent to the original problem C), is obtained by making copies of the variables X_{it}, the production quantity variables, Y_{it}, the setup variables and O_{pt}, the overtime requirement variables. Creating copies of variables in this manner enables us to use the original variables in one sub-problem and copies of these variables in another. The sub-problems are linked by the requirement that the copies (of variables) created must be identical to the original variables. Later the last requirement will be relaxed in order to exploit the special underlying structure of the individual sub-problems for developing efficient solution algorithms. $C_{Ldecomp}$ can be represented as follows:

Problem $C_{Ldecomp}$

Minimize $C_{Ldecomp} =$

$$\overbrace{\sum_{i=m+1}^{m+n}\sum_{t=1}^{T} h_i.I_{it} + \sum_{t=1}^{T} f_{2t}.z_{2t} + \sum_{t=1}^{T} v_{2t}.O_{2t}}^{PlantB} + \underbrace{\sum_{i=1}^{m}\sum_{t=1-L}^{T-L} h_i.I_{it} + \sum_{t=1-L}^{T-L} f_{1t}.z_{1t} + \sum_{t=1-L}^{T-L} v_{1t}.O_{1t}}_{PlantA}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \quad \forall t = 1, \ldots, T, \quad i = m+1, \ldots, m+n \tag{4.24}$$

$$\sum_{i=m+1}^{m+n} (s_i.Y'_{it} + b_i.X'_{it}) - O'_{2t} \leq Cap_{2t} \quad \forall t = 1, \ldots, T \tag{4.25}$$

$$X_{it} \leq M_{i2t}.Y_{it} \quad \forall i, t = 1, \ldots, T, \quad i = m+1, \ldots, m+n \tag{4.26}$$

$$O_{2t} \leq OT_{2t}.z_{2t} \quad \forall t = 1, \ldots, T \tag{4.27}$$

$$I_{it-1} + X_{it} - I_{it} - P_{it} = 0 \quad \forall t = 1-L, \ldots, T-L, \quad i = 1, \ldots, m \tag{4.28}$$

$$\sum_{i=1}^{m}(s_i.Y'_{it} + b_i.X'_{it}) - O'_{1t} \leq Cap_{1t} \quad \forall t = 1-L, \ldots, T-L \quad (4.29)$$

$$X_{it} \leq M_{i1t}.Y_{it} \quad \forall i, t = 1-L, \ldots, T-L, \quad i = 1, \ldots, m \quad (4.30)$$

$$O_{1t} \leq OT_{1t}.z_{1t} \quad \forall t = 1-L, \ldots, T-L \quad (4.31)$$

$$\sum_{i=m+1}^{m+n} u_{ik}.X_{it} = P_{kt-L} \quad \forall t = 1, \ldots T, k = 1, \ldots m \quad (4.32)$$

$$X_{it} = X'_{it} \quad \forall i, t \quad (4.33)$$

$$Y_{it} = Y'_{it} \quad \forall i, t \quad (4.34)$$

$$O_{2t} = O'_{2t} \quad \forall t = 1, \ldots, T \quad (4.35)$$

$$O_{1t} = O'_{1t} \quad \forall t = 1-L, \ldots, T-L \quad (4.36)$$

$$z_{pt}, Y_{it} \in (0,1), \quad X_{it}, I_{it} \geq 0, \; I_{i0} = 0 \quad \forall t, i, p \quad (4.37)$$

It can be easily seen that C and $C_{Ldecomp}$ are equivalent. In the formulation, $C_{Ldecomp}$, constraints (4.24), (4.25), (4.26) and (4.27) represent the problem for Plant B while constraints (4.28), (4.29), (4.30) and (4.31) represent the problem for Plant A. Constraints (4.33), (4.34), (4.35) and (4.36) represent copies of variables as described before. The Lagrangian problem is obtained by dualizing constraints (4.32), (4.33), (4.34), (4.35) and (4.36), and adding the appropriate penalty term (comprising the product of the violations of the dualized constraints and the Lagrange multipliers) to the objective function. The dualizing procedure is detailed below:

a. Dualize constraints (4.32) with dual variables π_{kt} and add the penalty function $\pi_{kt}.(P_{kt-L} - \sum_{i=m+1}^{m+n} u_{ik}.X_{it})$ for all $k = 1, \ldots$,m and $t = 1, \ldots$T to the objective function.

b. Dualize constraints (4.33) with dual variables ρ_{it} and add a penalty function $\rho_{it}(X'_{it}$ - $X_{it})$ for all i, t to the objective function

c. Dualize constraints (4.34) with dual variables γ_{it} and add a penalty function $\gamma_{it}(Y'_{it}$ - $Y_{it})$ for all i, t to the objective function

d. Dualize constraints (4.35) with dual variables δ_{2t} and add a penalty function $\delta_{2t}(O'_{2t}$ - $O_{2t})$ for all t = 1, ...,T to the objective function

e. Dualize constraints (4.36) with dual variables δ_{1t} and add a penalty function $\delta_{1t}(O'_{1t}$ - $O_{1t})$ for all t = 1-L, ...,T-L to the objective function

Problem $C_{Ldecomp}$ with the penalty terms added in the objective function and the constraint set defined by the constraints (4.24) to (4.31) and (4.37) (i.e., constraints that are not relaxed), represents the Lagrangian problem for the current Lagrangian decomposition based procedure. Separating subsets of constraints, the Lagrangian problem decomposes into three independent problems for each plant. We call these decomposed problems $B_{Ldecomp1}$, $B_{Ldecomp2}$, $B_{Ldecomp3}$, $A_{Ldecomp1}$, $A_{Ldecomp2}$ and $A_{Ldecomp2}$ respectively. These problems can be represented as follows:

Decomposed Problems for $C_{Ldecomp}$

Problem $B_{Ldecomp1}$

Minimize $C_{BLdecomp1}$ =

$$\sum_{i=m+1}^{m+n}\sum_{t=1}^{T} h_i.I_{it} - \sum_{i=m+1}^{m+n}\sum_{t=1}^{T}\rho_{it}.X_{it} - \sum_{i=m+1}^{m+n}\sum_{t=1}^{T}\sum_{k=1}^{m}\pi_{kt}.u_{ik}.X_{it} - \sum_{i=m+1}^{m+n}\sum_{t=1}^{T}\gamma_{it}.Y_{it}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n$$

$$X_{it} \leq M_{i2t}.Y_{it} \qquad \forall t = 1,\ldots,T, \quad i = m+1,\ldots,m+n$$

Problem $B_{Ldecomp2}$

Minimize $C_{BLdecomp2}$ =

$$\sum_{t=1}^{T} v_{2t}.O_{2t} + \sum_{t=1}^{T} f_{2t}.z_{2t} - \sum_{t=1}^{T}\delta_{2t}.O_{2t}$$

subject to

$$O_{2t} \leq OT_{2t}.z_{2t} \qquad \forall t = 1, \ldots, T$$

Problem $B_{Ldecomp3}$

Minimize $C_{BLdecomp3} =$

$$\sum_{i=m+1}^{m+n} \sum_{t=1}^{T} \rho_{it}.X'_{it} + \sum_{i=m+1}^{m+n} \sum_{t=1}^{T} \gamma_{it}.Y'_{it} + \sum_{t=1}^{T} \delta_{2t}.O'_{2t}$$

subject to

$$\begin{aligned}
\sum_{i=m+1}^{m+n} (s_i.Y'_{it} + b_i.X'_{it}) - O'_{2t} &\leq Cap_{2t} \qquad \forall t = 1, \ldots, T \\
O'_{2t} &\leq OT_{2t} \qquad \forall t = 1, \ldots, T \\
X'_{it} &\leq M_{i2t}.Y'_{it} \qquad \forall t = 1, \ldots, T, \\
&\qquad i = m+1, \ldots, m+n
\end{aligned}$$

Problem $A_{Ldecomp1}$

Minimize $C_{ALdecomp1} =$

$$\sum_{i=1}^{m} \sum_{t=1-L}^{T-L} h_i.I_{it} - \sum_{i=1}^{m} \sum_{t=1-L}^{T-L} \rho_{it}.X_{it} + \sum_{i=1}^{m} \sum_{t=1-L}^{T-L} \pi_{it}.P_{it} - \sum_{i=1}^{m} \sum_{t=1-L}^{T-L} \gamma_{it}.Y_{it}$$

subject to

$$\begin{aligned}
I_{it-1} + X_{it} - I_{it} - P_{it} &= 0 \qquad \forall t = 1-L, \ldots, T-L, \\
&\qquad i = 1, \ldots, m \\
X_{it} &\leq M_{i1t}.Y_{it} \qquad \forall t = 1-L, \ldots, T-L, \\
&\qquad i = 1, \ldots, m \\
\sum_{t'=1-L}^{t} P_{kt'} &\geq \sum_{i=m+1}^{m+n} \sum_{t'=1}^{t+L} u_{ik}.d_{it'} \qquad \forall t = 1-L, \ldots, T-L, \\
&\qquad k = 1, \ldots, m
\end{aligned}$$

Problem $A_{Ldecomp2}$

Minimize $C_{ALdecomp2}$ =

$$\sum_{t=1-L}^{T-L} v_{1t}.O_{1t} + \sum_{t=1-L}^{T-L} f_{1t}.z_{1t} - \sum_{t=1-L}^{T-L} \delta_{1t}.O_{1t}$$

subject to

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1-L, \ldots, T-L$$

Problem $A_{Ldecomp3}$

Minimize $C_{ALdecomp3}$ =

$$\sum_{i=1}^{m} \sum_{t=1-L}^{T-L} \rho_{it}.X'_{it} + \sum_{i=1}^{m} \sum_{t=1-L}^{T-L} \gamma_{it}.Y'_{it} + \sum_{t=1-L}^{T-L} \delta_{2t}.O'_{1t}$$

subject to

$$\sum_{i=1}^{m}(s_i.Y'_{it} + b_i.X'_{it}) - O'_{1t} \leq Cap_{1t} \qquad \forall t = 1-L, \ldots, T-L$$

$$O'_{1t} \leq OT_{1t} \qquad \forall t = 1-L, \ldots, T-L$$

$$X'_{it} \leq M_{i1t}.Y'_{it} \qquad \forall t = 1-L, \ldots, T-L, \quad i = 1, \ldots, m$$

Problems $B_{Ldecomp1}$ and $A_{Ldecomp1}$ can be efficiently solved by using the dynamic programming based algorithm of Wagner-Whitin (1958). As for the previous relaxations, an additional constraint has been added in Problem $A_{Ldecomp1}$ to improve bounds. Problems $B_{Ldecomp2}$ and $A_{Ldecomp2}$ can be solved easily by the same inspection procedure described in the previous subsection for the problems $B_{LRelax2}$ and $A_{LRelax2}$.

If $\qquad (v_{pt} + \delta_{pt}).OT_{pt} + f_{pt} < 0$

then

$$z_{pt} = 1, O_{pt} = OT_{pt}$$

else

$$z_{pt} = 0, O_{pt} = 0$$

Problems $B_{Ldecomp3}$ and $A_{Ldecomp3}$ represent continuous knapsack problems which can be easily solved for large instances. It is evident from our decompositions that we are unable to solve both capacity and demand feasibility constraints in the same sub-problem. It was this procedure that enabled Millar and Yang (1993) to reduce the gap between the upper and lower bounds considerably. However those authors did not consider setup times in their problems. The presence of non-zero setup times eliminates the possibility of simultaneously solving capacity and demand constraints in a single problem. We tested the three relaxations described above, i.e., Lagrangian relaxation of coupling constraints, Lagrangian relaxation of capacity and coupling constraints and the Lagrangian decomposition, for a randomly generated problem. Our computational experience with this problem indicates improved lower bounds for Lagrangian decomposition than those obtained by relaxing the capacity and the coupling constraints. This confirms our earlier belief that for the current problem Lagrangian decomposition is likely to give a better lower bound than Lagrangian relaxation (see Figure 4.1 for a comparison of the bounds for the Lagrangian relaxations and the Lagrangian decomposition considered).

Another benefit of implementing Lagrangian decomposition for the current problem was that it helped uncover hidden underlying structures in the problem. The knapsack problems that emerge as a consequence of the decomposition (i.e., $B_{Ldecomp3}$ and $A_{Ldecomp3}$) were obscured in both the original problem as well as previous case of Lagrangian relaxation of capacity and coupling constraints. This issue has been discussed earlier in Guignard and Kim (1987) for a produc-

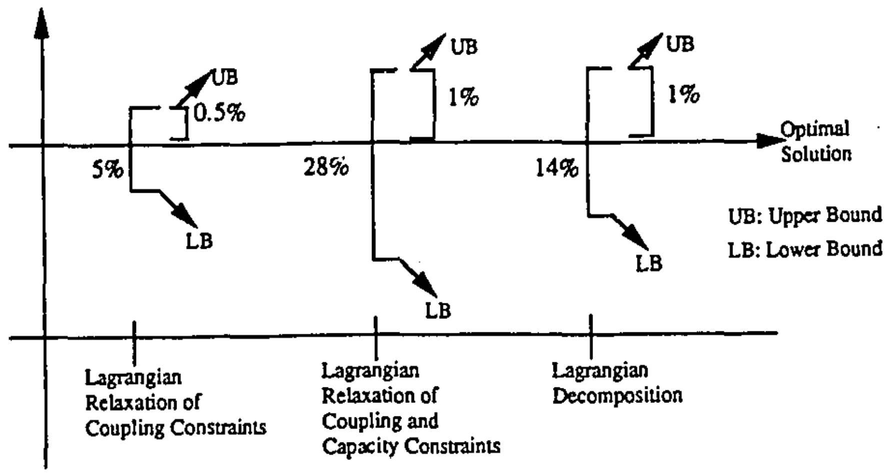

Figure 4.1: Comparison of Bounds for competing relaxations

tion scheduling problem in the process industry environment. The above provides evidence for the assertion of Guignard and Kim (1987) that Lagrangian decomposition may be viewed as a tool for extracting out of complex real world problems various well studied, specially structured, simpler problems for which efficient algorithms exist.

As stated earlier, Problems C, C_{LRelax} and $C_{LDecomp}$ given below are equivalent. From the theory of Lagrangian relaxation/decomposition, $LDualLRelax$ and $LDual_{LDecomp}$ provide a lower bound on C. For proof, see Geoffrion(1974), Fisher(1981) for Lagrangian relaxation and Guignard (1984) and Guignard and Kim (1987) for Lagrangian decomposition.

Problem C	Min C
(see page 66	subject to
	(3.11) to (3.20)
Problem	Min C_{LRelax}
C_{LRelax}	subject to
(see page 74	(4.12) to (4.23)
Problem	Min $C_{LDecomp}$

$C_{LDecomp}$	subject to
(see page 81)	(ldinvb) to (ldnon)
Problem	**Max** (**Min** $C_{BLRelax1} + C_{BRelax2}$
$LDual_{LRelax}$	$+ C_{ARelax1} + C_{ARelax2}$)
	subject to
	(4.13), (4.14), (4.15)
	(4.17), (4.18), (4.19), (4.23)
Problem	**Max** (**Min** $C_{BLdecomp1} + C_{BLdecomp2} + C_{BLdecomp3}$
$LDual_{LDecomp}$	$+ C_{ALdecomp1} + C_{ALdecomp2} + C_{ALdecomp2}$)
	subject to
	(4.24), (4.25), (4.26), (4.27),
	(4.28), (4.29), (4.30), (4.31),(4.37)

$LDual_{LRelax}$ is the Lagrangian dual for the case where the capacity and coupling constraints are relaxed and represents the greatest lower bound that can be achieved by the dualization of these complicating constraints. $LDual_{LDecomp}$ is the Lagrangian dual for the Lagrangian decomposition case (where only variable copying constraints are relaxed) and its optimal value represents the best lower bound that can be achieved for this procedure. The inner minimization in $LDual_{LRelax}$, above, has the s_1 - block angular structure i.e., the minimization decomposes naturally into s_1 independent problems. In our case, s_1 is equal to $m + n + 2T$ i.e., number of chips *plus* number of modules *plus* twice the number of time periods in the planning horizon. For Plant B, this yields n uncapacitated single item lotsizing problems (one for each module) *plus* T problems for overtime decision, one in each period. For Plant A, we get m uncapacitated single item lotsizing problems (one for each chip) *plus* T problems for overtime decision, one in each period. The inner minimization in $LDual_{LDecomp}$ has an s_2 - block angular structure with the minimization decomposing naturally into $s_2 = m + n + 4T$ independent problems. Of these problems, $m + n$ problems are single item uncapacitated lotsizing problems which can be solved by the dynamic programming based Wagner-Whitin (1958) algorithm while $2T$ problems can be solved trivially by the inspection procedure explained. Additionally $2T$ continuous knapsack problems must be solved for Problems $B_{Ldecomp3}$ and $A_{Ldecomp3}$.

While the number of products and time periods considered determine the problem dimensions in terms of the number of constraints and variables, there may be some variations between different formulations of the problem. For example, the Lagrangian relaxation formulation i.e., C_{LRelax}, has $T * (3m + 2n + 6)$ constraints and $T * (m + n + 4)$ variables (m here represents the number of chips, n the number of modules and T the number of time periods). For the case with 10/20 (modules/chips) and 4 time periods, this represents 344 constraints and 376 variables (128 of which are required to be binary). The Lagrangian decomposition formulation i.e., $C_{Ldecomp}$, has $T * (5m + 4n + 6)$ constraints and $T * (5m + 5n + 4)$ variables. For the case with 10/20 (modules/chips) and 4 time periods, this represents 584 constraints and 616 variables (248 of which are required to be binary).

Therefore the optimal solution to the inner minimization in $LDual_{LRelax}$ and $LDual_{LDecomp}$ can be obtained by solving s_1 and s_2 sub-problems respectively, which are *much smaller and much easier* than the original problem. The approach taken to solve $LDual_{LRelax}$ and $LDual_{LDecomp}$ is subgradient optimization in both cases, the details of which are explained later. In the next section we describe, in detail, the approximation algorithm that is used to solve $LDual_{LRelax}$ and $LDual_{LDecomp}$ using the relaxations/reformulations described in this section.

4.3 Algorithm for obtaining an Approximate Solution

We now describe the approximation algorithm that is used to solve Problems $LDual_{LRelax}$ and $LDual_{LDecomp}$, described in the previous section. As mentioned in Section 2.2 (in our discussion of the generic form of Problem C), for a given λ, the optimal solution to the Lagrangian problem provides a lower bound on the optimal solution to Problem C. The solution to the Lagrangian problem is then perturbed heuristically to generate a feasible solution which provides an upper bound on the optimal solution to Problem C. The optimal solution to Problem C is thus bounded between the upper and the lower bounds. The updating of λ is done by the subgradient optimization procedure. A description of algorithm to solve the Lagrangian dual problems must therefore include the following:

(a) Heuristic for generating feasible solution.

(b) Operational details of subgradient optimization implementation.

We first present detailed outlines of the procedures used for each of the problems, $Ldual_{LRelax}$ and $Ldual_{LDecomp}$. We discuss the heuristics used to perturb the optimal solutions to the Lagrangian problems C_{LRelax} and $C_{LDecomp}$ to generate a feasible solution to Problem C. These heuristics are called Heuristic 1 and Heuristic 2 respectively. Next we describe the implementation of subgradient optimization procedure which is similar for both the Lagrangian dual problems.

4.3.1 Heuristic 1: Lagrangian Relaxation of Capacity and Coupling Constraints

A flowchart of the algorithm used for Problem $Ldual_{LRelax}$ is given in Figure 4.2. The steps that are implemented may be described as follows:

Step 1 Initialize Lagrange multipliers γ_{pt}, δ_{pt} and π_{kt}.

Set current upper bound to $+\infty$.

Step 2 Solve independent, uncapacitated lotsizing problems A_{dec1} and B_{dec1} using the Wagner-Whitin (1958) algorithm.

Step 3 Solve overtime decision sub-problems A_{dec2} and B_{dec2} by the inspection procedure described in the previous section.

Step 4 Generate a feasible solution using the feasibility heuristic. If the feasible solution found is better than the "current" upper bound, store this feasible solution.

Step 5 Update the Lagrange multipliers using subgradient optimization.

Step 6 If the termination condition is satisfied, go to Step 7.

Else, go to Step 2.

Step 7 Print the solution.

A good starting point for the Lagrange multiplier for each relaxed constraint could be the value of the dual price associated with that constraint in the optimal solution to Problem C_{LP} i.e., LP relaxation of the primal problem C. Problem C_{LP} is equivalent to Problem C with the binary constraints relaxed.

At step 2 we solve problems A_{dec1} and B_{dec1}. Since the relaxed problems are uncapacitated, they are separable into independent problems for each individual item and hence solvable using the dynamic programming based Wagner-Whitin (1958) algorithm.

At step 3 we solve the overtime decision sub-problems for each time period at both the plants using the inspection procedure described earlier. The sum of optimal solutions for all these problems ($C_{Bdec1} + C_{Adec1} + C_{Bdec2} + C_{Adec2}$) provides a lower bound to problem C_{eq}. The results at steps 2 and 3 are likely to violate the capacity constraints and the link constraints. The capacity violations at the two plants are:

Solution Methodology for Lagrangian Relaxation

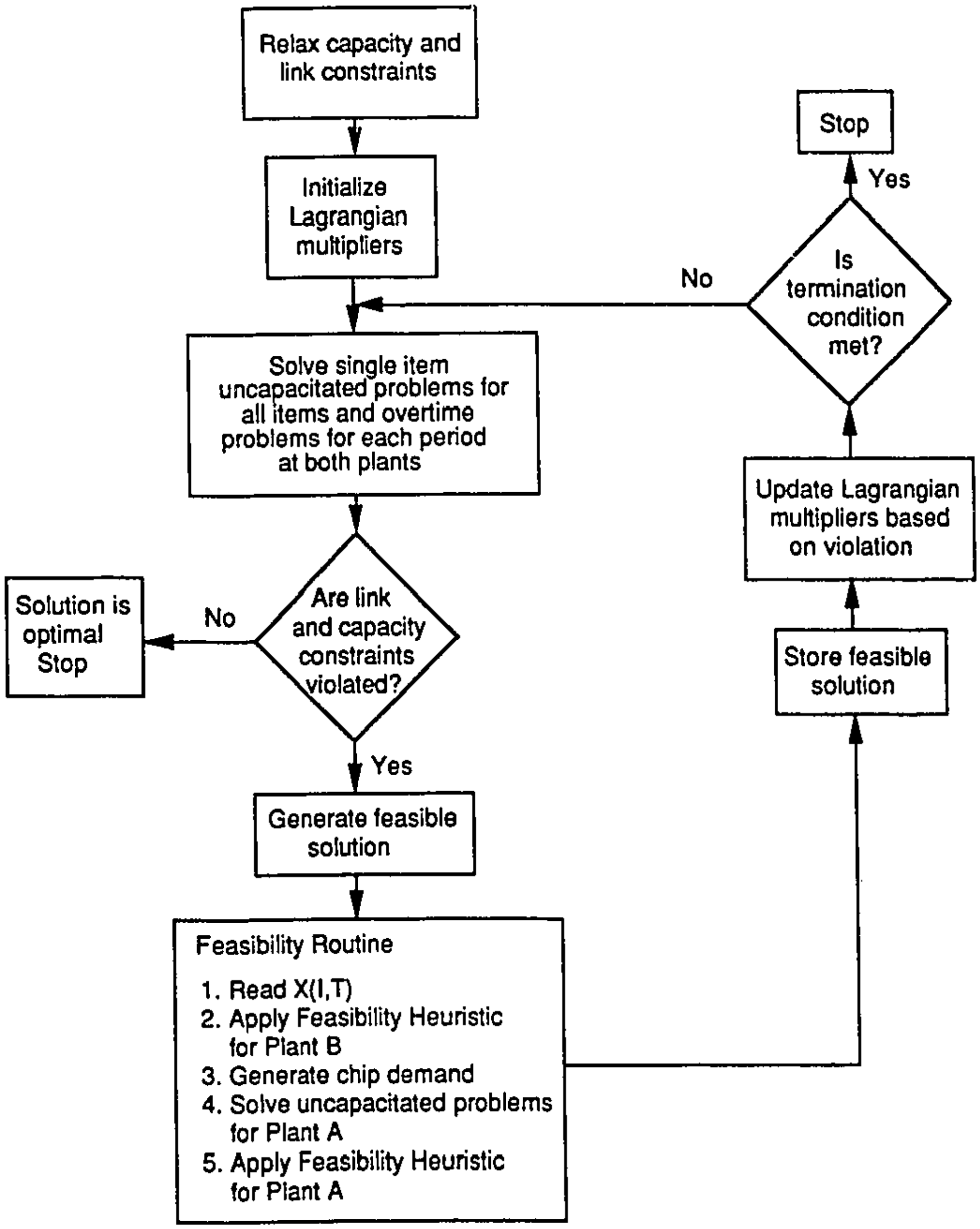

Figure 4.2: Flowchart for Lagrangian Relaxation based Algorithm

Plant A $\quad Cap_{1t} - \sum_{i=1}^{m}(s_i.Y_{it} + b_i.X_{it}) + O_{1t}$ for every t

Plant B $\quad Cap_{2t} - \sum_{i=m+1}^{m+n} s_i.Y_{it} + b_i.X_{it}) + O_{2t}$ for every t

The link constraint violation is ($\sum_{i=m+1}^{m+n} u_{ik}.X_{it}$ - P_{kt}) for each chip k in every period t.

At step 4, a feasibility heuristic is used to generate a feasible solution. The feasibility heuristic is somewhat inspired by the heuristic of Trigeiro et al.,(1989) which has been modified to incorporate the multi-plant structure of our problem. The heuristic begins with the uncapacitated results obtained at Step 2 for one of the plants and attempts to generate a solution that is capacity feasible at the given plant in every period. This is done by shifting production from the period in which there is capacity violation to another period in which capacity is available, while ensuring that demand requirements are satisfied in all periods. Once a feasible solution is found at the selected plant, the demand (or availability as the case may be) for chips is generated for the other plant and the Wagner-Whitin lotsizing problems and overtime sub-problems are re-solved at this second plant. The feasibility heuristic is then applied to the second plant in the same manner as for the first plant. Two scenarios are considered and these are explained later in this section. The objective of the feasibility heuristic is to minimize the Lagrangian cost of total shifting. The Lagrangian cost L_i of shifting a quantity X_1 of item i from period t to period t' has the following components – inventory, processing, setup, overtime and coupling – as detailed below. These costs are defined on the set of chips for Plant A, and on the set of modules for Plant B.

$$\underline{\text{Plant B}}\ L_i = \begin{cases} X_1.h_i(t-t') & \text{Inventory} \\ -s_i(\gamma_{t2}.(1-Y_{it}^a) - \gamma_{t'2}(1-Y_{it'}^b)) & \text{Setup} \\ -b_i.X_1(\gamma_{t'2} - \gamma_{t2}) & \text{Processing} \\ +f_{2t'}(z_{2t'}^a - z_{2t'}^b) - f_{2t}(z_{2t}^b - z_{2t}^a) & \text{Overtime-fixed} \\ +(v_{2t'} + \delta_{2t'}).(O_{2t'}^a - O_{2t'}^b) - (v_{2t} + \delta_{2t}).(O_{2t}^a - O_{2t}^b) & \text{Overtime-variable} \\ + \sum_{k=1}^{m} X_1.u_{ik}(\pi_{kt'} - \pi_{kt}) & \text{Coupling} \end{cases}$$

$$
\underline{\text{Plant A}}\ L_i = \begin{cases} X_1.h_i.(t-t') & \text{Inventory} \\ -s_i.(\gamma_{t1}(1-Y_{it}^a) - \gamma_{t'1}(1-Y_{it'}^b)) & \text{Setup} \\ -b_i.X_1(\gamma_{t'1} - \gamma_{t1}) & \text{Processing} \\ +f_{1t'}(z_{1t'}^a - z_{1t'}^b) - f_{1t}(z_{1t}^b - z_{1t}^a) & \text{Overtime-fixed} \\ +(v_{1t'} + \delta_{1t'}).(O_{1t'}^a - O_{1t'}^b) - (v_{1t} + \delta_{1t}).(O_{1t}^a - O_{1t}^b) & \text{Overtime-variable} \\ +X_1(\pi_{kt} - \pi_{kt'}) & \text{Coupling} \end{cases}
$$

The *inventory* term represents the holding cost that is either saved or expended depending on whether the shift is to an earlier period or to a later period. The *setup* term is the difference between the cost of setting up in t *versus* the cost of setting up in t'. This term will come into play for an item only if the shift results in either the elimination of the setup for this item in t or the incurring of a new setup in t' (or both of these simultaneously). Y_{it}^a here represents the setup variable for item i in period t after the proposed shift of X_1 units of that item from period t to t' while Y_{it}^b represents the setup variable for item i in period t' before the proposed shift. The *processing* term represents the impact of shifting of the processing (excluding the impact on setup time) of the item from the current period t to t'. The next two terms measure the impact on the overtime cost as a result of the shift. The superscript b refers to the value of the variable before the shift, while the superscript a refers to the modified value of the variable after the shift. The term relating to fixed overtime cost measures the net impact of shifting the item, on the fixed overtime cost. This term will become active only if the overtime status (as indicated by the value of the z_{1t} or z_{2t}) changes in either/both periods t and t' as a result of the shift. In a similar spirit, the variable overtime term measures the impact of the shift on the variable overtime costs. Finally the *coupling* term measures the effect of the shift in terms of the changes it produces on the demand for chips in the periods t and t'. The Lagrangian cost of shifting is determined by summing up each factor for all items as described above.

The heuristic comprises four passes : two backward passes and two forward passes. The first backward pass starts with the result of the uncapacitated problems obtained at Step 2. Beginning with the last period, production is shifted to earlier periods if there is a capacity violation in the current period. The objective as earlier stated is to minimize the Lagrangian cost of all shifting. Production can be shifted to either the previous period or to the previous setup of the same item. The item with the least Lagrangian cost is shifted first, and if capacity violation persists, other items are considered in ascending order of Lagrangian cost of shifting. Shifting halts when feasibility is achieved in the current period. No attempt is made to achieve feasibility in the period into which production is being shifted. This pass will ensure feasibility in all periods except *at most* the first one. The first forward pass begins with the result of the earlier pass. Beginning from the first period, production is shifted from periods with capacity violation to future periods. The target period is always the next period while the shift quantity is the end of period inventory. The objective once again is to minimize Lagrangian cost of shifting. Shifting halts when cumulative feasibility is achieved. The second backward pass is similar to the first backward pass, except for the initial state of the system. The second forward pass is similar to the first forward pass, except that it is more rigorous and continues shifting of production till feasibility (versus cumulative feasibility in the first pass) is achieved in the current period.

We note that this procedure will not necessarily find a feasible solution even if one exists. Our overall approach is different from that of Trigeiro et al., (1989) in that for each iteration we use an alternative heuristic to generate a feasible solution. Starting with the solutions of Problems $B_{LRelax1}$ (uncapacitated lotsizing problems for downstream module plant), and $B_{LRelax2}$ (overtime problems), we retain the setup decisions and the overtime scheduling decisions but discard the production quantities and the amount of overtime recommended. The following problem is then solved, using Y'_{it} and O''_{pt} as known parameters obtained from the solution of problems $B_{LRelax1}$ and $B_{LRelax2}$.

Problem B_{Heur2}

Minimize $C_{BHeur2} =$

$$\sum_{i=m+1}^{m+n} \sum_{t=1}^{T} h_i.I_{it} + \sum_{t=1}^{T} v_{2t}.O_{2t}$$

subject to

$$\begin{aligned} \sum_{i=m+1}^{m+n} b_i.X_{it}) - O_{2t} &\leq Cap_{2t} + O''_{2t} - \sum_{i=m+1}^{m+n} s_i.Y'_{it} \qquad \forall t = 1, \ldots, T \\ I_{it-1} + X_{it} - I_{it} &= d_{it} \qquad \forall t = 1, \ldots, T, \\ & \qquad\qquad i = m+1, \ldots, m+n \end{aligned}$$

Problem B_{Heur2} is a linear program with $T(n+1)$ constraints and can be easily solved for practical problems. If a feasible solution is found for B_{Heur2}, the demand for chips P_{kt} is generated by summing up over the set of all modules $\sum_{i=m+1}^{m+n} u_{ik}.X_{it}$. The setup and overtime decisions are retained from the uncapacitated lotsizing and overtime problems while the production quantities and the amount of overtime are discarded. A feasible solution for the problem at Plant A can be obtained by solving A_{Heur2}

Problem A_{Heur2}

Minimize $C_{AHeur2} =$

$$\sum_{i=1}^{m} \sum_{t=1-L}^{T-L} h_i.I_{it} + \sum_{t=1-L}^{T-L} v_{1t}.O_{1t}$$

subject to

$$\begin{aligned} \sum_{i=1}^{m} b_i.X_{it} - O_{1t} &\leq Cap_{1t} + O''1t - \sum_{i=1}^{m} s_i.Y'_{it} \qquad \forall t = 1-L, \ldots, T-L \\ I_{it-1} + X_{it} - I_{it} &= P''_{it} \qquad \forall t = 1-L, \ldots, T-L, \end{aligned}$$

$$i = 1, \ldots, m$$

$P''_{it}(= \sum_{i=m+1}^{m+n} u_{ik}.X_{it})$ represents the number of chips demanded by Plant B in period t and is obtained from the result of Problem B_{Heur2}. Problem A_{Heur2} is a linear program with $T(m+1)$ constraints and can be easily solved for practical problems. The solution to A_{Heur2} gives a feasible solution for Plant B. If feasible solutions are found at both plants, the sum of the two solutions gives a feasible solution to Problem C. Once again we note that the above procedure is not bound to find a feasible solution for either plants. Our procedure is superior to that of Trigeiro et al., (1989) because two potentially feasible solutions are generated at each iteration and the least cost solution is stored as the best known upper bound. Since the objective is to find the best feasible solution, it could pay off to generate a larger set of feasible solutions, starting from different initial conditions.

At step 5, the Lagrange multipliers for the capacity constraints and the link constraints are updated using subgradient optimization. The operational details of the subgradient optimization procedure will be explained in Section 5.1.2

At step 6 the algorithm checks if a pre-specified termination condition is satisfied. If the condition is satisfied, the algorithm terminates. If the stopping criterion is not met, the control loops back to Step 2 and the entire procedure is repeated. Termination criterion may either be the number of iterations of subgradient optimization or a pre-specified gap between the lower (Step 2 and 3) and the upper bounds (Step 4). Specifying number of iterations as a stopping criterion runs the risk of stopping subgradient optimization when we are still far from optimizing the Lagrangian dual. This may leave some of the reducible gap unexploited and as a consequence also affect the quality of the feasible solution. Specifying a target gap (between upper and lower bounds) as a stopping criterion may create problems if the actual duality gap between the optimal lower bound and the optimal primal solution is large. We discuss this issue in greater detail in the next chapter and propose alternatives to resolve it.

4.3.2 Heuristic 2: Lagrangian Decomposition

A flowchart of the algorithm used for Problem $Ldual_{LDecomp}$ is given in Figure 4.3. The steps that are implemented may be described as follows:

Step 1	Initialize Lagrange multipliers π_{kt}, ρ_{it}, γ_{it}, and δ_{pt}. Set current upper bound to $+\infty$.
Step 2	Solve independent, uncapacitated lotsizing problems $A_{LDecomp1}$ and $B_{LDecomp1}$ using Wagner-Whitin (1958) algorithm.
Step 3	Solve overtime decision sub-problems $A_{LDecomp2}$ and $B_{LDecomp2}$ by inspection procedure described in previous section.
Step 4	Solve knapsack sub-problems $A_{LDecomp2}$ and $B_{LDecomp2}$ using any standard continuous knapsack algorithm
Step 5	Generate feasible solution using feasibility heuristic. If feasible solution found is better than "current" upper bound, store feasible solution.
Step 6	Update Lagrange multipliers using subgradient optimization.
Step 7	If termination condition is satisfied, go to Step 7. Else, go to Step 2.
Step 8	Print solution.

All the above steps are identical to the procedure described for the Lagrangian relaxation case except that the sub-problems $B_{LDecomp1}$, $B_{LDecomp2}$, $A_{LDecomp1}$ and $A_{LDecomp2}$ are substituted for their corresponding counterparts of the Lagrangian relaxation case. We therefore omit the details of the heuristic for Lagrangian decomposition

Solution Methodology for Lagrangian Decomposition

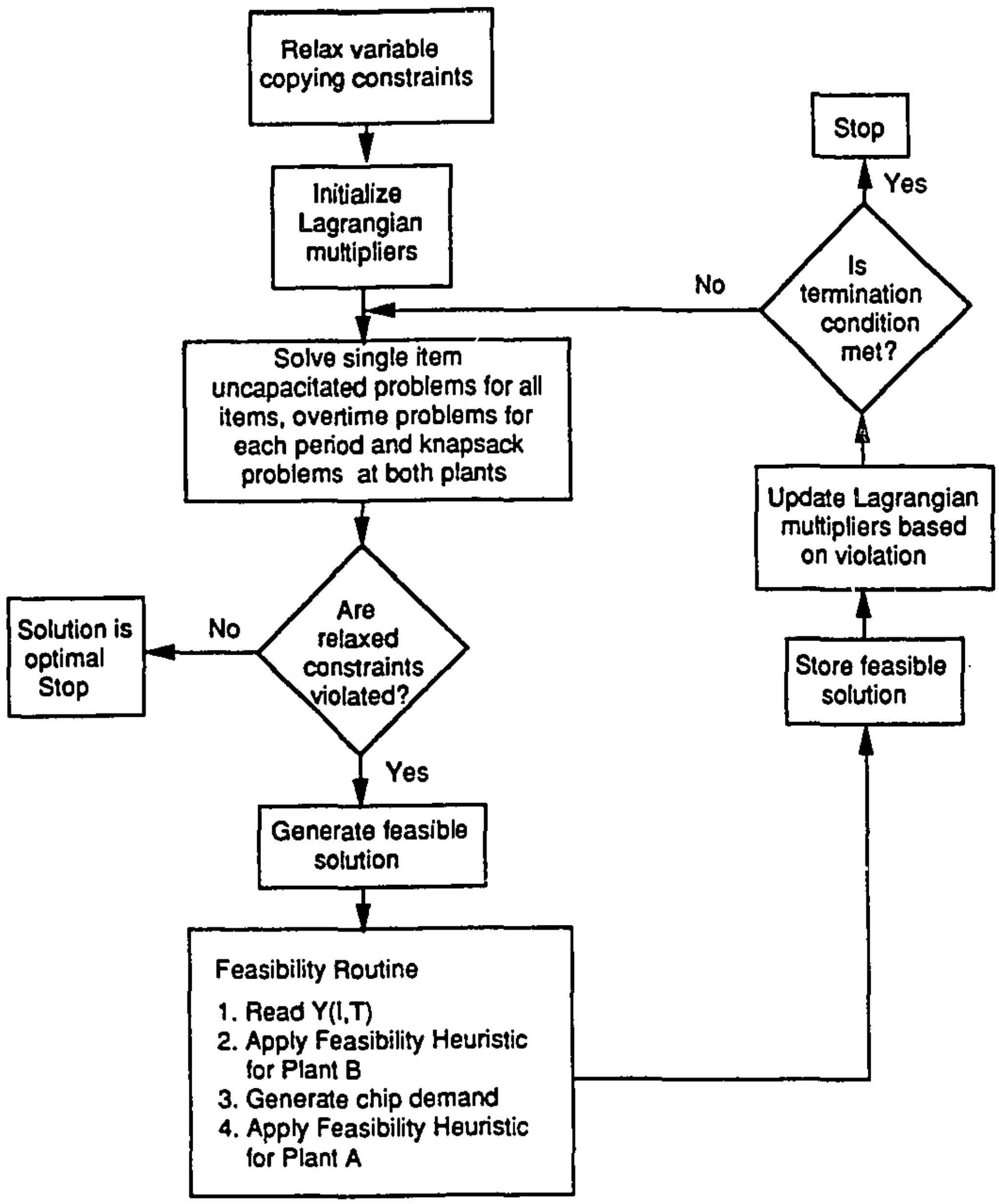

Figure 4.3: Flowchart for Lagrangian Decomposition based Algorithm

4.3.3 Computational Requirements

The computational requirements of a given algorithm may be characterized by its worst case running time. For details of analysis of algorithms, see Cormen et al., (1990). For the coordinated model, the running time needs to be characterized for several problems. For the Lagrangian problem using Lagrangian relaxation, the sub-problems that arise are the single item lotsizing problems and overtime problems. The complexity of the lotsizing problems (solvable by the Wagner-Whitin dynamic programming algorithm) is $O(T^2)$, while the overtime problems are solvable in $O(T)$ (T represents the total number of periods in the planning horizon). The feasibility heuristic can be characterized as a sorting problem and hence its worst case running time will be $O(N \log N)$, where N is the number of products (Cormen et al., (1990), page 172). In the scenarios that the coordinated model will be typically concerned with, N will be considerably larger than T and therefore the running time of the overall model will be bounded by $O(N \log N)$. When Lagrangian decomposition is used, the sub-problems are the single item lotsizing problems ($O(T^2)$), the overtime problems ($O(T)$), and fractional knapsack problems, solvable by a greedy heuristic ($O(N \log N)$). As for the previous case, the feasibility heuristic has a worst case running time of $O(N \log N)$. The overall model therefore has the same worst case running time as in the case of Lagrangian relaxation.

4.3.4 Subgradient Optimization

The procedure used for solving the Lagrangian dual problems i.e., $Ldual_{LRelax}$ and $Ldual_{LDecomp}$ is subgradient optimization. An important consideration that guided our selection of this procedure was its successful use for updating Lagrange multipliers in models relating to single stage problems, particularly in Trigeiro et al., (1989) and Diaby et al., (1992b). Given an initial value μ^0 of the multiplier (where the general symbol μ is used to represent δ_{pt}, γ_{pt} and π_{kt}), a sequence μ^k of multipliers is generated by the following rule.

$$\mu^{k+1} = \mu^k + t_k(b - Ax^k)$$

where $(b_p - Ax_p^k)$ is a generic representation of the violation of the p^{th} violated constraint (capacity constraint at Plants A/B or link constraint) on the k^{th} iteration, x^k is an optimal solution to the decomposed problems A_{dec1}, A_{dec2}, B_{dec1} and B_{dec2} and t_k is a positive scalar step size for the k^{th} iteration. The step size we have used is the following:

$$t_k = \frac{\lambda_k(Z^* - (C_{Adec1}(\mu_k) + C_{Adec2}(\mu_k) + C_{Bdec1}(\mu_k) + C_{Bdec2}(\mu_k))}{\sum_p (b - Ax^k)^2}$$

λ_k is a positive scalar (≤ 2) and Z^* is an upper bound on the decomposed problems obtained usually by applying a heuristic (in our case we use the result from the uncoordinated problem and replace it whenever a better feasible solution is obtained by the cost of this improved solution). The denominator term represents the sum of squares of the violation (of the p violated constraints). Justification for this stepsize is provided in Held, Wolfe and Crowder(1974). The sequence λ_k is determined by setting $\lambda_0 = 2$, and reducing it by a fixed multiplier whenever the lower bound does not improve after a given number of iterations. The multiplier and the number of iterations used, were determined through preliminary testing. The above procedure terminates after the duality gap between the upper and lower bounds is within specified limits or after reaching a specified iteration limit. Specifying the number of iterations as a stopping criterion runs the risk of stopping subgradient optimization when we are still far from optimizing the Lagrangian dual. This may leave some of the reducible gap unexploited and as a consequence also affect the quality of the feasible solution. Specifying a target gap (between upper and lower bounds) as a stopping criterion may create problems if the actual duality gap between the optimal lower bound and the optimal primal solution is large. We discuss this issue in greater detail in the next chapter. We also propose an experimental design for our study and present results relating to our experimentation.

It has been established that a divergent series of multipliers would lead to the Lagrangian dual converging to its optimal value (Goffin (1977) and Fisher (1985)). We implemented one example using a divergent series of multipliers for 6000 iterations. Convergence however was extremely slow and this series did not outperform the sequence of multipliers explained above.

Chapter 5

Results

From a managerial perspective, it is important to know the extent of benefits that can accrue by adopting superior models and to identify critical parameters that significantly affect these benefits. We take this as the primary focus of this research. We take the view, as in Geoffrion (1974), that the Lagrangian dual need not be solved optimally and need not be devoid of duality gap in order to be useful. An important aspect of Lagrangian decomposition/relaxation methodology is to generate improved feasible solutions for complex integer programs like the one being considered here. Nevertheless, from an algorithmic point of view, it is important to evaluate the performance of the model against the true optimal. If the model gives near optimal results, we can be confident about the validity of our conclusions and the managerial implications derived from them. In this chapter, we address both these issues of analysis of cost benefits of coordination and the analysis of the performance of Lagrangian relaxation and Lagrangian decomposition methodologies as compared to the true optimal. The first section of this chapter provides a summary of the results relating to the cost benefits of implementing the coordinated model, and managerial implications of these results. Cost benefits are defined as the percentage reduction in total costs (comprising inventory and fixed and variable overtime costs) for the coordinated versus uncoordinated model. Detailed results are presented for a typical problem to provide a flavour for the underlying source of cost benefits accruing from multi-plant coordination. Next, the experimental design and the choice of the experimental factors are explained. The experimental design is implemented for both Lagrangian relaxation and the Lagrangian decomposition cases, as discussed in the previous chapter. For the purpose of analysis of the cost benefits of coordination, we use the higher cost improvement obtained from the two methods. Cost improvement results are provided for a variety of different cases and the implementation aspects of the coordinated model are discussed. Statistical analysis of the variance attributable to the effect of each of the experimental factors is presented. Insights from this analysis are used for defining guidelines for managerial action. At a broad level, we address how and where the results from this research can be applied by managers. In the second section of this chapter, we focus on the convergence results of Lagrangian relaxation versus Lagrangian decomposition for the multi-plant coordination problem. The performance measure used for evaluating convergence is the residual gap between the best lower bound obtained and the best known feasible solution on reaching a pre-specified termination condition. Details are discussed later in the chapter.

5.1 Summary of Cost Benefits of Coordination and Managerial Implications

5.1.1 Structure of Solution: Importance of Coordination

We first present detailed results for a typical problem (see Table 5.1). The objective of this exercise is to provide insights into the mechanism of cost improvement that accrues from coordination. Numerical details of this problem, i.e., processing and setup times, holding costs, bill of material etc., are given in Appendix 1.

	Uncoordinated	Coordinated	% Change in Cost (as % of TOTAL Coordinated cost)
Plant B			
Inventory Holding	\$5746.60	\$6301.08	+6.4
Fixed Overtime	\$ 120.00	\$ 120.00	0.0
Variable Overtime	\$ 526.53	\$ 557.55	+0.4
Total(Plant B)	\$6393.13	\$6978.63	+6.8
Plant A			
Inventory Holding	\$ 986.23	\$ 23.1	-11.2
Fixed Overtime	\$ 120.00	\$120.00	0.0
Variable Overtime	\$ 1444.14	\$1475.76	+0.4
Total(Plant A)	\$2550.37	\$1618.86	-10.8
TOTAL	\$8943.50	\$8597.49	-4.0

Table 5.1: Results for sample problem

The total costs of inventory holding and fixed and variable overtime costs for Plants A and B for this example (shown in Table 5.1) are \$8943.50 for the uncoordinated approach and \$8597.49 for the coordinated approach. The % change in cost shown in the last column of Table 5.1 represents the difference between the uncoordinated and the coordinated approaches as a percentage of the coordinated solution (i.e., as a % of \$8597.49). For this example, the coordinated approach results in a total cost reduction of 4.0%. As a percentage of the cost of the coordinated solution, the coordinated approach leaves Plant B worse off by 6.8% but results in a reduction of 10.8% in the costs for Plant A and consequently a reduction of 4% in the total costs of both plants. We have observed that this trend in cost changes is generalizable, i.e., in all instances where there is a reduction in total cost by implementing the coordinated model, this reduction comes from a decrease in the costs at Plant A which more than offsets the increase in costs at Plant B. Some reflection will convince the reader that this is intuitively meaningful. Recall that in the uncoordinated approach, costs are first optimized for Plant B independently, the resulting demand for chips is communicated to Plant A, and the costs for Plant A are then optimized subject to this demand. Since the uncoordinated model is biased in favour of Plant B, which drives the model, there is greater concern for module inventory and constraints on the process at Plant B. This is really the crux of the issue. With the uncoordinated model, the firm will gain at one end of its supply chain but may lose considerably at another so that overall, the firm may be worse off. In contrast, the coordinated model, through the mechanism of Lagrange multipliers, considers the interaction between the costs and the processes at both the plants by associating a Lagrangian cost with the violation of every constraint. Coordination will therefore diminish Plant B's cost performance while improving that of Plant A. The cost increase at Plant B will at most be equal to the cost reduction at Plant A, in which case both approaches perform equally well. However, if the cost saving at Plant A exceeds the cost increase at Plant B (as is the case in the example above), the coordinated approach outperforms the uncoordinated approach. The coordinated model ensures benefits for the firm by viewing the total costs globally for both the plants, and by capturing the interdependence of the processes at both plants.

We further explore the distinguishing features of the coordinated approach vis-a-vis the uncoordinated approach by studying the behaviour of inventory costs and setups at both the plants in the above example. Inventory of modules and chips and the setups incurred at both plants for the two approaches are shown in Tables 5.2 and 5.3.

Two important points can be observed from Tables 5.2 and 5.3. First, the reduction in total cost comes through a counter-intuitive interchange of inventory between the two plants. Carrying a larger inventory of the higher value modules actually results in a reduction of overall costs for the two plants taken together. This is because in the coordinated approach, the increased cost for Plant B, as a consequence of carrying a higher inventory of modules is compensated by higher cost savings for Plant A due to reduced inventory of chips. Savings at Plant A are achieved mainly through reduced inventory carrying requirements due to better match between the demand for chips and the available processing resources. For instance, in the uncoordinated model, the total inventory (summed over all periods) at Plant B for modules 2 and 3 was 262 and 32, respectively. The larger inventory of module 2 was mainly due to its lower cost compared to module 3. In the coordinated model, the total inventory of modules 2 and 3 was 39 and 273, respectively. In effect, the coordinated model swapped the inventory from the lower value module 2 to the more costly module 3 (and increased the total inventory of modules 2 and 3 marginally). The reason why this swap makes sense is that it drastically reduces the inventory of chips 3 and 4 at Plant A (71 and 210 for the uncoordinated model versus 0 and 7 for the coordinated model), thereby reducing the total costs at both plants. The coordinated approach has therefore

Uncoordinated					Coordinated				
Production									
	Periods					*Periods*			
Modules	1	2	3	4	*Modules*	1	2	3	4
1	50	0	0	0	1	50	0	0	0
2	56	180	0	57	2	10	89	97	97
3	37	0	183	121	3	93	88	80	80
Inventory									
	Periods					*Periods*			
Modules	1	2	3	4	*Modules*	1	2	3	4
1	35	8	2	0	1	35	8	2	0
2	48	171	43	0	2	2	34	3	0
3	1	0	31	0	3	57	144	72	0
Setups									
	Periods					*Periods*			
Modules	1	2	3	4	*Modules*	1	2	3	4
1	1	0	0	0	1	1	0	0	0
2	1	1	0	1	2	1	1	1	1
3	1	0	1	1	3	1	1	1	1

Table 5.2: Inventory and Setup Analysis for Plant *B*

Notes

1. Production and inventory figures represent number of modules for each period.
2. Setup = 1 if there is a setup for module i in period t, = 0 otherwise.
3. Numerical details of holding cost, processing and setup times etc., are in Appendix 1.

Uncoordinated					Coordinated				
Production									
	Periods					*Periods*			
Chips	1	2	3	4	*Chips*	1	2	3	4
1	50	0	0	0	1	50	0	0	0
2	106	180	0	57	2	60	89	97	97
3	130	251	295	299	3	196	265	257	257
4	119	0	147	75	4	100	81	80	80
Inventory									
	Periods					*Periods*			
Chips	1	2	3	4	*Chips*	1	2	3	4
1	0	0	0	0	1	0	0	0	0
2	0	0	0	0	2	0	0	0	0
3	0	71	0	0	3	0	0	0	0
4	82	82	46	0	4	7	0	0	0
Setups									
	Periods					*Periods*			
Chips	1	2	3	4	*Chips*	1	2	3	4
1	1	0	0	0	1	1	0	0	0
2	1	1	0	1	2	1	1	1	1
3	1	1	1	1	3	1	1	1	1
4	1	0	1	1	4	1	1	1	1

Table 5.3: Inventory and Setup Analysis for Plant A

Notes

1. Production and inventory figures represent number of chips for each period.
2. Setup = 1 if there is a setup for chip i in period t, = 0 otherwise.
3. Numerical details of holding cost, processing and setup times etc., are in Appendix 1.

been effective in capturing the interaction between the processes at the two plants. The key to effective multi-plant coordination is therefore to determine the optimal location of inventory.

The second observation from Tables 5.2 and 5.3 relates to productive resources expended in setups in the two approaches. In the uncoordinated approach, there are a total of 11 setups for chips at Plant A and 7 setups for modules at Plant B, while in the coordinated approach there are 13 setups for chips and 9 setups for modules. This observation makes intuitive sense. In the event of variable capacity requirements in different periods coupled with tight capacity restrictions, the uncoordinated approach tends to combine demands of two or more periods in attempting to reduce the number of setups at Plant B. This forces reduced number of setups at Plant A too, in order to satisfy Plant B's requirements for chips. The coordinated model by contrast, is concerned with obtaining the least cost fit between the capacity requirements and the available capacity, at the two plants. In general therefore, the coordinated model recommends a larger number of setups as compared to the uncoordinated model.

To recapitulate, coordination achieves cost savings over the uncoordinated approach by taking a global view of the inventory and by considering the implications of the processes at both the plants. This technique results in increased costs at Plant B but reduces the costs at Plant A. The inventory at the two plants changes in such a way that the overall costs at the two plants are reduced. Simultaneously, there is a trend in the coordinated approach towards an increased number of setups. Overall, the coordinated approach captures the interaction between the processes at the two plants and hence outperforms the uncoordinated approach. An important issue that arises is the identification of critical parameters whose presence or absence enhances benefits of coordination. This will assist in establishing guidelines for managerial planning. We consider this issue next when we present the experimental design for our study.

5.1.2 Experimental Design

The objective of a good experimental design in the context of this research is to evaluate the quality of results obtained for the coordinated versus the uncoordinated approach . For validity of this evaluation process, a variety of data which has practical implications for the multi-plant coordination problem must be considered. Ideally we would like to experiment on existing data sets from literature in order to obtain comparative results relative to the existing methods. Unfortunately, since this problem has not been adequately addressed, no standardized data sets are available. All experimentation was therefore carried out on randomly generated problems. The experimental design showing the major parameters of interest, is shown in Table 5.4. The choice of the experimental factors and their different levels were guided by their relevance to the real life situation at IBM as well as practical considerations of time. Details of each parameter are described below.

A. Demand

The average demand level across all products was selected as 50 units/period. With this value, the demand for each product in every period was generated for different levels of coefficient of variation (c.v.) in two steps. First, the average demand for each product was generated by obtaining values from a uniform distribution with the given c.v. and average demand of 50. Next for each product, the period-wise demand was generated by obtaining values for given c.v. and the average obtained in the first step. For each of the steps referred to above, two levels of c.v. were considered - high (0.57) and low (0.10). The high c.v. case represents the situation where the range (i.e., maximum value *minus* minimum value) is twice the average value. The low c.v. case displays much less variability with range equal to 35% of the mean. We therefore obtained four different scenarios each representing different levels of variability of demand between products and between periods. These scenarios are a good approximation of the real

Parameter	Value
Demand	Average demand = 50 units/period Average demand for products is uniformly distributed with coefficient of variation = 0.10, 0.57 Demand between periods is uniformly distributed with coefficient of variation = 0.10, 0.57
Capacity	Average lot-for-lot utilization = 0.85, 0.95
Setup/ Processing Time	Setup time to processing time ratio = 2, 5 coefficient of variation of setup times = 0.10, 0.57 coefficient of variation of processing times = 0.10, 0.57
Holding Cost	Holding cost of chips is uniformly distributed with average = 3.0 and coefficient of variation = 0.10, 0.57 Holding cost of module i = X_i*Y where X_i is the sum of holding costs of all chips required for module i and Y ≡ U(1,2)
Overtime Cost	Fixed overtime cost: High (= 100*AHC), Low (= 10*AHC) Variable overtime cost: High(=5*AHC per unit of overtime), Low(= AHC per unit of overtime) where AHC is the average holding cost of chips (= 3.0)
Number of Modules/ Chips	3/4, 10/20

Table 5.4: Experimental Design

situation in fashion based industries (e.g., toys and clothing) and high technology industries (e.g., computers and consumer electronics) in which product life cycles are characterized by rapid growth, maturity and decline phases as well as substantial seasonal variation (Kurawarwala and Matsuo, 1992). Demand variability between products and between periods will be high if different products show dissimilar seasonality effects at distinct times of the year. For example, during the high seasonality phase (around Christmas in North America) the sales of novelty items may rise and that of regular items may dip giving high variability between products. In order to avoid the number of combinations from exploding, we retained only two combinations i.e., high variability of demand (c.v. = 0.57 for both steps) and low variability of demand (c.v. = 0.10 for both steps) for our experimentation.

B. Capacity

Capacity utilization was specified on a lot-for-lot basis i.e., assuming that there is a setup for each item in the period in which the demand occurs. Processing and setup time requirements of all periods were added and the available regular and overtime resources were computed by dividing these total requirements by medium and high target capacity utilization levels of 0.85 and 0.95. Overtime resources were taken as 25% of the regular resources. In order to ensure cumulative feasibility of the problem, demand was reduced in the first few periods and corresponding amounts added back in the later periods. As observed by Trigeiro et al., (1989) this would approximate the real life situation somewhat since some of the demand of the initial periods can be satisfied through the starting inventory carried over from the past.

C. Costs

(i) Holding Cost

For Plant A, holding costs for different chips were generated by obtaining values from a uniform distribution with a mean of 3.0, with c.v. being high (0.57) or low (0.10). For Plant B, the holding cost for each module i was computed as the product of a multiplier, Y, and the sum of holding costs of all chips required by this module, X_i. The value of the multiplier, Y for each module was chosen randomly over the uniform interval (1.0, 2.0), i.e., the value added at Plant B ranged from a minimum of 0.0 to a maximum of the total value added at Plant A to all the chips used in this module, and was uniformly distributed. The zero value addition represents the case where a single chip is mounted on a plastic board at Plant B and hence the increase in value (cost of the plastic board) at Plant B is negligible.

(ii) Overtime Costs

Varying levels of fixed (5.0 and 100.0 times average holding cost) and variable overtime costs (1.0 and 5.0 times average holding cost per unit of overtime) were selected, providing four different options. The high fixed overtime cost case corresponds to the scenario where additional inter-plant transportation costs are included in the fixed overtime costs if overtime is scheduled.

D. Processing and Setup Times

(i) Processing Times

The average unit processing time of chips at Plant A and of modules at Plant B was taken to be 1.0 time unit with the actual processing times being computed from a uniform distribution with c.v. high (0.57) or low (0.10). The choice of different processing time c.v.s was guided by an important characteristic of high-tech industries like computers and telecom equipment manufacturing etc. These industries are characterized by continuously changing product mix and product designs over the product life cycle. In this kind of a manufacturing environment, there are occasions when certain products which are not fully integrated into the manufacturing system due to unstable product design require significantly higher processing times than the other products on a critical workcenter. This results in a high c.v. of processing times. Once the designs stabilize, through learning, the variability of processing times across products and hence the c.v. of processing times decreases. In Appendix 2, we demonstrate how the presence of even a few such products can considerably affect the c.v. of processing times. Testing the coordinated model under different c.v.'s of processing times therefore increases the appropriateness of the experimental design since in a real life system, it is likely that the processing time c.v.'s would show a cyclical trend over a period of time as product designs go through periods of relative stability or instability.

(ii) Setup Times

Setup times were specified through two parameters, the setup/processing time ratio and the c.v. of setup times. Given the emphasis on setup time reduction, it is important to test how cost improvements of coordination are affected as the setup/processing time ratio changes. Using different c.v.s (0.57, 0.10) of setup times, reflects once again the concern that the industries which are the focus of this study exhibit rapid product changes which might necessitate different setup requirements for different products.

E. Problem Size

Different number of products (3 modules/4 chips, 10 modules/20 chips) were selected at the two plants. The length of planning horizon was fixed at 4 periods as this was the planning horizon that was being used at IBM. Longer planning horizons are not very useful given the dynamic nature of demand in the kind of manufacturing systems under study.

The above experiments were run on VAX3100 using a computer code developed in FORTRAN. Initial experimentation revealed that the overtime costs did not have a major impact on our results. We therefore limited our study to the consideration of only one of the four possible combinations for overtime costs i.e., low fixed and low variable cost of overtime. These experiments were also useful from the point of view of exploring good parameters for use in subgradient optimization (see section 4.3.4 on page 100). The parameters of interest relate to the number of iterations after which the multiplier λ_k must be reduced and the multiplier which must be used for the reduction. In general, we found that the lower bounds obtained by using different values of the above parameters seem to be robust i.e., results were not affected drastically by the values chosen for these parameters. The best results were obtained for $\lambda_0 = 2.0$, and dividing λ by 1.05 after every 15 iterations. Hence these values were adopted for subgradient optimization in all the subsequent experimentation.

For a given problem size (in terms of number of products and periods), our experimental design comprised 2^6 factorial experiments incorporating the main effects of the six factors and 57 interaction effects of different combinations of the factors and their levels. The six factors being investigated were measures of demand variability, capacity utilization, setup/processing time ratio, setup time variability, processing time variability and holding cost variability. As depicted in Table 5.4, we used the coefficient of variation as a measure of the variability of the factors. Each experimental factor was considered at two different levels, as explained before. We therefore analyze the effect of $2^6(=64)$ treatments. The statistical validity of the results increases as the number of replications for each treatment increases. However due to time considerations, we limited the number of independent replications to 4 for each treatment in the case of Lagrangian relaxation (3 modules/4 chips) and 2 in the case of Lagrangian decomposition (3 modules/4 chips). These experiments suffice for studying the relative impact of the different experimental factors. For the larger problems (10 modules/20 chips), we restricted ourselves to about 32 problems without replication using Lagrangian relaxation only. This yields 416 ($= 256 + 128 + 32$) different problems encompassing a wide variety of scenarios differing in terms of the values of the experimental factors.

5.1.3 Summary of Cost Improvement Results

In this section, we present the summary of cost improvement results of coordination for our experimentation in order to determine the value of coordination as a function of different problem parameters. The measure of cost improvement is the reduction in the cost of the coordinated solution over the uncoordinated solution, expressed as a percent of the coordinated solution. For the analysis of the cost benefits of coordination the following procedure was adopted. The cost benefits of coordination were determined for each treatment (or combination of experimental factors) using both Lagrangian relaxation and Lagrangian decomposition methodologies as described before, and the higher cost reduction obtained from the two methods was used for the analysis that is presented. As mentioned earlier, four independent replications were made for Lagrangian relaxation (3 modules/4 chips) and two independent replications were made for Lagrangian decomposition (3 modules/4 chips). We have classified these problems into four distinct cases as follows:

Case 1 High c.v.(demand), High Capacity Utilization, for 3 modules/4 chips case

Case 2 Low c.v.(demand), High Capacity Utilization, for 3 modules/4 chips case

Case 3 High c.v.(demand), Low Capacity Utilization, for 3 modules/4 chips case

Case 4 Low c.v.(demand), Low Capacity Utilization, for 3 modules/4 chips case

Case 5 High c.v.(demand), High Capacity Utilization, for 10 modules/20 chips case

Case 6 High c.v.(demand), Low Capacity Utilization, for 10 modules/20 chips case

Cases 1 to 4 summarize the cost improvement results for 64 problems while Cases 5 and 6 each represent the results of 16 problems. These results and the values of parameters used in each case are shown in Tables 5.5 to 5.10 (see pages 119 to 124). These results show that the coordinated model outperforms the uncoordinated model in approximately 65% of the problem instances while in the rest of the problems, the two approaches perform equally well. There seems to be considerable variability in the cost improvement, with average improvement for Cases 1 to 4 being 1.42%. However in certain cases, the improvement exceeds 11.0%. There also appears to be significant variability between independent replications of the same set of experimental conditions.

Insert Tables 5.5 to 5.10 here.

We formally investigated whether the means of various treatments were different, using the analysis of variance (ANOVA) approach. The results of this analysis are presented in Appendix 3. The F value for the ANOVA model is 0.79 which leads us to interpret that the average cost improvements of different treatments in our experiments are not significantly different. Some variability in the ANOVA model was artificially damped because the value of inputs for two different treatments were distinct only for those variables that had different input values. For the variables that had the same input values in two treatments, the actual values for the variables were identical. For example, in Table 5.5, the input data sets for the first two problems in the first row are identical, except for the values of the setup times of items (as a consequence of the change in the variable spratio). Under these conditions, the results of ANOVA provide evidence of the robustness of our results. As noted by Fisher and Rinnooy Kan (1988), one measure of the robustness of heuristics is the effect that data transformations have on performance - the smaller this effect, the more robust this method might be called.

Among the individual factors, we note that the main effects of the setup/processing time ratio and of capacity utilization are significant at $\alpha = 0.02$. Care however must be taken in interpreting the main effects because the three way interaction effect of (capacity utilization by c.v.(demand) by setup/processing time ratio) is also significant at $\alpha = 0.02$. We first interpret the main effects and then present

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	0.2	0.0	1.1	0.0
		5.9	3.4	2.7	5.0
(proc.		1.7	0.7	0.0	0.2
time)		0.0	0.0	1.1	0.0
	c.v.(set)=0.10	0.4	0.0	0.0	0.0
=		5.7	3.7	3.2	5.0
		0.0	0.3	0.0	0.0
0.57		0.9	0.0	2.7	0.0
c.v.	c.v.(set)=0.57	0.0	0.2	0.0	0.0
		10.6	0.7	0.0	0.1
(proc.		2.2	0.0	2.9	0.6
time)		0.6	0.0	1.2	0.5
	c.v.(set)=0.10	0.0	0.0	0.0	0.0
=		1.6	1.1	1.5	0.0
		4.0	0.0	4.3	0.3
0.10		0.7	0.0	0.5	0.0

Table 5.5: Percentage Cost Improvement for Case 1

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	0.0	0.0	1.3	0.0
		2.7	3.8	5.0	5.0
(proc.		0.0	0.0	1.3	0.0
time)		0.0	0.0	0.5	0.0
	c.v.(set)=0.10	0.0	1.0	0.0	3.2
=		3.2	2.5	5.0	5.0
		0.0	0.0	1.2	0.0
0.57		0.0	0.0	0.0	0.0
c.v.	c.v.(set)=0.57	0.0	0.0	0.0	0.0
		2.2	1.2	0.0	0.0
(proc.		0.9	1.5	0.0	0.0
time)		1.3	0.0	0.0	0.0
	c.v.(set)=0.10	0.0	0.0	0.0	0.0
=		1.6	0.5	0.0	0.0
		0.7	1.9	0.0	0.2
0.10		0.0	0.0	1.0	0.0

Table 5.6: Percentage Cost Improvement for Case 2

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.10)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	2.3	0.0	1.1	0.2
		5.1	4.9	2.8	0.6
(proc.		0.0	0.0	0.0	0.9
time)		0.0	0.0	2.6	0.0
	c.v.(set)=0.10	0.0	0.0	0.9	0.5
=		5.9	5.1	3.1	0.1
		0.3	0.0	0.2	0.0
0.57		0.0	0.0	2.8	1.1
c.v.	c.v.(set)=0.57	5.6	10.4	0.9	0.9
		1.4	6.4	0.0	0.0
(proc.		0.0	0.2	1.4	5.0
time)		1.6	0.7	2.5	0.4
	c.v.(set)=0.10	4.3	10.1	0.0	0.0
=		0.0	0.4	0.0	0.0
		0.0	0.3	1.6	1.3
0.10		2.4	1.4	1.6	0.9

Table 5.7: Percentage Cost Improvement for Case 3

Legend

c.v.(hcost) : coefficient of variation of holding cost

c.v.(set) : coefficient of variation of setup times

c.v.(proctime) : coefficient of variation of processing times

spratio : setup/processing time ratio

c.v.(demand) : coefficient of variation of demand (0.57)

capacity utilization : 0.85

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	3.7	0.7	6.4	1.2
		7.2	7.2	0.2	0.0
(proc.		1.0	0.0	2.6	1.2
time)		0.3	0.2	5.0	0.2
	c.v.(set)=0.10	2.0	0.0	0.4	0.3
=		11.2	4.6	3.8	2.0
		0.0	0.0	2.6	1.2
0.57		0.6	0.2	5.0	0.2
c.v.	c.v.(set)=0.57	2.9	2.1	0.0	0.0
		4.9	0.0	1.4	0.0
(proc.		2.1	2.0	3.2	5.4
time)		0.7	0.6	3.1	0.4
	c.v.(set)=0.10	2.4	2.1	1.1	0.0
=		3.5	0.0	0.6	0.5
		2.3	2.6	8.0	1.6
0.10		1.8	0.0	1.0	1.6

Table 5.8: Percentage Cost Improvement for Case 4

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.10)
capacity utilization	:	0.85

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v. (proc. time) = 0.57	c.v.(set)=0.57	2.1	0.0	3.7	0.0
	c.v.(set)=0.10	0.0	0.0	1.6	0.2
c.v. (proc. time) = 0.10	c.v.(set)=0.57	10.2	0.2	3.7	1.7
	c.v.(set)=0.10	2.6	1.4	3.8	1.8

Table 5.9: Percentage Cost Improvement for Case 5

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v. (proc. time) = 0.57	c.v.(set)=0.57	2.7	1.9	5.1	4.3
	c.v.(set)=0.10	2.3	0.5	4.3	2.1
c.v. (proc. time) = 0.10	c.v.(set)=0.57	5.9	4.5	3.2	1.2
	c.v.(set)=0.10	3.2	1.4	4.0	2.8

Table 5.10: Percentage Cost Improvement for Case 6

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.85

an analysis of the interaction effects in Tables 5.11 and 5.12 respectively.

To highlight the impact of different experimental factors, we summarize average cost improvement results for all of Cases 1 to 4 in Table 5.11. Table 5.11 is arranged so that inferences may be made about the relative impact of the experimental factors on cost improvements due to coordination. An important observation relates to the effect of capacity utilization. For all factors, we observe that the cost improvements are significantly higher when the capacity utilization is 0.85 than when it is 0.95. This is not surprising because when capacity utilization is lower, the set of production schedules which are jointly feasible in terms of capacity at both the plants is larger. Hence the likelihood increases that one of these would yield a better solution compared to the uncoordinated approach. With high capacity utilization, there is a lesser number of jointly feasible schedules and hence the flexibility to choose schedules is also lower. Of course the value of coordination will reduce if the available capacity is very high. In the extreme case with infinite capacity at both the plants (uncapacitated case), the value of coordination will be zero because Plant A can accommodate the optimal schedule of Plant B at no greater cost than its own optimal schedule. Coordination therefore exhibits enhanced value at intermediate levels of capacity utilization but relatively little value when the capacity utilization is very high or very low. From the practical point of view this is important because managers can be assured of benefits of coordination in the intermediate range of capacity utilization (80% to 95% on average, on critical resources) which they typically encounter in real life (see Williams et al., 1992).

The other significant observation from Table 5.11 relates to setup to processing time ratios. On an average, cost improvement with a high setup/processing time ratio (spratio=5) is 77% higher for the high capacity utilization case and 53% higher for the low capacity utilization case than with a low setup/processing time ratio (spratio=2). This has important implications for firms where the setup/processing time ratio is high. Cost benefits arising from coordination between plants can be invested in a setup time reduction program. If the firm is successful in reducing the setup/processing time ratio, it will see a reduced value for the coordinated model. However, reduced setup times will free capacity resulting in overall reduction in costs. The philosophy of coordination therefore fits in with the long term perspective of improving the competitive position of the firm.

The previous discussion on the main effects of the significant factors must be viewed in conjunction with an analysis of the three way interaction effect of (capacity utilization by c.v.(demand) by setup/processing time ratio). In order to highlight the interaction effect between the three factors, we analyze these factors exclusively in Table 5.12.

We can observe from Table 5.12 that benefits are significantly higher for capacity utilization = 0.85 as compared to capacity utilization = 0.95, under three out of the four possible combinations of the other two factors. However, when c.v.(demand) = 0.57 and spratio = 5, there is very little difference between the two levels of capacity utilization. Similarly, the cost benefits are higher for high value of spratio (=5), for two out of the four possible combinations of the other two factors, while in the other two scenarios, there is little difference in the cost benefits with high and low values of spratio. The highest benefits occur for the case when spratio = 5, capacity utilization is 0.85 and c.v.(demand) = 0.10. In general, the interpretation of the main effects of capacity utilization and the setup to processing time ratio that was described earlier, holds true because in none of the cases does the effect become completely reversed.

Insert Tables 5.11 and 5.12 here.

Of the other factors, the variability of demand and processing times show different trends in relation to cost improvements depending on the capacity utilization. With high capacity utilization, the benefits are higher for high coefficient

c.v. (demand)	c.v. (hcost)	c.v. (proctime)	c.v. (set)	spratio	Number of Problems	% Cost Improvement	
						a	b
0.57					128	1.21	1.60
0.10					128	0.84	2.05
	0.57				128	1.09	2.15
	0.10				128	0.96	1.49
		0.57			128	1.32	1.78
		0.10			128	0.73	1.87
			0.57		128	1.07	1.97
			0.10		128	0.98	1.68
				5	128	1.31	2.21
				2	128	0.74	1.44

Table 5.11: Analysis of Experimental Factors - Average Effect

Legend

c.v.(demand)	:	coefficient of variation of demand
c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(proctime)	:	coefficient of variation of processing times
c.v.(set)	:	coefficient of variation of setup times
spratio	:	setup/processing time ratio
a	:	capacity utilization = 0.95
b	:	capacity utilization = 0.85

	c.v.(demand)=0.57		c.v.(demand)=0.10	
	cap. util = 0.95	cap. util = 0.85	cap. util = 0.95	cap. util = 0.85
spratio = 5	1.74	1.58	0.87	2.84
spratio = 2	0.68	1.62	0.81	1.19

Table 5.12: Percentage Cost Analysis for Factors with Interaction

Legend

spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
cap. util.	:	capacity utilization

of variation of these factors while for low capacity utilization, the benefits are higher when the coefficient of variation is low. Finally the benefits of coordination are higher for high coefficient of variation of holding cost and setup time. These results demonstrate that benefits of coordination accrue in a wide variety of conditions, although the actual numerical value of these benefits may vary. As we have argued in our discussion of the experimental design, real life production systems in high-tech industries like computers, telecom equipment manufacturing, consumer electronics etc., have to encounter a dynamically changing environment characterized by compressed product life cycles, rapid product design changes and product obsolescence, and widely fluctuating demands. In such an environment, the same firm may observe variability in relation to the above factors over a period of time. For instance we show in Appendix 1 that if even a few products have significantly different processing times, the overall c.v. of processing time will be high. This can happen when a new product with an unstable product design is not fully integrated into the manufacturing system and hence has a considerably higher processing time on a critical resource, than the other products. As process learning occurs, the manufacturing of this product stabilizes and its processing time may approach that of the other products resulting in reduced c.v. of processing times. Similarly, for firms that have a well defined high sales season for a specific type of products (e.g., novelty items around Christmas), there may be differing variability of demand across products at different times of the year. Table 5.11 demonstrates that coordination benefits accrue for a wide range of values of the above factors, incorporating some of the variability that exists in real life. This should serve as a motivating factor for the adoption of the coordinated model by managers.

A comparison of the results in Tables 5.5 to 5.8 with the results in Tables 5.9 and 5.10 indicates that the cost improvement increases with increase in number of products. For Cases 1 to 4, the average cost improvement is 1.4% while for Cases 5 and 6, the average cost improvement is 2.6%. The reason for this increase is that with a larger set of products, there is an enlarged set of possibilities for transferring production of items from one period to another.

Table 5.13 depicts what we call the "maximum effect" of a treatment. Here we choose the maximum cost improvement over the independent replications as the representative effect of this treatment as opposed to the average of the replications in Table 5.11. Most of the trends described for the average cost improvements in Table 5.11 are also true for the maximum cost improvements. Managers can jointly look at both these performance measures, when considering the justification of the coordinated system.

Insert Table 5.13 here.

c.v. (demand)	c.v. (hcost)	c.v. (proctime)	c.v. (set)	spratio	Number of Problems	% Cost Improvement	
						a	b
0.57					32	3.69	4.36
0.10					32	2.54	4.81
	0.57				32	3.41	5.92
	0.10				32	2.83	3.24
		0.57			32	4.18	4.66
		0.10			32	2.06	4.51
			0.57		32	3.25	4.73
			0.10		32	2.98	4.44
				5	32	3.75	5.02
				2	32	2.48	4.14

Table 5.13: Analysis of Experimental Factors - Maximum Effect

Legend

c.v.(demand) : coefficient of variation of demand

c.v.(hcost) : coefficient of variation of holding cost

c.v.(proctime) : coefficient of variation of processing times

c.v.(set) : coefficient of variation of setup times

spratio : setup/processing time ratio

a : capacity utilization = 0.95

b : capacity utilization = 0.85

5.1.4 Implementation Aspects of Coordination

In this chapter, we have argued that the coordinated model is beneficial for the firm as it reduces its total cost of operations. We have demonstrated the actual cost benefits that accrue from coordination for a subset of problems which in terms of some of the aspects we are modelling, are representative of real life systems. We have also discussed the theoretical implications of our findings for managers. We now address the implications of these results from the point of view of implementation of the coordinated model in a firm. Broadly our objective here is to identify how different management levels of the firm can use these results in justifying the implementation of the coordinated model. We discuss the trade-offs that must be considered for such an analysis i.e., the potential benefits that accrue to each level of the firm versus the costs of making organizational changes and adopting the incentive structures that reinforce coordination. In keeping with our discussion of the issues in Chapter 1, we broadly discuss the implementation aspects at the strategic and the operational levels in the firm.

At the strategic level, the top management of the firm would like to assess the long term costs and benefits of adopting a coordinated approach to production planning. The typical questions that would have to be addressed at this level would be, how the coordinated approach would enhance the competitive position of the firm, what investment would be required for adoption of this approach, what would be the payback period of this investment, what kind of incentive structures would have to be put in place to motivate managers of the plant which will perform poorly according to the current performance measures etc. One of the most important prerequisites for the success of this approach is the commitment of the top management. This is a critical ingredient for the success of any coordination effort, because without this any attempt towards coordination will become very quickly mired in the set of current performance measures which are clearly oriented towards the uncoordinated approach as they put a high premium

1.	Annual Sales	$ 21.4b(17.1b)	
2.	Annual Profits	$ 1.3b(0.5b)	
3.	Average Value of Inventory	$ 3.0b(2.4b)	
4.	Cost of Inventory	@10% per annum	@15% per annum
		$ 300.0m(240.0m)	$ 450.0m(360.0m)
5.	Annual Benefits of Coordination		
	Average Effect (1.42%)	$ 4.3m(3.4m)	$ 6.4m(5.1m)
	Maximum Effect (3.85%)	$ 11.6m(9.2m)	$ 17.3m(13.9m)

Table 5.14: Benefits of Coordination: Case of Hewlett-Packard
Figures are for 4 quarters upto 1/31/94(1/31/93)

Sources:

1. "How H-P continues to grow and grow", by Alan Deutschman in *Fortune*, May 2, 1994.
2. "Strategic Supply Chain Management", by Corey Billington, *OR/MS Today*, April, 1994.

on the performance of the individual units, seen in isolation. We believe that large firms are ideal candidates for implementing coordination as they stand to make potentially large savings. With this as the frame of reference, we estimate the potential annual cash benefits of coordination for a large company. The example that we present is that of Hewlett-Packard (H-P) Company whose innovative performance has been commended in a recent review in *Fortune* magazine. The details of our analysis are provided in Table 5.14.

Using the cost improvements obtained from our average and maximum analysis (across independent replications) in Tables 5.11 and 5.13, the potential benefits of the coordinated model for H-P are shown in Table 5.14. Some computations will demonstrate that these cost improvements would have added 2% - 3% to H-P's net profits in the previous two years. We *emphasize* that this estimate is on the conservative side because as we discussed earlier, the cost benefits of coordination increase with increase in number of products. Therefore, in a real life scenario, where the number of products is much larger, the cost benefits would be larger too. We have also demonstrated that for certain instances, the cost improvements exceeded 10%. If these instances were to apply to H-P, the impact on its net profits would be in the range of 4% to 8%. In addition, it must be noted that the cost improvements alone do not reveal the true potential of coordination. The positive spin-off effects of an effective coordination strategy would be many. For instance, the savings could be channelled into a focused setup reduction program which as we earlier discussed will reinforce the benefits of coordination. In addition, as we discussed in Chapter 1, multi-plant coordination is a subset of the overall supply chain coordination domain. Effective multi-plant coordination would likely inspire supply chain coordination and hence strengthen the competitive position of the firm. An important concern for top managers is to put in place an incentive structure which reflects the objective of coordination. There would therefore be an important need to incorporate the concerns of the entire firm in performance evaluation measures as opposed to those of individual entities.

The cost implications of coordination would be quite minimal for large companies, e.g., H-P, because they already have sophisticated information systems linking manufacturing and distribution sites. In any case, any investment that is made in improving the information system will be a one time cost while the benefits would accrue perennially. The more important concern would be the reaction of managers to what they may see as limiting their autonomy. One alternative would be to evaluate the true benefits of the coordinated versus the uncoordinated approach and equitably allocate these between different units. If the overall benefits are positive and the reallocation is equitable, all units will have incentive to cooperate as this would improve each unit's performance vis-a-vis the uncoordinated approach. The same would apply to a buyer-supplier relationship and a relationship based on coordination could lead to a cooperative partnership. As discussed by Womack et al., (1990), such a relationship based on a rational framework rather than the relative bargaining power of the different units could be an important source of future competitiveness of firms.

Many of the organizational changes for implementing the coordinated model would be done at the higher levels of management. Nevertheless, these changes would have to percolate to the plant level workers who are responsible for the day-to-day production activities. Individual plant managements would have to be more sensitive to the concerns at the other plants. One important aspect that would go a long way in increasing such sensitivity would be to encourage exposure across plants both at the management and the worker levels. Benefits of coordination also need to be communicated down to the lowest levels in order to improve the awareness about the importance of coordination.

In summary therefore, the critical strategic inputs for the success of coordination are commitment of the top managers, the instituting of a reinforcing incentive structure, and an effective mechanism for channelling the savings from coordination. At the operational level, while the day to day responsibilities would not change, cross training and sensitivity towards the benefits of coordination would be important determinants of the success of the coordinated system.

5.2 Convergence Results for Lagrangian Relaxation and Lagrangian Decomposition

Till now, we have largely focused on the practical benefits and the implementation aspects of the coordinated model. In this section we turn to the theoretical concern of comparing the results obtained from the Lagrangian relaxation and Lagrangian decomposition methodologies. The measure of performance we use for this comparison is the **gap** between the best lower bound and the best feasible solution obtained in each case expressed as a percentage of the best lower bound (hereafter referred to as the gap). The gap signifies the maximum deviation of the feasible solution from optimality. The gap was measured in each case when a pre-specified termination condition was achieved. The termination condition was to stop when either 500 iterations were completed or when the gap was reduced to below 0.5%. The iteration limit was determined largely by the time restrictions. We present results for the six cases discussed on page 117. Tables 5.15 to 5.20 summarize the results for the gaps obtained with Lagrangian relaxation while the results for Lagrangian decomposition are in Tables 5.21 to 5.24. A comparative analysis of the approaches is presented in Table 5.25 for the 128 cases where Lagrangian relaxation and Lagrangian decomposition were used for the same set of problems i.e., 2 independent replications of 64 treatments referred to in our experimental design.

Insert Tables 5.15 to 5.25 here.

Lagrangian decomposition outperforms Lagrangian relaxation in 84% (107 out of 128) instances, while in the remaining 16% cases, Lagrangian relaxation yields a smaller gap. In the 107 instances where Lagrangian decomposition gives a lower gap, it outperforms Lagrangian relaxation by 3.9% . In contrast, when Lagrangian relaxation is better, the average reduction in gap compared to Lagrangian decomposition is 1.1%. Finally, as outlined in Table 5.25, in terms of the average gap, the maximum gap and the standard deviation of the gap, Lagrangian decomposition is distinctly superior to Lagrangian relaxation.

Another important concern from the algorithmic point of view is the probable behaviour of the gap with increased problem size. For the two problem sizes considered in this dissertation (3/4 and 10/20 modules/chips respectively), we observed that the percentage gaps for the same set of input data sets were noticeably similar. This can be verified by comparing the results in Tables 5.15 and 5.17, with those in Tables 5.19 and 5.20. Tables 5.15 and 5.17 summarize the percentage gaps for the smaller problems under high c.v. of demand using Lagrangian relaxation, while Tables 5.19 and 5.20 are the corresponding cases for the larger problems. The average percentage gaps turn out to be 8.7chips. With an approximately five fold increase in problem size, the percentage gap has remained remarkably stable. Based on our experience with the two sizes considered, it is reasonable to assume that the percentage gaps do not increase with increase in problem size. Another observation is that while the maximum gap (for the cases considered above) turned out to be 27.4only 15.5

From a practical point of view, an alternate comparison of the two approaches could be based on the values of the heuristic solutions generated. We have noted that in a vast majority of cases, Lagrangian decomposition yields heuristic solutions that are at least as good as those obtained from Lagrangian relaxation. Lagrangian relaxation yields better feasible solutions in only a handful of instances. Hence the conjecture that better lower bounds also yield improved feasible solutions. Our results provide an empirical proof for the above assertion.

Lagrangian decomposition therefore gives a superior performance in terms of both the gap and the value of the feasible solutions. Nevertheless, there is some concern with the maximum gap which in some problem instances turns out to be close to 20%. In the next subsection, we address how this gap can be reduced further, using our existing results from subgradient optimization.

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	4.4	3.8	6.2	5.9
		7.9	10.3	12.1	12.0
(proc.		5.7	4.3	7.8	5.3
time)		1.9	4.2	1.2	4.4
	c.v.(set)=0.10	4.1	4.3	6.6	7.4
=		8.8	10.4	11.0	12.8
		5.2	4.4	5.4	5.0
0.57		6.2	5.1	2.5	5.5
c.v.	c.v.(set)=0.57	3.3	2.3	2.2	2.2
		0.6	1.7	2.1	3.0
(proc.		2.8	4.4	0.7	3.7
time)		1.2	3.2	3.7	3.0
	c.v.(set)=0.10	2.1	2.6	1.7	2.2
=		1.6	1.7	1.9	3.0
		3.3	4.2	2.1	5.7
0.10		1.5	4.6	2.4	4.1

Table 5.15: Percentage Gap for Case 1 using Lagrangian Relaxation

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	2.9	3.0	3.4	4.9
		14.2	14.0	15.8	13.5
(proc.		3.8	3.9	2.5	4.9
time)		4.6	3.4	4.3	3.7
	c.v.(set)=0.10	1.7	3.2	3.4	4.8
=		15.1	12.9	20.2	16.3
		6.1	4.9	4.2	5.4
0.57		6.9	4.7	6.5	4.7
c.v.	c.v.(set)=0.57	6.0	4.4	5.0	4.0
		3.4	3.8	3.9	4.4
(proc.		2.6	3.4	2.4	3.9
time)		2.9	4.0	2.4	3.5
	c.v.(set)=0.10	3.9	3.9	3.5	3.6
=		4.8	5.2	5.8	5.0
		6.1	4.0	5.8	4.1
0.10		7.9	5.4	5.1	4.1

Table 5.16: Percentage Gap for Case 2 using Lagrangian Relaxation

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.10)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	27.4	23.2	18.4	17.6
		13.3	13.5	18.3	16.3
(proc.		7.7	7.6	9.4	9.5
time)		19.6	17.4	11.2	14.2
	c.v.(set)=0.10	26.1	22.7	18.7	17.1
=		13.9	14.0	18.8	16.1
		7.8	7.5	12.0	10.3
0.57		21.7	18.8	12.4	13.9
c.v.	c.v.(set)=0.57	19.2	16.8	8.6	7.7
		9.0	8.9	12.4	11.2
(proc.		5.3	5.2	5.1	6.2
time)		14.3	14.0	6.7	7.5
	c.v.(set)=0.10	18.4	16.4	7.8	7.3
=		8.7	8.7	13.3	11.3
		6.0	4.5	7.2	6.5
0.10		17.4	14.5	9.6	8.2

Table 5.17: Percentage Gap for Case 3 using Lagrangian Relaxation

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.85

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v.	c.v.(set)=0.57	33.0	26.0	20.4	16.8
		16.6	16.4	21.2	19.6
(proc.		5.1	6.2	8.8	8.8
time)		17.5	17.9	20.4	11.0
	c.v.(set)=0.10	26.7	24.1	19.8	16.1
=		13.7	17.5	22.9	20.0
		7.8	7.3	9.9	10.0
0.57		20.4	18.2	13.6	12.9
c.v.	c.v.(set)=0.57	28.7	23.2	12.0	9.3
		8.8	11.0	16.1	13.9
(proc.		4.9	3.7	8.9	6.0
time)		17.2	14.0	7.4	7.7
	c.v.(set)=0.10	25.8	22.1	9.2	9.2
=		10.1	11.0	16.2	13.7
		4.2	3.5	5.9	5.3
0.10		16.9	16.1	11.6	7.3

Table 5.18: Percentage Gap for Case 4 using Lagrangian Relaxation

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.10)
capacity utilization	:	0.85

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v. (proc. time) = 0.57	c.v.(set)=0.57	13.2	12.1	12.0	3.7
	c.v.(set)=0.10	12.0	10.9	8.4	8.8
c.v. (proc. time) = 0.10	c.v.(set)=0.57	15.5	9.8	3.7	5.0
	c.v.(set)=0.10	9.8	10.9	3.2	5.2

Table 5.19: Percentage Gap for Case 5 using Lagrangian Relaxation

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v. (proc. time) = 0.57	c.v.(set)=0.57	12.4	11.3	7.0	7.2
	c.v.(set)=0.10	11.2	10.6	7.2	8.4
c.v. (proc. time) = 0.10	c.v.(set)=0.57	8.9	10.5	5.9	5.9
	c.v.(set)=0.10	8.9	10.8	5.6	5.2

Table 5.20: Percentage Gap for Case 6 using Lagrangian Relaxation

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.85

<table>
<tr><td rowspan="2"></td><td rowspan="2"></td><td colspan="2">c.v.(hcost)=0.57</td><td colspan="2">c.v.(hcost)=0.10</td></tr>
<tr><td>spratio=5</td><td>spratio=2</td><td>spratio=5</td><td>spratio=2</td></tr>
<tr><td rowspan="2">c.v. (proc. time) = 0.57</td><td>c.v.(set)=0.57</td><td>1.0
7.3</td><td>2.8
8.0</td><td>0.5
10.6</td><td>0.5
8.7</td></tr>
<tr><td>c.v.(set)=0.10</td><td>0.5
8.0</td><td>0.6
7.7</td><td>0.8
8.9</td><td>2.1
8.5</td></tr>
<tr><td rowspan="2">c.v. (proc. time) = 0.10</td><td>c.v.(set)=0.57</td><td>2.3
1.4</td><td>1.0
2.5</td><td>1.2
3.0</td><td>0.6
1.4</td></tr>
<tr><td>c.v.(set)=0.10</td><td>0.8
2.5</td><td>0.6
3.3</td><td>0.7
1.7</td><td>0.6
1.3</td></tr>
</table>

Table 5.21: Percentage Gap for Case 1 using Lagrangian Decomposition

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v. (proc. time) = 0.57	c.v.(set)=0.57	2.8 8.4	2.8 13.7	2.5 7.1	4.0 12.5
	c.v.(set)=0.10	3.0 8.8	3.7 12.7	3.4 11.0	0.5 7.9
c.v. (proc. time) = 0.10	c.v.(set)=0.57	5.3 2.6	5.8 5.2	3.2 3.1	4.7 4.0
	c.v.(set)=0.10	3.1 2.6	3.8 2.1	1.9 1.8	4.2 4.0

Table 5.22: Percentage Gap for Case 2 using Lagrangian Decomposition

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.10)
capacity utilization	:	0.95

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v. (proc. time) = 0.57	c.v.(set)=0.57	14.3 13.8	20.3 12.7	12.4 12.1	13.1 15.3
	c.v.(set)=0.10	19.4 12.1	19.2 12.8	13.6 12.0	12.2 14.4
c.v. (proc. time) = 0.10	c.v.(set)=0.57	4.4 9.3	0.6 6.7	2.5 13.8	1.7 8.7
	c.v.(set)=0.10	5.4 9.5	1.1 8.5	4.6 11.4	2.5 8.4

Table 5.23: Percentage Gap for Case 3 using Lagrangian Decomposition

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.57)
capacity utilization	:	0.85

		c.v.(hcost)=0.57		c.v.(hcost)=0.10	
		spratio=5	spratio=2	spratio=5	spratio=2
c.v. (proc. time) = 0.57	c.v.(set)=0.57	18.8 12.6	19.2 17.5	19.1 18.6	18.4 15.9
	c.v.(set)=0.10	18.6 8.8	19.1 18.4	17.7 17.7	17.9 15.6
c.v. (proc. time) = 0.10	c.v.(set)=0.57	14.7 10.4	8.9 9.4	8.8 14.8	4.6 13.4
	c.v.(set)=0.10	15.1 11.5	7.5 10.3	6.4 14.5	7.7 10.5

Table 5.24: Percentage Gap for Case 4 using Lagrangian Decomposition

Legend

c.v.(hcost)	:	coefficient of variation of holding cost
c.v.(set)	:	coefficient of variation of setup times
c.v.(proctime)	:	coefficient of variation of processing times
spratio	:	setup/processing time ratio
c.v.(demand)	:	coefficient of variation of demand (0.10)
capacity utilization	:	0.85

	Lagrangian Decomposition	Lagrangian Relaxation
Average Gap	8.1%	11.2%
Max-Min Range	0.5% to 20.3%	0.6% to 33.0%
Standard Deviation of Gap	5.9%	7.5%
Number of times gap lower	107(= 84%)	21(= 16%)
Average Gap Reduction when Approach superior	3.9%	1.1%

Table 5.25: Comparative Analysis of Gaps

5.2.1 Discussion of duality gap

(in the rest of this chapter, $v(a)$ represents the optimal value of problem a.)

The quality of results from Lagrangian relaxation has traditionally been judged in terms of the gap between the best lower bound obtained by solving the Lagrangian dual and the upper bound obtained from the feasible solutions that are generated at each iteration. If this gap is reasonably small, we have found a provably near optimal feasible solution. However, as observed by Trigeiro et al., (1989), in complex integer programs, the optimal value of the Lagrangian dual may be substantially lower than the optimal solution to the primal problem. If the gap between the upper and the lower bounds is large, it is difficult to ascertain whether it is associated with a large duality gap or a large deviation of the feasible solution from optimal or an incomplete computation of the optimal lower bound. This ambiguity is unfortunate and has been recognized by several researchers (e.g., Geoffrion, 1974, Trigeiro et al., 1989).

We extend the above analysis with the objective of clarifying the issue of duality gap. For this purpose, we use an adapted version of a linear programming formulation of a production planning problem which is due to Manne (1958). We first re-state Problem C in the generic form, C_{primal}, earlier presented in Chapter 4 (see page 69).

Problem C_{primal}

Minimize

$$C_{primal} = cx$$

subject to

$$\begin{aligned} Ax &= b & (5.1) \\ Dx &= e & (5.2) \\ x &\geq 0 \quad \text{and integral} & (5.3) \end{aligned}$$

In the above formulation, cx represents the objective function (inventory holding and overtime cost terms). Constraints (5.1) represent the complicating constraints (i.e., capacity constraints and link constraints). Constraints (5.2) represent the special structure constraints (i.e., inventory balance, setup enforcement and overtime enforcement constraints). An adapted version of Manne's (1958) linear programming formulation, C_{dual}, which has also been reported in Fisher (1981) is reproduced below (the relationship between C_{primal} and C_{dual} is explained later).

Problem C_{dual}

Minimize

$$C_d = \sum_{n=1}^{N} \lambda_n . cx^n$$

subject to

$$\begin{aligned}\sum_{n=1}^{N} \lambda_n . Ax^n &\geq b \\ \sum_{n=1}^{N} \lambda_n &= 1 \\ \lambda_n &\geq 0 \\ x &\in X\{X : Dx \geq e, x \geq 0 \text{ and integral}\}\end{aligned}$$

The superscript n (n going from 1 to N) defined over the set N_{feas} indexes all feasible vectors of decision variables for the decomposed problems (represented by the special structure constraints (5.2) in C_{primal} above). The set X can be represented as $X = \{X^n, n = 1, \ldots, N\}$ (see Fisher (1981)). This set is known to be finite because of the integrality and boundedness of the variables. However the cardinality of X is extremely large. The λ's are optimal non-negative weights allocated to each individual vector of decision variables in X. Additionally, if the λ's are constrained to be binary, C_{dual} will be recognized as being equivalent to C_{primal}. However, constraining λ's to be binary is not very meaningful as it makes C_{dual} as difficult to solve as C_{primal}. For this reason, C_{dual} is formulated as a linear program with continuous λ's. To emphasize the physical significance of this problem, we re-state its interpretation.

Minimize Weighted Primal Costs

subject to

(1) Sum of weighted violations for all relaxed complicating constraints is eliminated

(2) Sum of non-negative weights equals 1

$v(C_{dual})$ is a provable lower bound on $v(C)$. However given the large cardinality of X, complete enumeration is ruled out. We therefore follow a modified approach towards solving C_{dual}. We call this procedure, the **Bound Improvement Procedure (BIP)**. A vector of decision variables relating to special structure constraints is included in a set called X' as it is generated in the course of implementing subgradient optimization (refer to methodology outlined in Chapter 4). After every $\mathcal{F}$ iterations of subgradient optimization, Problem C_{dual} is solved in a modified form, C'_{dual} (defined over X' instead of over X) which has substantially fewer columns as compared to C_{dual}. $\mathcal{F}$ can be any convenient number of iterations. From the structure of the problem, it is evident that each successive feasible solution to C'_{dual} decreases monotonically since addition of new variables cannot make the result any worse. C'_{dual} differs from C_{dual} in an important respect. It lacks some of the columns(variables) which would be present in C_{dual} which results in $v(C'_{dual})$ being higher than $v(C_{dual})$. This prevents C'_{dual} from being a provable lower bound on $v(C_{eq})$. However the convergence of $v(C'_{dual})$ as a function of the cardinality of X' can be empirically investigated. Initially, X' is not representative of X as C'_{dual} lacks several columns (variables) of C_{dual} and hence it is probable that $v(C'_{dual})$ exceeds $v(C)$. However, as vectors are added into X' (based on additional iterations of subgradient optimization), it becomes increasingly representative of X and $v(C'_{dual})$ will decrease till it finally drops below $v(C)$. At this point, $v(C'_{dual})$ becomes a lower bound on $v(C)$. Evidently, it is difficult to know <u>when</u> $v(C'_{dual})$ becomes a lower bound, since we do not know $v(C)$. We note however the following relationship between $v(C'_{dual})$ and $v(C_{dual})$.

$$v(C'_{dual}) \quad = \quad v(C_{dual}) \quad = \quad v(Ldual_{LRelax})$$

$$\text{as } X' \rightarrow X$$

(See Fisher,1981)

Recall that $Ldual_{LRelax}$ is the Lagrangian dual problem for the case of Lagrangian relaxation of capacity and coupling constraints. We empirically investigated the behaviour of Problem C'_{dual} as a function of increasing iterations of subgradient optimization. The above analysis was implemented on several problems and the results are presented in Figures 5.1 and 5.2. As the number of iterations increases, X' approaches X. After every 100 iterations (i.e., $\mathcal{F} = 100$), the results (i.e., production quantity X_{it} etc.,) relating to the relaxed problems $B_{Lrelax1}$, $B_{Lrelax1}$, $A_{Lrelax1}$ and $A_{Lrelax2}$, (see page 76 onwards) were added to X' and C'_{dual} was optimized as explained before. Addition of the results of successive iterations of subgradient optimization in this manner assures a monotonic decrease in $v(C'_{dual})$. $v(C'_{dual}$ represents an upper bound on the optimal value of the Lagrangian dual and hence we can ascertain the maximum deviation of the lower bound (obtained from subgradient optimization) from the optimal value of the Lagrangian dual. Without the Bound Improvement Procedure, no such indication about the quality of the lower bound was available (except the percentage gap itself). One important use of $C'dual$ is therefore to determine the proper termination condition for subgradient optimization. C'_{dual} may be used as an indicator as to when the solution of $Ldual_{LRelax}$ may be stopped, i.e., when $v(C'_{dual})$ - $v(Ldual_{LRelax})$ is sufficiently small. For example, subgradient optimization to solve $Ldual_{LRelax}$ may be continued till the gap between $v(C'_{dual})$ and $v(Ldual_{LRelax})$ is less than 5%.

Another use of C'_{dual} is to utilize its solution to construct a feasible solution for C. For this purpose we used the following heuristic procedure. Whenever a feasible solution to C'_{dual} was obtained, the production variable for each product i in every period t was determined by taking a weighted average of the optimal λ's, i.e.,

$$\mathcal{X}_{it} = \sum_{n \in \mathcal{X}'} \lambda_n X_{it}{}^n$$

where $X_{it}{}^n$ represents the production quantity for product i in period t in the n^{th} iteration of subgradient optimization. $\mathcal{X}_{it}$ is therefore a synthesis of all the iterations of subgradient optimization. $\mathcal{X}_{it}$ was rounded downwards to obtain integer production quantities. Setup and overtime enforcement variables were obtained according to the following rules:

$$
\begin{aligned}
\mathcal{Y}_{it} &= 1 \quad \text{if } \mathcal{X}_{it} \neq 0 \\
& \quad 0 \quad \text{otherwise} \\
\mathcal{Z}_{1t} &= 1 \quad \text{if } \textstyle\sum_{i=1}^{m} (\mathcal{X}_{it}.b_i + \mathcal{Y}_{it}.s_i) \geq Cap_{1t} \\
& \quad 0 \quad \text{otherwise} \\
\mathcal{Z}_{2t} &= 1 \quad \text{if } \textstyle\sum_{i=m+1}^{m+n} (\mathcal{X}_{it}.b_i + \mathcal{Y}_{it}.s_i) \geq Cap_{2t} \\
& \quad 0 \quad \text{otherwise}
\end{aligned}
$$

The determination of overtime quantities in each period at either plant is a simple computation based on the above quantities. We call this the composite solution represented by $\mathcal{C}$. $\mathcal{C}$ maybe both demand infeasible and capacity infeasible for the original problem, i.e., C_{primal}. However the infeasibility may not be very large and a feasible solution maybe heuristically obtained. We adopt the following two step procedure.

(a) Demand Infeasibility: A single pass procedure is used to determine cumulative demand feasibility for each period, i.e., to determine whether cumulative production for each product is at least as large as the cumulative demand in every period. If the cumulative demand exceeds cumulative production, $\mathcal{X}_{it}$ is increased appropriately to ensure cumulative demand feasibility.

(a) Capacity Infeasibility: A modified form of the feasibility heuristic described on page 93 is used. The procedure used here differs from the earlier heuristic in that the cost of shifting items is limited to considering the inventory holding and overtime costs only. This heuristic goes through the four passes described for the feasibility heuristic on page 95.

In this section, we have outlined two different procedures. The first procedure relates to solving Problem C'_{dual} which is a reasonably sized linear programming problem and is the dual of the Lagrangian dual, $Ldual_{LRelax}$, for the Lagrangian relaxation procedure. The solution to this problem yields a measure of the convergence of the lower bound obtained by optimizing the Lagrangian dual using subgradient optimization. The above technique has been widely reported in literature (see Fisher, 1981) and the effectiveness of the dual problem in determining the quality of the lower bound has been recognized. However we have found no evidence in literature of the second procedure that we have proposed i.e., utilizing the solution of Problem C'_{dual} to construct an improved feasible solution. This attempt therefore signifies a fundamental shift and seeks to establish the importance of the dual problem C'_{dual} from the point of view of improving the upper bound. The practical significance of this procedure is considerable because any improvement in the upper bound leads to an actual reduction in costs.

With this objective, the Bound Improvement Procedure (BIP) was applied to 25 problems selected from Tables 5.17 and 5.18. The choice of these cases was guided by the relatively large percentage gaps that remained unresolved after implementing 500 iterations of subgradient optimization. The procedure was implemented for the Lagrangian relaxation case (since the gaps were higher) and the number of problems selected represented approximately 10% of the total problem set on which Lagrangian relaxation was initially implemented. We first diagrammatically present the details of the implementation for two different cases to contrast the different conditions relating to the lower bound convergence, under which the cost of the feasible solution could be improved. Next we summarize the improved cost benefits for the 25 problems on which it was implemented.

The detailed results for two examples are presented in Figures 5.1 and 5.2 and incorporate the results of 4000 iterations. The important measures presented in these figures are defined as follows:

a = (Heuristic 1 *minus* Lagrangian Dual)

b = (Bound Imp. Proc. *minus* Lagrangian Dual)

c = (Heuristic 1 *minus* Bound Imp. Proc.)

d = (Dual of Lag. Dual *minus* Lagrangian Dual) *a* represents the percentage gap between the upper bound obtained from Heuristic 1 and the best lower bound obtained from the Lagrangian dual (gap is defined exactly as on page 135 except that it is measured after 4000 iterations). *b* is the percentage gap after the BIP has been applied to obtain a better upper bound. *c* is the improvement in the cost of the coordinated solution after implementing BIP over Heuristic 1 (expressed as a percentage of the cost of Heuristic 1). This represents the actual cost savings as a result of implementing BIP. Finally, *d* represents the convergence of the lower bound obtained by using subgradient optimization for maximizing the Lagrangian dual i.e., in terms of the maximum deviation from the optimal value.

In the first example (Diagram 5.1), we observe that implementing BIP results in a reduction of the percentage gap between the best known upper and lower bounds. The actual cost improvement in the value of the feasible solution is approximately 7%, while the lower bound is shown to converge to 5.4% within its optimal value. For this problem, we also obtained the true optimal solution by direct application of branch and bound, and the best feasible solution deviated by 2.6% from the true optimal. Hence for this particular example, we can say that out of the final unresolved gap of approximately 20.1%, 2.6% is attributable to the heuristic procedures that we have used, while a maximum of 5.4% is due to the incomplete optimization of the Lagrangian dual. The remaining approximately 12% of the gap represents the duality gap inherent in the problem about which nothing can be done (the percentages may not match exactly as they are defined over different bases).

In the second example (Diagram 5.2), while *a*, *b* and *c* behave as explained above for the first example, the behaviour of *d* is different. Here, the dual of the Lagrangian dual converges to an evidently non-optimal value which is even above the value of the upper bound. In such a situation, *d* is virtually meaningless. Nevertheless, despite the poor behaviour of the dual, we note that a cost improvement of 3.1% was obtained, as can be observed from c. Once again the true optimal solution was obtained using branch and bound and this indicated a deviation of 3.2% between the best feasible solution and the true optimal. For this example, no assessment can be made about the allocation of the unresolved gap to the three sources i.e., upper bounding and lower bounding procedures and the duality gap, because of the poor value of the dual.

Table 5.26 represents the summary of results for the 25 problems that were considered. x measures the cost improvement over 500 iterations (these are taken from Tables 5.7 and 5.8). y represents the benefits of coordination after increasing the number of iterations to 4000. z represents the coordination benefits after the BIP has been implemented (over 4000 iterations). It can be observed that while a substantial increase in the number of iterations yield no benefit on the average, the implementation of BIP boosts the cost benefits of the coordinated model by over 40% (3.0% versus 2.10%).

The Bound Improvement Procedure therefore provides an improved feasible solution as compared to the feasibility heuristic described in Chapter 4 (page 93). We have not used this procedure for the entire set of problems discussed in Section 5.1.3 because solving C'_{dual} necessitates solving a large LP problem and hence increases the computational effort considerably. However we have demonstrated that with better computational resources, improved feasible solutions can be obtained using the Bound Improvement Procedure. Our results relating to cost improvements in Section 5.1.3 are therefore conservative. This coupled with the fact that cost improvement increases with increased number of products should reassure firms that the cost benefits of coordination will in fact be larger than those reported in Section 5.1.3.

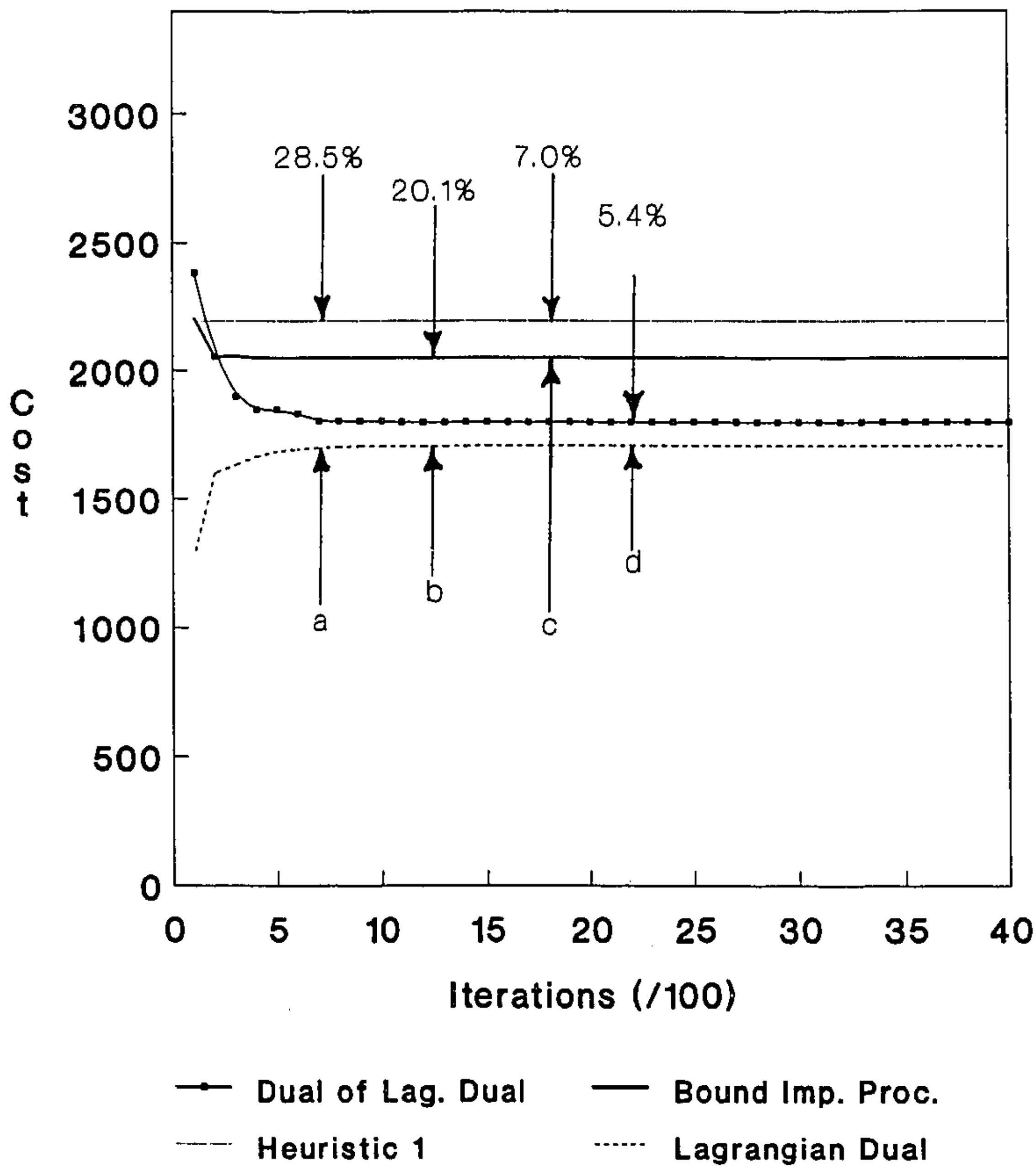

Figure 5.1: Bound Improvement Procedure: Example 1

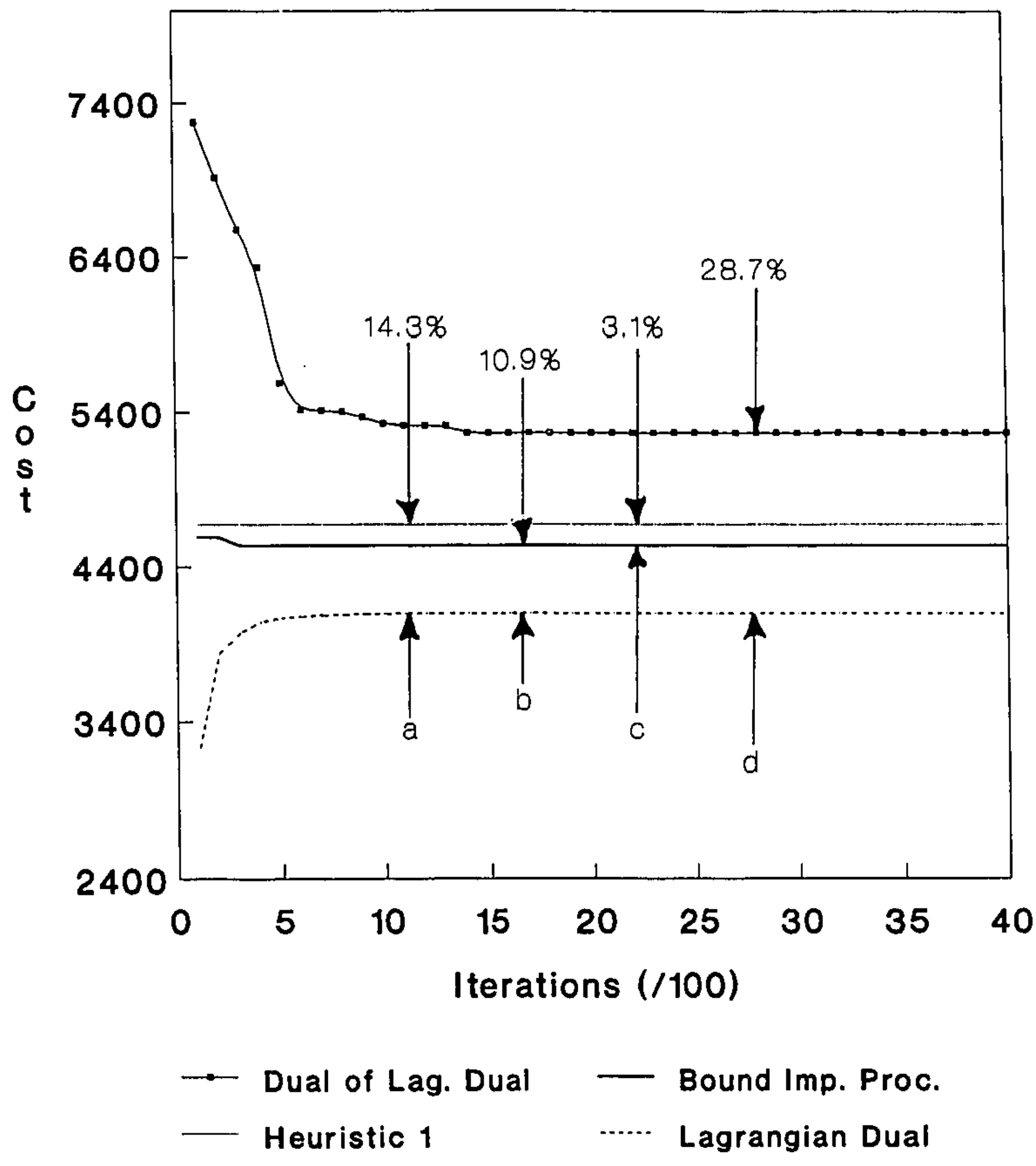

Figure 5.2: Bound Improvement Procedure: Example 2

Problem Number	500 Iterations	4000 Iterations	4000 Iterations with BIP
1	3.7	4.1	4.1
2	0.7	0.7	1.4
3	6.4	6.4	6.4
4	1.2	1.2	1.2
5	7.2	7.2	7.3
6	7.2	7.2	7.2
7	0.2	0.2	0.2
8	0.0	0.0	1.4
9	2.9	3.5	5.5
10	2.1	2.1	3.1
11	0.0	0.0	0.0
12	0.0	0.0	0.0
13	2.4	2.4	4.6
14	1.1	1.1	1.1
15	0.0	0.0	0.0
16	0.0	0.0	0.0
17	0.0	0.0	1.3
18	0.9	0.9	2.3
19	0.5	0.8	0.8
20	0.0	0.0	3.0
21	2.1	2.1	6.7
22	2.3	2.3	4.9
23	5.1	5.1	5.7
24	4.9	4.9	5.7
25	0.6	0.6	1.3
Average	2.1	2.1	3.0

Table 5.26: Percentage Cost Improvements with Bound Improvement Procedure

5.3 Summary

The analysis outlined in the first section of this chapter indicates that the coordinated approach gives better cost performance as compared to the uncoordinated approach. A general trend observed was a reduction in costs at Plant A and an increase in the costs at Plant B. The cost decrease at Plant A was larger than the cost increase at Plant B and hence the superior performance of the coordinated model. The analysis of inventory and setup behaviour at the two plants indicates that the coordinated model recommends a larger number of setups as compared to the uncoordinated model. A range of cost improvements were observed for different levels of experimental factors. Two factors i.e., capacity utilization at the two plants and the setup to processing time ratio were observed to be significant.

The second section of this chapter focused on the relative performance of Lagrangian relaxation versus Lagrangian decomposition. Our results demonstrate the superiority of Lagrangian decomposition in terms of both lower gaps between the upper and lower bounds and better values of feasible solutions. A LP based Bound Improvement Procedure was presented which serves as a meaningful indicator for terminating the subgradient optimization procedure and also yields superior feasible solutions.

Chapter 6

Conclusions and Future Work

6.1 Conclusions and Contributions

At the outset we observe that coordination as a philosophy has been widely advocated in the operations management literature. As discussed in Chapter 1, the main reason for the widespread support for this approach is that it helps to integrate the sub-units of the organization by providing an optimal direction for the firm as a whole on several critical issues (see Table 1.1 on page 9). Unfortunately, there are very few success stories to study or emulate on the coordination front and hence there is no empirical evidence relating to actual implementation in the industry. This study represents a step in that direction as it successfully illustrates the use of analytical tools to bring out the cost benefits of coordination.

A multi-plant coordination model which captures the interaction between plants and explicitly incorporates setup times was proposed for a two-plant firm. Using a real life production system at IBM as the motivating example, the objective of this research was to compare the relative performance of the coordinated model and the uncoordinated model which ignores the interaction between the plants. The model bridges an important gap in the literature, for this problem has not been addressed until now. Our model is realistic in terms of the cost structure as it considers the important aspects of inventory holding costs and overtime costs. Since the problem was computationally intractable in its original form, several equivalent formulations were proposed and competing relaxations were discussed to solve the problem. Efficient solution techniques based on Lagrangian relaxation and Lagrangian decomposition were implemented for solving the model. The experimental design reflected the concern for realism and focused on studying the effect of several key parameters on coordination. The parameters that were investigated were measures of demand variability, capacity utilization, setup/processing time ratio, setup time variability, processing time variability and holding cost variability. Variability in each case was represented by the coefficient of variation of the parameter and hence the data used in the experimentation was representative of an underlying distribution. Independent replications were conducted for each set of parameter combination.

Our experimentation led to two general set of findings, one relating to cost improvements and the other to the performance of the alternative methodologies, Lagrangian relaxation and Lagrangian decomposition.

In terms of the cost improvements, the coordinated model outperformed the uncoordinated approach in 62% of the problem instances, while for the other problems the two approaches performed equally well. A general trend observed in cases where the coordinated approach was better was that the costs decreased for the upstream plant (A) while they increased for the downstream plant B. The decrease at A were larger than the increase at B resulting in cost improvements. There was also a tendency towards increased number of setups in the solution of the coordinated model. A wide range of cost improvements were observed, with the maximum improvement exceeding 11%. Using the average of the independent replications, the cost benefits were 1.42%. The improvement obtained by averaging the maximum cost benefit for the independent replications of the same parameters was 3.85%. We have estimated that for large companies like Hewlett-Packard this could mean cost savings in excess of $ 17 million which is approximately 2% to 3% of its net profits in the previous two years. We note that these figures are conservative due to two distinct reasons. First, we have demonstrated that cost benefits of coordination improve as the number of products increase. Second, additional improvements could accrue by implementing the Bound Improvement Procedure. The actual benefits of coordination in a real life scenario are therefore likely to be higher than those reported above. An important contribution of this research is therefore to establish the use of coordination as a strategy towards improving the firm's competitive position.

Statistical analysis established that the cost improvements obtained from different parameter combinations (treatments) are not significantly different. Benefits of coordination will therefore accrue over a wide variety of values of parameters. Two factors out of the above i.e., capacity utilization and setup/processing time ratio were found to be significant at $\alpha = 0.02$. The importance of coordination is greater at medium ranges of capacity utilization since there is a greater flexibility at the two plants to accommodate the other. Coordination also assumes importance when setup/processing time ratio is high.

The above results relating to cost benefits of coordination have important implications from the implementation point of view. Organizational changes need to be initiated and new structures that support coordination need to be adopted at the highest levels in the organization, if coordinated operations are to be successfully implemented. Appropriate incentive schemes, cross training of workers and managers across different plants and exchange of information between plants were identified as the important issues that need to addressed in this regard.

In terms of the alternative methodologies, Lagrangian decomposition provided a superior performance, on the measure of the gap between the best known feasible solution and the best known lower bound, for 84% instances. On the criterion of the average gap, maximum gap and the standard deviation, Lagrangian decomposition was consistently better than Lagrangian relaxation. The Lagrangian decomposition was also superior to Lagrangian relaxation from the point of view of generating better feasible solutions. While the superiority of Lagrangian decomposition has been established theoretically, our contribution is in demonstrating its implementation for the multi-plant coordination problem. The underlying knapsack sub-problems which were exposed as a result of implementing the Lagrangian decomposition represent a unique method of solving this problem which has not been reported before.

The LP based Bound Improvement Procedure strengthens our analyses from two important viewpoints. First, it provides an upper bound on the optimal value of the Lagrangian dual, thereby establishing its quality. This, as we have earlier argued, could be used as a meaningful termination condition for subgradient optimization. The second application of the Bound Improvement Procedure that we have demonstrated in this dissertation is the generation of improved feasible solutions. This to our knowledge has not been reported in the literature and represents another important contribution of this research.

In conclusion, this dissertation establishes the importance of coordination in improving the performance of the firm and in strengthening its competitive position. In addition, the methodological constructs proposed.

6.2 Future Directions

In this section we bring together the insights accumulated from the modelling process and the computational experience to identify future research directions that integrate the overall research framework for multi-plant coordination, outlined in Chapter 2. In the current research, we have considered a two-plant model with a single workcenter at each plant. A significant direction for future research is to consider a model with a more realistic production structure at each plant and also the case with more than two plants. These two situations are depicted in Figures 6.1 and 6.2. We present the outline of a generic model which may be used to represent both the above cases. We will later examine how the solution methodology proposed in this dissertation can be used for solving the above model. Problem $C_{Ldfuture}$ represents the case with three plants depicted in Figure 6.2, but we will demonstrate that this model is general enough to incorporate the situation with n plants with N different workcenters at each plant (Figure 6.1). Ld in the subscript of the problem name signifies that this is a formulation for using Lagrangian decomposition. We consider three plants A, B and C which successively add value to products. The additional notation used is defined below.

Notation

$$i = \begin{cases} \text{index for processing at Plant A} & = 1, \ldots, \text{m} \\ \text{index for processing at Plant B} & = \text{m+1}, \ldots, \text{m+n} \\ \text{index for processing at Plant C} & = \text{m+n+1}, \ldots, \text{m+n+q} \end{cases}$$

t = index for time periods = 1-L_A, ..., T

p = index for plants = 1 ... Plant A

= 2 ... Plant B

= 3 ... Plant C

Constant manufacturing lead times of L_A and L_B are assumed for components to get from Plants A and B respectively to Plant C. Consequently, the planning horizon is set from period 1 to T for Plant C, from period $1 - L_B$ to $T - L_B$ for Plant B and from period $1 - L_A$ to $T - L_A$ for Plant A. A bill of material index u_{ij} is used in the coupling constraints between each pair of upstream and downstream plants. This index represents the number of components j, manufactured at the upstream plant, which are required for each product i at the downstream plant. Without loss of generality, u_{ij} can have a value 1 for all i and j and this could represent two workcenters at the same plant. Our model can therefore be generalized to the situation of n plants with N different workcenters at each plant (Figure 6.1). The model is presented below.

Problem $C_{Ldfuture}$

Minimize $C_{Ldfuture} =$

$$\overbrace{\sum_{i=m+n+1}^{m+n+q} \sum_{t=1}^{T} h_i.I_{it} + \sum_{t=1}^{T} f_{3t}.z_{3t} + \sum_{t=1}^{T} v_{3t}.O_{3t}}^{PlantC}$$

$$+ \overbrace{\sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} h_i.I_{it} + \sum_{t=1-L_B}^{T-L_B} f_{2t}.z_{2t} + \sum_{t=1-L_B}^{T-L_B} v_{2t}.O_{2t}}^{PlantB}$$

$$+ \overbrace{\sum_{i=1}^{m} \sum_{t=1-L_A}^{T-L_A} h_i.I_{it} + \sum_{t=1-L_A}^{T-L_A} f_{1t}.z_{1t} + \sum_{t=1-L_A}^{T-L_A} v_{1t}.O_{1t}}^{PlantA}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \qquad \forall t = 1, \ldots, T, \quad i = m+n+1, \ldots, m+n+q \tag{6.1}$$

$$\sum_{i=m+1}^{m+n} (s_i.Y'_{it} + b_i.X'_{it}) - O'_{3t} \leq Cap_{3t} \qquad \forall t = 1, \ldots, T \tag{6.2}$$

$$X_{it} \leq M_{i3t}.Y_{it} \qquad \forall t = 1, \ldots, T, \quad i = m+n+1, \ldots, m+n+q \tag{6.3}$$

$$O_{3t} \leq OT_{3t}.z_{3t} \qquad \forall t = 1, \ldots, T, \tag{6.4}$$

$$I_{it-1} + X_{it} - I_{it} - P_{it}{}^{BC} = 0 \qquad \forall t = 1 - L_B, \ldots, T - L_B, \quad i = m+1, \ldots, m+n \tag{6.5}$$

$$\sum_{i=m+1}^{m+n} (s_i.Y'_{it} + b_i.X'_{it}) - O'_{2t} \leq Cap_{2t} \qquad \forall t = 1 - L_B, \ldots, T - L_B \tag{6.6}$$

$$X_{it} \leq M_{i2t}.Y_{it} \qquad \forall t = 1 - L_B, \ldots, T - L_B, \quad i = m+1, \ldots, m+n \tag{6.7}$$

$$O_{2t} \leq OT_{2t}.z_{2t} \qquad \forall t = 1-L_B,\ldots,T-L_B, \quad (6.8)$$

$$\sum_{i=m+n+1}^{m+n+q} u_{ij}.X_{it} = P^{BC}_{jt-L_B} \qquad \forall t = 1,\ldots T, \quad \forall j = m+1,\ldots,m+n \quad (6.9)$$

$$I_{it-1} + X_{it} - I_{it} - P_{it}^{AB} = 0 \qquad \forall t = 1-L_A,\ldots,T-L_A, \quad i = 1,\ldots,m \quad (6.10)$$

$$\sum_{i=1}^{m}(s_i.Y'_{it} + b_i.X'_{it}) - O'_{1t} \leq Cap_{1t} \qquad \forall t = 1-L_A,\ldots,T-L_A \quad (6.11)$$

$$X_{it} \leq M_{i1t}.Y_{it} \qquad \forall t = 1-L_A,\ldots,T-L_A, \quad i = 1,\ldots,m \quad (6.12)$$

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1-L_A,\ldots,T-L_A \quad (6.13)$$

$$\sum_{j=m+1}^{m+n} u_{jk}.X_{jt-L_B} = P^{AB}_{kt-L_A} \qquad \forall t = 1,\ldots T, k = 1,\ldots,m \quad (6.14)$$

$$X_{it} = X'_{it} \qquad \forall i,t \quad (6.15)$$

$$Y_{it} = Y'_{it} \qquad \forall i,t \quad (6.16)$$

$$O_{pt} = O'_{pt} \qquad \forall p,t \quad (6.17)$$

$$z_{pt}, Y_{it} \in (0,1), \qquad X_{it}, I_{it} \geq 0, \quad I_{i0} = 0 \qquad \forall t,i,p \quad (6.18)$$

The cost terms in the objective function representing each of the plants A, B and C are indicated above. These terms represent the total cost comprising inventory holding, fixed overtime and variable overtime costs at each plant. Constraints 6.1 to 6.4 represent the problem for Plant C, 6.5 to 6.9, the problem for Plant B and 6.10 to 6.14, the problem for Plant A. The physical significance of these constraints is identical to that explained in our earlier uncoordinated and coordinated models and is not repeated here. Constraints 6.9 represent the coupling constraints between the processes at Plant C and Plant B while constraints 6.14 couple Plants B and A. Constraints 6.15, 6.16, 6.17 represent respectively copies of the production variable X_{it}, setup variable Y_{it} and the overtime variable O_{pt} at each plant. In order to obtain tractable sub-problems, the variable copying constraints are dualized and a penalty term is added to the objective function as explained below.

a. Dualize constraints (6.9) with dual variables π_{jt} and add the penalty function $\pi_{jt}.(P^{BC}_{jt-L_B} - \sum_{i=m+n+1}^{m+n+q} u_{ij}.X_{it})$ for all j = m+1, ...,m+n and t = 1, ...,T to the objective function.

b. Dualize constraints (6.14) with dual variables ξ_{kt} and add the penalty function $\xi_{kt}.(P^{AB}_{kt-L_A} - \sum_{j=m+1}^{m+n} u_{jk}.X_{jt-L_B})$ for all k = 1, ...,m and t = 1, ...,T to the objective function.

c. Dualize constraints (6.15) with dual variables ρ_{it} and add a penalty function $\rho_{it}(X'_{it}$ - $X_{it})$ for all i, t to the objective function

d. Dualize constraints (6.16) with dual variables γ_{it} and add a penalty function $\gamma_{it}(Y'_{it}$ - $Y_{it})$ for all i, t to the objective function

e. Dualize constraints (6.17) with dual variables δ_{pt} and add a penalty function $\delta_{pt}(O'_{pt}$ - $O_{pt})$ for all p, t = 1, ...,T to the objective function

Separating subsets of constraints, the Lagrangian problem decomposes into three independent problems for each plant which can be solved in a manner identical to that explained in Chapter 4.

Decomposed Problems for $C_{Ldfuture}$

Problem $C_{Ldfuture1}$

Minimize $C_{CLdfuture1}$ =

$$\sum_{i=m+n+1}^{m+n+q} \sum_{t=1}^{T} h_i.I_{it} - \sum_{i=m+n+1}^{m+n+q} \sum_{t=1}^{T} \rho_{it}.X_{it} - \sum_{i=m+n+1}^{m+n+q} \sum_{t=1}^{T} \sum_{k=m+1}^{m+n} \pi_{kt}.u_{ik}.X_{it} - \sum_{i=m+n+1}^{m+n+q} \sum_{t=1}^{T} \gamma_{it}.Y_{it}$$

subject to

$$I_{it-1} + X_{it} - I_{it} = d_{it} \quad \forall t = 1, \ldots, T, \quad i = m+n+1, \ldots, m+n+q$$

$$X_{it} \leq M_{i3t}.Y_{it} \quad \forall t = 1, \ldots, T, \quad i = m+n+1, \ldots, m+n+q$$

Problem $C_{Ldfuture2}$

Minimize $C_{CLdfuture2} =$

$$\sum_{t=1}^{T} v_{3t}.O_{3t} + \sum_{t=1}^{T} f_{3t}.z_{3t} - \sum_{t=1}^{T} \delta_{3t}.O_{3t}$$

subject to

$$O_{3t} \leq OT_{3t}.z_{3t} \qquad \forall t = 1, \ldots, T$$

Problem $C_{Ldfuture3}$

Minimize $C_{CLdfuture3} =$

$$\sum_{i=m+n+1}^{m+n+q} \sum_{t=1}^{T} \rho_{it}.X'_{it} + \sum_{i=m+n+1}^{m+n+q} \sum_{t=1}^{T} \gamma_{it}.Y'_{it} + \sum_{t=1}^{T} \delta_{3t}.O'_{3t}$$

subject to

$$\begin{aligned}
\sum_{i=m+n+1}^{m+n+q} (s_i.Y'_{it} + b_i.X'_{it}) - O'_{3t} &\leq Cap_{3t} \qquad \forall t = 1, \ldots, T \\
O'_{3t} &\leq OT_{3t} \qquad \forall t = 1, \ldots, T \\
X'_{it} &\leq M_{i3t}.Y'_{it} \qquad \forall t = 1, \ldots, T, \\
& \qquad i = m+n+1, \ldots, m+n+q
\end{aligned}$$

Problem $B_{Ldfuture1}$

Minimize $C_{BLdfuture1} =$

$$\sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} h_i.I_{it} - \sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} \rho_{it}.X_{it} + \sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} \pi_{it}.P_{it}^{BC} - \sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} \gamma_{it}.Y_{it}$$

$$- \sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} \sum_{k=1}^{m} \xi_{kt}.u_{ik}.X_{it}$$

subject to

$$I_{it-1} + X_{it} - I_{it} - P_{it}^{BC} = 0 \qquad \forall t = 1 - L_B, \ldots, T - L_B, \quad i = m+1, \ldots, m+n$$

$$X_{it} \leq M_{i2t}.Y_{it} \qquad \forall t = 1 - L_B, \ldots, T - L_B, \quad i = m+1, \ldots, m+n$$

$$\sum_{t'=1-L_B}^{t} P_{kt'}^{BC} \geq \sum_{i=m+n+1}^{m+n+q} \sum_{t'=1}^{t+L_B} u_{ik}.d_{it'} \qquad \forall t = 1 - L_B, \ldots, T - L_B, \quad k = m+1, \ldots, m+n$$

Problem $B_{Ldfuture2}$

Minimize $C_{BLdfuture2} =$

$$\sum_{t=1-L_B}^{T-L_B} v_{2t}.O_{2t} + \sum_{t=1-L_B}^{T-L_B} f_{2t}.z_{2t} - \sum_{t=1-L_B}^{T-L_B} \delta_{2t}.O_{2t}$$

subject to

$$O_{2t} \leq OT_{2t}.z_{2t} \qquad \forall t = 1 - L_B, \ldots, T - L_B$$

Problem $B_{Ldfuture3}$

Minimize $C_{BLdfuture3} =$

$$\sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} \rho_{it}.X'_{it} + \sum_{i=m+1}^{m+n} \sum_{t=1-L_B}^{T-L_B} \gamma_{it}.Y'_{it} + \sum_{t=1-L_B}^{T-L_B} \delta_{2t}.O'_{2t}$$

subject to

$$\sum_{i=m+1}^{m+n} (s_i.Y'_{it} + b_i.X'_{it}) - O'_{2t} \leq Cap_{2t} \qquad \forall t = 1 - L_B, \ldots, T - L_B$$

$$O_{2t} \leq OT_{2t} \qquad \forall t = 1 - L_B, \ldots, T - L_B$$

$$X'_{it} \leq M_{i2t}.Y'_{it} \qquad \forall t = 1 - L_B, \ldots, T - L_B, \quad i = m+1, \ldots, m+n$$

Problem $A_{Ldfuture1}$

Minimize $C_{ALdfuture1} =$

$$\sum_{i=1}^{m}\sum_{t=1-L_A}^{T-L_A} h_i.I_{it} - \sum_{i=1}^{m}\sum_{t=1-L_A}^{T-L_A} \rho_{it}.X_{it} + \sum_{i=1}^{m}\sum_{t=1-L_A}^{T-L_A} \xi_{it}.P_{it}^{AB} - \sum_{i=1}^{m}\sum_{t=1-L_A}^{T-L_A} \gamma_{it}.Y_{it}$$

subject to

$$\begin{aligned}
I_{it-1} + X_{it} - I_{it} - P_{it}^{AB} &= 0 \qquad \forall t = 1 - L_A, \ldots, T - L_A, \\
&\qquad\qquad i = 1, \ldots, m \\
X_{it} &\leq M_{i1t}.Y_{it} \qquad \forall t = 1 - L_A, \ldots, T - L_A, \\
&\qquad\qquad i = 1, \ldots, m \\
\sum_{t'=1-L_A}^{t} P_{jt'}^{AB} &\geq \sum_{k=m+1}^{m+n}\sum_{t'=1-L_B}^{t+L_A-L_B} u_{kj}.\sum_{i=m+n+1}^{m+n+q}\sum_{t'=1}^{t+L_A} u_{ik}.d_{it'} \\
&\qquad\qquad \forall t = 1 - L_A, \ldots, T - L_A, \\
&\qquad\qquad j = 1, \ldots, m
\end{aligned}$$

Problem $A_{Ldfuture2}$

Minimize $C_{ALdfuture2} =$

$$\sum_{t=1-L_A}^{T-L_A} v_{1t}.O_{1t} + \sum_{t=1-L_A}^{T-L_A} f_{1t}.z_{1t} - \sum_{t=1-L_A}^{T-L_A} \delta_{1t}.O_{1t}$$

subject to

$$O_{1t} \leq OT_{1t}.z_{1t} \qquad \forall t = 1 - L_A, \ldots, T - L_A$$

Problem $A_{Ldfuture3}$

Minimize $C_{ALdfuture3} =$

$$\sum_{i=1}^{m}\sum_{t=1-L_A}^{T-L_A} \rho_{it}.X'_{it} + \sum_{i=1}^{m}\sum_{t=1-L_A}^{T-L_A} \gamma_{it}.Y'_{it} + \sum_{t=1-L_A}^{T-L_A} \delta_{1t}.O'_{1t}$$

subject to

$$\sum_{i=1}^{m}(s_i.Y'_{it} + b_i.X'_{it}) - O'_{1t} \leq Cap_{1t} \qquad \forall t = 1 - L_A, \ldots, T - L_A$$

$$O_{1t} \leq OT_{1t} \qquad \forall t = 1 - L_A, \ldots, T - L_A$$

$$X'_{it} \leq M_{i1t}.Y'_{it} \qquad \forall t = 1 - L_A, \ldots, T - L_A, \quad i = 1, \ldots, m$$

It is evident from above that the Lagrangian decomposition methodology as described in Chapter 4, can be used to solve the above sub-problems. We thus demonstrate the application of Lagrangian decomposition to the most general form of the multi-plant coordination problem, i.e., n plants with N different workcenters at each plant.

Some other directions emerging from this dissertation can be summarized as follows:

- The results obtained here need to be confirmed by real life case studies from the industry. The insights from this experience will be useful in strengthening the proposed research framework. This is an important direction for future research.
- Reducing the gap between the upper and lower bounds could make use of other methodologies than subgradient optimization. It has been suggested (see Guignard and Kim, 1987) that the dual ascent method may be more efficient in cases where a large number of dual multipliers are involved. The suitability of the dual ascent method for optimizing the Lagrangian dual of this problem needs to be studied. A promising methodological direction that must be explored further is the use of **interior point methods**. See for example Goffin et al., 1992.
- The exploitation of $v(C'_{dual})$ as a stopping criterion for the solution of the Lagrangian dual, and especially for generating improved solutions, needs to be explored further.

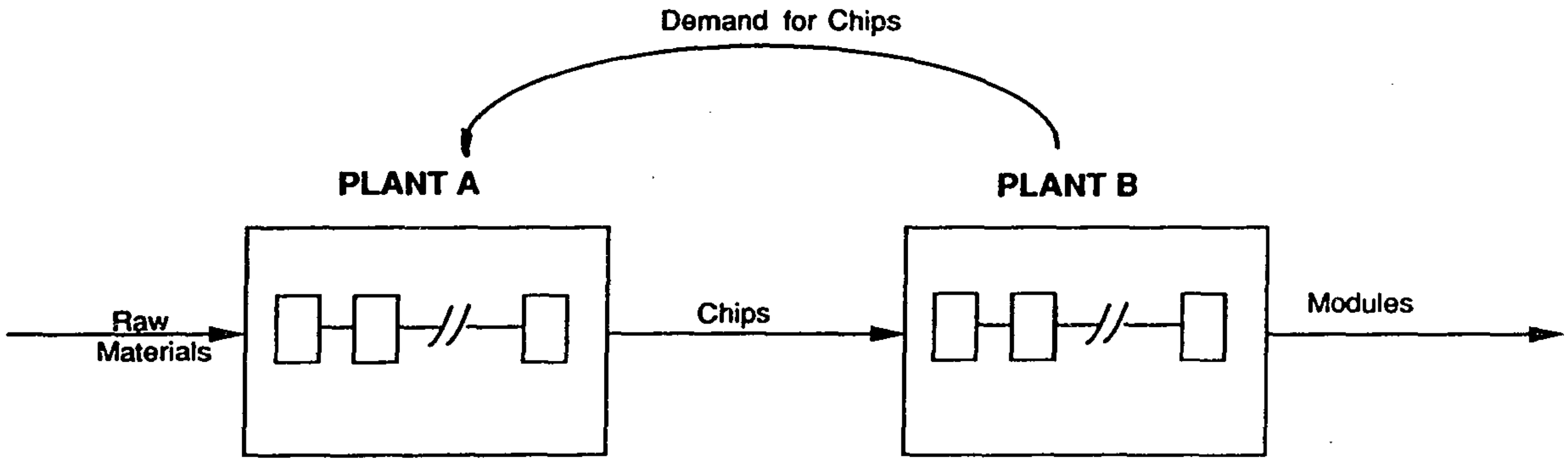

Figure 6.1:Multi-Plant Coordination Problem with Multiple Workcenters

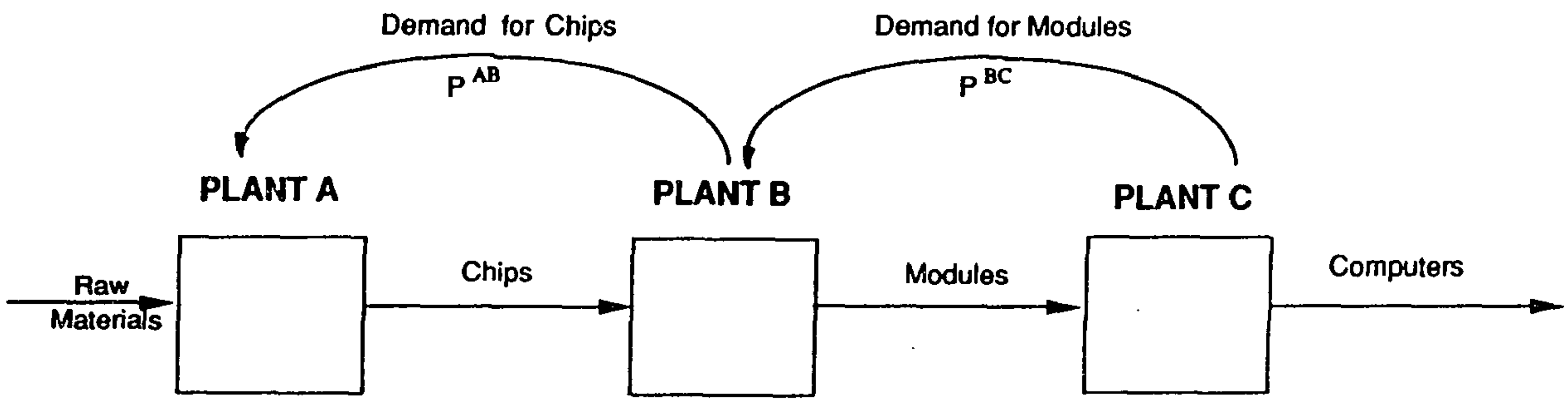

Figure 6.2: Multi-Plant Coordination Problem with Three Plants

Bibliography

Afentakis, P., Gavish, B., and Karmarkar, U.S. (1984), "Computationally efficient optimal solution to the lot sizing problem in multi-stage assembly systems," *Management Science*, 30, 222-239.

Anily, S., and Federgruen, A. (1990), "One warehouse multiple retailer systems with vehicle routing costs," *Management Science*, 36, 92-114.

Baker, K.R. (1977), "An experimental study of the effectiveness of rolling schedules in production planning," *Decision Sciences*, 8, 19-27.

Baker, K.R., Dixon, P., Magazine, M.J., and Silver, E.A. (1978), " An algorithm for the dynamic lotsize problem with time varying production capacity constraints," *Management Science*, 24, 1710-1720.

Baker, K.R., and Peterson D.W. (1979), "An analytic framework for evaluating rolling schedules," *Management Science*, 25, 341-351.

Banerjee, A. (1986a), "On A quantity discount pricing model to increase vendor's profits," *Management Science*, 32, 1513-1517.

Banerjee, A. (1986b), "A joint economic lotsize model for purchaser and vendor," *Decision Sciences*, 17, 292-311.

Barany, I., Van Roy T.J., and Wolsey, L.A. (1984), "Strong formulations for multi-item, capacitated lotsizing," *Management Science*, 30, 1255-1261.

Bassok, Y., and Akella, R. (1991), "Ordering and Production Decisions with Supply Quality and Demand Uncertainty," *Management Science*, 37, 1556-1574.

Beek, P., Bremer, A., and Putten, C. (1985), "Design and optimization of multi-echelon assembly networks: Savings and Potentialities," *European Journal of Op-*

erational Research, 19, 57-67.

Bell, W., Dalberto, L.M., Fisher, M.L., Greenfield, A.J., Jaikumar, R., Kedia, P., Mack, R.G., and Plutzman, P.J. (1983), "Improving the distribution of industrial gases with an on-line computerized routing and scheduling optimizer," *Interfaces*, 13, 14-23.

Bendiner, J. (1993), "Integrated Logistics Management: Benefits and Challenges," *OR/MS Today*, 20, 34-36.

Berry, W.L. (1972), "Lot sizing procedures for requirement planning systems : A framework for analysis," *Production and Inventory Management*, 13, 19-35.

Bhatnagar, R., Chandra, P., and Goyal, S. (1992), "Models for Multi-Plant Coordination," *European Journal of Operational Research*, 67, 141-160.

Bhatnagar, R., Chandra, P., Loulou, R., and Qiu, J. (1993), "Order Release and Product Mix Coordination in a Complex PCB Manufacturing Line with Batch Processors," Research Working Paper, McGill University.

Billington, C. (1994), "Supply Chain Management," *OR/MS Today*, 21, 20-27.

Billington, P.J., McClain, J.O., and Thomas, L.J. (1983), "Mathematical Programming approaches to capacity constrained MRP systems:Review, formulation and problem reduction," *Management Science*, 29, 1126-1141.

Bitran, G.R., and Matsuo, H. (1986), "Approximation formulations for the single product capacitated lot size problem," *Operations Research*, 34, 63-74.

Blackburn, J.D., and Millen, R.A. (1980), "Heuristic Lot-Sizing Performance in a Rolling-Schedule Environment," *Decision Sciences*, 11, 691-701.

Blackburn, J.D., and Millen, R.A. (1982), "Improved heuristics for multi-stage requirement planning systems," *Management Science*, 28, 44-56.

Blackburn, J.D., Kropp, D.H., and Millen, R.A. (1986), "A comparison of strategies to dampen nervousness in MRP systems," *Management Science*, 32, 413-429.

Blumenfeld, J.D., Burns, L.D., Daganzo, C.F., Frick, M.C., and Hall, R.W. (1987), "Reducing logistics cost at General Motors," *Interfaces*, 17, 26-47.

Burns, L.D., Hall R.W., Blumenfeld, D.E., and Daganzo, C.F. (1985), "Distribution strategies that minimize transportation and inventory costs," *Operations Research*, 31, 469-490.

Carlson, R.C., Jucker, J.V. and Kropp, D.H. (1979), "Less Nervous MRP Systems: A Dynamic Economic Lot Sizing Approach," *Management Science*, 25, 754-761.

Carlson, R.C., Beckman, S.L. and Kropp, D.H. (1982), "The effectiveness of extending the horizon in rolling production scheduling," *Decision Sciences*, 13, 129-146.

Carlson, R.C., and Yano, C.A. (1986), "Safety stocks in MRP systems with emergency setups for components," *Management Science*, 32, 403-412.

Chand, S. (1983), "Rolling Horizon Procedures for the facilities in series inventory model with nested schedules," *Management Science*, 29, 237-249.

Chandra, P. (1993), "A dynamic distribution model with warehouse and customer replenishment requirements," *Journal of Operational Research Society*, 44, 681-692.

Chandra, P., and Fisher, M.L. (1994), "Coordination of Production and distribution planning," *European Journal of Operational Research*, 72, 503-517.

Chandra, P. and Gupta, S. (1993), "An Analysis of a Last Station-Bottleneck semiconductor Packaging Line," Research Working Paper, McGill University.

Chen, W.-H., and Thizy, J-M. (1990), "Analysis of relaxations for the multi-item capacitated lotsizing problem," *Annals of Operations Research*, ed. M. Queyranne, JC Baltzer AG Scientific Publishing Company.

Chung, C., and Lin, C.M. (1988), "An $O(T^2)$ algorithm for the NI/G/NI/ND capacitated lotsizing problem," *Management Science*, 34, 420-426.

Clark, A.J., and Scarf, H. (1960), "Optimal policies for a multi-echelon inventory problem," *Management Science*, 6, 475-490.

Clark, A.J., and Scarf, H. (1962), "Approximate solutions to a simple multi-echelon inventory problem," in: K.J. Arrow et al.(eds), *Studies in Applied Probability and Management Science*, Stanford University Press, Stanford, CA, 88-100.

Cohen M.A., and Lee, H.L. (1988),"Strategic Analysis of integrated production distribution systems: Models and Methods," *Operations Research*, 36, 216-228.

Cohen M.A., and Lee, H.L. (1989),"Resource Deployment analysis of global manufacturing and distribution networks," *Journal of Manufacturing and Operations Management*, 2, 81-104.

Coleman, B.J., and McKnew, M.A. (1991), "An improved heuristic for multi-level lotsizing in materials requirement planning," *Decision Sciences*, 22, 136-156.

Cooper, W.W., Sinha, K.K., and Sullivan, R.S. (1992), "Measuring Complexity in High- Technology Manufacturing: Indexes for Evaluation," *Interfaces*, 22, 4, 38-48.

Cormen, T.H., Leiserson, C.E., and Rivest, R.L. (1990), Introduction to Algorithms, MIT Press, Cambridge, Massachusetts.

Crowston, W.B., and Wagner, M. (1973), "Dynamic lot size models for multi-stage assembly systems," *Management Science*, 20, 14-21.

Crowston, W.B., Wagner, M. and Williams, J.F. (1973), "Economic lot size determination in multi stage assembly systems," *Management Science*, 19, 517-527.

Dantzig, G.B. and Wolfe, P. (1960), "Decomposition Principle for Linear Programs," *Operations Research*, 8, 101-111.

De Bodt, M., and Van Wassenhove, L.N. (1983), "Cost Increases due to Demand Uncertainty in MRP lotsizing," *Decision Sciences*, 14, 345-362.

De Bodt, M., Van Wassenhove, L.N., and Gelders, L. (1982), "Lot Sizing and safety stock decisions in an MRP system with demand uncertainty," *Engineering Costs and Production Economics*, 6, 67-75.

De Matteis, J.J. (1971), "An economic lot sizing technique, the part period algorithm," *IBM Systems Journal*, 7, 30-38.

Dertouzos, M.L., Lester, R.K., Solow, R.M. and The MIT Commission on Industrial Productivity (1990), "Made in America Regaining the Productive Edge," Harper Perennial Publishers, N.Y.

Diaby, M., Bahl, H.C., Karwan, M.H., and Zionts, S. (1992a), "Capacitated Lotsizing and scheduling by Lagrangian relaxation," *European Journal of Operational Research*, 59, 444-458.

Diaby, M., Bahl, H.C., Karwan, M.H., and Zionts, S. (1992b), "A Lagrangian Relaxation Approach for Very-Large-Scale Capacitated Lotsizing," *Management*

Science, 38, 1329-1340.

Dixon P., and Silver, E.A. (1981), "A heuristic solution procedure for the multi-item, single level, limited capacity, lotsizing problem," *Journal of Operations Management*, 2, 23-39.

Dogramaci, A., Panayiotopoulos, J.C., and Adam, N.R. (1981), "The dynamic lotsizing problem for multiple items under limited capacity," *AIIE Transactions*, 13, 294-303.

Dror, M., and Ball, M. (1987), "Inventory/routing: Reduction from an annual to a short period," *Naval Research Logistics Quarterly*, 34, 891-905.

Drucker, P.F. (1990), "The Emerging Theory of Manufacturing," *Harvard Business Review*, May-June, 94-102.

Dzielinski, B., and Gomory, R. (1965), "Optimal programming of lot sizes, inventories and labor allocations," *Management Science*, 11, 874-890.

Eisenhut, P.S. (1975), "A dynamic lotsizing algorithm with capacity constraints," *AIIE Transactions*, 7, 170-176.

Eppen, G.D., and Martin, R.K. (1987), "Solving multi-item capacitated lotsizing problems using variable redefinition," *Operations Research*, 35, 832-848.

Far Eastern Economic Review (1989), "Taiwan's Memory Test," 17 August 1989, p. 67.

Federgruen, A., and Zipkin, P. (1984), "A combined vehicle routing and inventory allocation problem," *Operations Research*, 32, 1019-1037.

Fisher, M.L. (1981), "The Lagrangian Relaxation Method for Solving Integer

Programming Problems," *Management Science*, 27, 1-18.

Fisher, M.L., and Rinnooy Kan A.H.G. (1988), "The design, analysis and implementation of heuristics," *Management Science*, 34, 263-265.

Florian, M., and Klein, M. (1971), "Deterministic Production Planning with concave costs and capacity constraints," *Management Science*, 18, 12-20.

Florian, M., Lenstra, J.K., and Rinnooy Kan A.H.G. (1980), "Deterministic production planning : algorithms and complexity," *Management Science*, 26, 669-679.

Geoffrion, A.M. (1974), "Lagrangian Relaxation and its Uses in Integer Programming," *Mathematical Programming Study*, 2, 82-114.

Glover, F., and Mulvey, J. (1980), "Equivalence of the 0-1 integer programming problem to discrete generalized and pure network," *Operations Research*, 28, 829-835.

Goffin, J.L. (1977), "On the convergence rates of subgradient optimization methods," *Mathematical Programming*, 13, 329-347.

Goffin, J.L., Haurie, A., and Vial, J.P. (1992), "Decomposition and Nondifferentiable Optimization with the Projective Algorithm," *Management Science*, 38, 284-302.

Goyal, S.K. (1988), "A joint economic lot size model for purchaser and vendor: A comment," *Decision Science*, 19, 236-241.

Goyal, S.K., and Gupta, Y.P. (1989), "Integrated Inventory and Models: The buyer vendor coordination," *European Journal of Operations Research*, 41, 261-269.

Graves, S.C. (1988), "Safety Stocks in Manufacturing Systems," *Journal of Man-*

ufacturing and Operations Management, 1, 67-101.

Guignard, M., and Kim, S. (1987), "Lagrangian Decomposition: A Model yielding stronger Lagrangian Bounds," *Mathematical Programming*, 39, 215-228.

Guignard, M. (1984), "Lagrangian Decomposition: An Improvement over Lagrangean and Surrogate Duals," *Department of Statistics Report*, No. 62, University of Pennsylvania.

Hansmann, F. (1959), "Optimal inventory location and control in production and distribution networks," *Operations Research*, 7, 483-498.

Held, M., Wolfe, P., and Crowder, H.D. (1974), "Validation of subgradient optimization," *Mathematical Programming*, 6, 62-88.

Ho, C. (1989), "Evaluating the impact of operating environments on MRP system nervousness," *International Journal of Production Research*, 27, 1115-1135.

Ishii, K., Takahashi, K., and Muramatsu, R. (1988), "Integrated production, inventory, and distribution systems," *International Journal of Production Research*, 26, 473-482.

Jagannathan, R. and Rao, M.R. (1973), "A class of deterministic production planning problems," *Management Science*, 19, 1295-1300.

Joglekar, P.N. (1988), "Comments on A Quantity discount pricing model to increase vendor's profits, " *Management Science*, 34, 1391-1398.

Karmarkar, U.S. (1987), "Lotsize, leadtimes and in process inventories," *Management Science*, 33, 409-418.

Karmarkar, U.S. (1989), "Getting control of just-in-time," *Harvard Business Re-*

view, 5, 122-131.

Karmarkar, U.S. and Schrage, L. (1985), "The deterministic dynamic product cycling problem," *Operations Research*, 33, 326-345.

Karmarkar, U.S., Kekre, S. and Kekre, S. (1983), "Multi Item Lotsizing and Manufacturing lead times," Working Paper No. QM8325, University of Rochester.

Karmarkar, U.S., Kekre, S., and Kekre, S. (1992), "Multi-item batching heuristics for minimization of queuing delays," *European Journal of Operational Research*, 58, 99-111.

Karni, R. and Roll, Y. (1982), "A heuristic algorithm, for the multi-item lotsizing problem with capacity constraints," *AIIE Transactions*, 14, 249-256.

King, R.H., and Love, R.R. (1980), "Coordinating decisions for increased profits," *Interfaces*, 10, 4-19.

Kropp, D.H., and Carlson, R.C. (1984), "A lot sizing algorithm for reducing nervousness in MRP systems," *Management Science*, 30, 240-244.

Kumar A., Akella, R., and Cornuejols, G. (1990), "Inventory/production decisions under uncertain but bounded demand conditions with implications for supply contracts," Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University.

Lambrecht, M.R., and VanderEcken, J. (1978a), "A facility in series capacity constrained single facility dynamic lotsize model," *European Journal of Operational Research*, 2, 42-49.

Lambrecht, M.R., and VanderEcken, J. (1978b), "A capacity constrained single facility dynamic lot size model," *European Journal of Operational Research*, 2, 132-136.

Lambrecht, M.R., and Vanderveken, H. (1979), "Heuristic Procedures for the single operation multi-item loading problem," *AIIE Transactions*, 18, 319-326.

Lambrecht, M.R., Muckstadt, J.A., and Luyten, R. (1984), "Protective stocks in multi-stage production systems," *International Journal of Production Research*, 22, 319-328.

Lambrecht, M.R., Luyten, R., and VanderEcken, J. (1985), "Protective inventories and bottlenecks in production systems," *European Journal of Operational Research*, 22, 319-328.

Lasdon, L.S. and Terjung, R.C. (1971), "An efficient algorithm for multi-item scheduling," *Operations Research*, 19, 946-969.

Lee, H.L., and Rosenblatt, M.J. (1986), "A generalized quantity discount pricing model to increase supplier's profits," *Management Science*, 32, 1177-1185.

Lee, H.L., and Billington, C. (1992), "Managing Supply Chain Inventory: Pitfalls and Opportunities," *Sloan Management Review*, Spring 1992, 65-72.

Lee, H.L., Billington, C., and Brent, C. (1993), "Hewlett-Packard Gains Control of Inventory and Service through Design for Localization," *Interfaces*, 23, 1-11.

Love, S.F. (1972), "A facilities in series inventory model with nested schedules," *Management Science*, 18, 327-338.

Love, S.F. (1973), "Bounded production and inventory models with piecewise concave costs," *Management Science*, 20, 313-318.

Lozano, S., Larraneta, J., and Onieva, L. (1991), "Primal-Dual approach to the single level capacitated lotsizing problem," *European Journal of Operational Re-*

search, 51, 354-366.

Luss, H. (1982), "Operations Research and Capacity Expansion Problems: A Survey," *Operations Research*, 30, 907-947.

Maes, J., and Van Wassenhove, L.N. (1986a), "Multi-item single level capacitated dynamic lot sizing heuristics : A computational comparison (Part I : Static Case)," *IIE Transactions*, 18, 114-123.

Maes, J., and Van Wassenhove, L.N. (1986b), "Multi-item single level capacitated dynamic lot sizing heuristics : A computational comparison (Part II : Rolling Horizon)," *IIE Transactions*, 18, 124-129.

Maes, J., and Van Wassenhove, L.N. (1988), "Multi-Item Single-Level Capacitated Dynamic Lot-Sizing Heuristics : A General Review," *Journal of the Operational Research Society*, 39, 991-1004.

Maes, J., McClain, J.O., and Van Wassenhove, L.N. (1991), "Multi-level capacitated lotsizing complexity and LP based heuristics," *European Journal of Operational Research*, 53, 131-148.

Magnanti, T.L. and Vachani, R. (1990), "A Strong Cutting Plane Algorithm for Production Scheduling with Changeover Costs," *Operations Research*, 38, 3, 456-473.

Manne, A.S. (1958), "Programming of economic lotsizes," *Management Science*, 4, 115-135.

Mather, H. (1977), "Reschedule the Reschedules You Just Rescheduled - Way of Life for MRP?," *Production and Inventory Management*, 18, 60-79.

Millar, H.H., and Yang, M. (1993), "An Application of Lagrangean Decomposition to the Capacitated Multi-item Lot Sizing Problem," Working Paper, Saint Mary's

University, Nova Scotia.

Miller, J.G. (1979), "Hedging the master schedule," in L.P. Ritzman, L.J. Krajewski, W.L. Berry, S.H. Goodman, S.T. Hardy and L.D. Vitts (eds.), *Disaggregation Problems in Manufacturing and service organizations*, Martinus Nijhoff, Boston, MA, 237-256.

Monahan, J.P. (1984), "A quantity discount pricing model to increase vendor's profits," *Management Science*, 30, 720-726.

Monden, Y. (1983), *Toyota Production System*, IIE Press, Norcross, GA.

More, S.M. (1974), "MRP and the Least total cost method of lot sizing," *Production and Inventory Management*, 15, 47-55.

New, C.C. (1974), "Lot sizing in multi-level requirements planning," *Production and Inventory Management*, 15, 57-71.

Newson, E.F.P. (1975a), "Multi-item lotsize scheduling by heuristic part I: with fixed resources," *Management Science*, 21, 1186-1193.

Newson, E.F.P. (1975b), "Multi-item lotsize scheduling by heuristic part II: with variable resources," *Management Science*, 21, 1194-1203.

OECD (1992), "Technology and Economy: The Key relationships," Organisation for Economic Co-operation and Development, OECD, Paris, CEDEX, 16, France.

OECD (1991), "Technology and Productivity: The Challenge for Economic Policy," Organisation for Economic Co-operation and Development, OECD, Paris, CEDEX, 16, France.

Pochet, Y., and Wolsey, L.A. (1991), "Solving multi-item lotsizing problems using strong cutting planes," *Management Science*, 37, 53-67.

Pyke, D.F., and Cohen, M.A. (1990a), "Performance Characteristics of Stochastic Integrated Production-Distribution System," Working Paper, No. 244, The Amos Tuck School of Business Administration, Dartmouth College, Hanover, New Hampshire 03755.

Pyke, D.F., and Cohen, M.A. (1990b), "Multi Product Integrated Production-Distribution Systems," Working Paper, No. 254, The Amos Tuck School of Business Administration, Dartmouth College, Hanover, New Hampshire 03755.

Rosenblatt, M.J., and Lee, H.L. (1985), "Improving profitability with quantity discounts under fixed demand," *IIE Transactions*, 17, 388-395.

Schmidt, C.P., and Nahmias, S. (1985), "Optimal Policy for a two stage assembly system under random demand," *Operations Research*, 33, 1130-1145.

Silver, E.A., and Meal, H.C. (1973), "A heuristic for selecting lot size quantities for the case of deterministic time varying demand rate and discrete opportunity for replenishment," *Production and Inventory Management*, 14, 64-77.

Simpson, K.F. (1958), "In-process inventories," *Operations Research*, 6, 863-873.

Skinner, W. (1974), "The focused factory," *Harvard Business Review*, 52, 113-121.

Sridharan, V., Berry, W.L., and Udayabhanu, V. (1987), "Freezing the Master Production Schedule Under Rolling Planning Horizons," *Management Science*, 33, 1137-1149.

Sridharan, V., and Berry, W.L. (1990), "Freezing the Master Production Schedule Under Demand Uncertainty," *Decision Sciences*, 21, 97-120.

Steinberg, E., and Napier, H.A. (1980), "Optimal multi-level lot sizing for requirements planning systems," *Management Science*, 26, 1258-1271.

Swoveland, C. (1975), "A deterministic multi-period production planning model with piecewise concave production and holding-backorder costs," *Management Science*, 21, 1007-1013.

Thizy, J.M. and Van Wassenhove, L.N. (1985), "Lagrangian Relaxation for the multi-item capacitated lotsizing problem : a heuristic approach," *AIIE Transactions*, 17, 308-313.

Thizy, J.M. (1991), "Analysis of Lagrangian decomposition for the multi item capacitated lotsizing problem," *INFOR*, 29, 271-283.

Trigeiro, W.W. Thomas., L.J. and McClain, J.O. (1989), "Capacitated lotsizing with setup times," *Management Science*, 35, 353-366.

Van Roy, T.J., and Wolsey, L.A. (1987), "Solving mixed integer programming problems using automatic formulation," *Operations Research*, 35, 45-57.

Wagner, H., and Whitin, T. (1958), "Dynamic version of the economic lotsize model," *Management Science*, 4, 89-96.

Wets, R.J.-B. (1989), "Stochastic Programming," Chapter VIII in Optimization, Handbooks in Operations Research and Management Science, Volume I, Nemhauser, G. L., Rinnooy Kan A. H. G., and Todd, M. J., (eds.), Elsevier Science Publishers B.V.

Wijngaard, J., and Wortmann, J.C. (1985), "MRP and inventories," *European Journal of Operational Research*, 20, 281-293.

Williams, K., Haslam, C., Williams, J., Cutler, T., Adcroft, A., and Johal, S. (1992), "Against lean production," *Economy and Society*, 21, 321-354.

Williams, J.F. (1981), "Heuristic techniques for simultaneous scheduling of production and distribution in multi-echelon structures: Theory and empirical comparisons," *Management Science*, 27, 336-352.

Womack, J., Roos, D., and Jones, D. (1990), "The machine that changed the world, the story of lean production," Rawson Associates, New York.

Yano, C.A. (1987), "Setting planned leadtimes in serial production systems with tardiness costs," *Management Science*, 33, 95-106.

Yano, C.A., and Carlson, R.C. (1985), "An analysis of scheduling policies in multi-echelon production systems," *IIE Transactions*, 17, 370-377.

Yano, C.A., and Carlson, R.C. (1987), "Interaction between frequency of rescheduling and the role of safety stock in material requirements planning systems," *International Journal of Production Research*, 25, 221-232.

Zahorik, A., Thomas, L.J., and Trigeiro W.W. (1983), "Network programming models for production scheduling in multi-stage, multi-item, capacitated systems," *Management Science*, 30, 302-325.

Zangwill, W. (1966), "A deterministic multi-product, multi-facility, production and inventory model," *Operations Research*, 14, 486-507.

Zangwill, W. (1969), "A backlogging model and a multi-echelon model of a dynamic lotsize production system: A network approach," *Management Science*, 15, 506-527.

Zipkin, P.H. (1986), "Models for design and control of stochastic multi-item batch production systems," *Operations Research*, 34, 91-104.

Zoller, K. (1990), "Coordination of replenishment activities," *International Journal of Production Research*, 28, 1123-1136.

Appendix 1

Numerical Details for Sample Problem on page 105

The following numerical details relate to the sample problem discussed on page 105.

Modules at Plant B			
Module	*1*	*2*	*3*
Unit Holding Cost (in $)	16.68	16.88	17.92
Processing Time (time units per module)	1.54	1.49	1.47
Setup Time (time units per module)	5.65	5.47	4.87

Chips at Plant A				
Chips	*1*	*2*	*3*	*4*
Unit Holding Cost (in $)	4.39	4.92	4.13	3.30
Processing Time (time units per chip)	1.33	1.50	1.66	1.34
Setup Time (time units per chip)	5.20	5.77	5.03	4.77

Regular and Overtime Capacity		
	Plant A	*Plant B*
Regular Capacity (time units per period)	558	219
Overtime Capacity (time units per period)	140	55

Module Demand

Modules	*Periods* 1	2	3	4
1	15	27	6	2
2	8	57	128	100
3	36	1	152	152

Bill of Materials

Modules	*Chips* 1	2	3	4
1	1	0	0	0
2	0	1	2	0
3	0	1	1	1

The above problem was randomly generated for the case with high c.v. of demand and high c.v. of holding cost, low c.v. for setup and processing times, high setup/processing time ratio and high capacity utilization.

Appendix 2

Processing Time Variability (page 115)

We demonstrate the effect that unstable product designs could have on processing time variability as discussed on page 115. The choice of different processing time c.v.s in our experimental design was guided by an important characteristic of high-tech industries like computers and telecom equipment manufacturing etc. These industries are characterized by continuously changing product mix and product designs over the product life cycle. Here we show that the presence of even a few such products can considerably affect the c.v. of processing times. The example presented here is motivated by our experience at the Fiberworld Division of Northern Telecom, Montreal (see Bhatnagar et al., 1993). We consider 9 representative products out of the set of products manufactured at the above plant. The actual process relates to the mounting of components on printed circuit boards (PCBs). Typically this process is automated and is carried out by "pick and place" surface mount machines. However when a new product is not fully integrated into the manufacturing system or the PCB design is unstable, manual modification work may be necessary for these specific products. We consider three different scenarios representing different levels of product design stability as depicted below. Scenario 1 represents the situation when the product designs for all products are stable and therefore most of the products take similar times on a critical workcenter. Scenario 2 is the case where two products require extra manual work and hence have considerably higher processing times than the rest of the products. Scenario 3 is the converse of scenario 2, with two products requiring significantly lower processing times than the other products. The processing times of different products and the resulting c.v.s of processing times are shown below.

Product	Scenario 1	Scenario 2	Scenario 3
1	7	7	7
2	8	8	8
3	7	7	7
4	8	8	8
5	9	9	9
6	5	5	5
7	8	8	8
8	6	20	0
9	9	30	0
Average Processing Times	7.4	11.3	5.8
s.d. of Processing Times	1.25	8.2	3.5
c.v. of Processing Times	0.16	0.73	0.60
c.v. = coefficient of variation			
s.d. = standard deviation			

It is evident from the above that the current product life cycle stage of each product could have a substantial impact on the c.v. of processing times. In addition, the c.v. may vary over time, e.g., each of the scenarios 1, 2 and 3 could be true of the same firm at different points in time. Consideration of different processing time c.v.s therefore makes our experimental design more valid in terms of its content.

Appendix 3

Analysis of Variance for Cost Improvement

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	887.71	192	4.62		
caputil	40.72	1	40.72	8.81	.003
Cvdem	.10	1	.10	.02	.882
Cvhcost	9.80	1	9.80	2.12	.147
Cvproc	4.03	1	4.03	.87	.352
Cvset	2.19	1	2.19	.47	.492
Spratio	28.56	1	28.56	6.18	.014
Caputil by cvdem	10.85	1	10.85	2.35	.127
Caputil by cvhcost	4.54	1	4.54	.98	.323
Caputil by cvproc	7.53	1	7.53	1.63	.203
Caputil by cvset	.65	1	.65	.14	.708
Caputil by spratio	.71	1	.71	.15	.695
Cvdem by cvhcost	7.60	1	7.60	1.64	.201
Cvdem by cvproc	11.35	1	11.35	2.45	.119
Cvdem by cvset	.87	1	.87	.19	.665
Cvdem by spratio	1.64	1	1.64	.36	.552
Cvhcost by cvproc	10.68	1	10.68	2.31	.130
Cvhcost by cvset	1.22	1	1.22	.26	.608
Cvhcost by spratio	.37	1	.37	.08	.778
Cvproc by cvset	.94	1	.94	.20	.653
Cvproc by spratio	1.20	1	1.20	.26	.612
Cvset by spratio	.07	1	.07	.01	.905

Table 6.1: Analysis of Variance for Cost Improvement

Source of Variation	SS	DF	MS	F	Sig of F
Caputil by cvdem by cvhcost	.69	1	.69	.15	.699
caputil by cvdem by cvproc	1.43	1	1.43	.31	.579
caputil by cvdem by cvset	.03	1	.03	.01	.933
caputil by cvdem by spratio	27.63	1	27.63	5.98	.015
caputil by cvhcost by cvproc	.69	1	.69	.15	.699
caputil by cvhcost by cvset	.52	1	.52	.11	.739
caputil by cvhcost by spratio	5.09	1	5.09	1.10	.295
caputil by cvproc by cvset	.00	1	.00	.00	.979
caputil by cvproc by spratio	14.77	1	14.77	3.20	.075
caputil by cvset by spratio	1.91	1	1.91	.41	.521
cvdem by cvhcost by cvproc	.22	1	.22	.05	.828
cvdem by cvhcost by cvset	.28	1	.28	.06	.805

Table : 6.1 Analysis of Variance for Cost Improvement (continued)

Source of Variation	SS	DF	MS	F	Sig of F
cvdem by cvhcost by spratio	.01	1	.01	.00	.961
cvdem by cvproc by Cvset	1.67	1	1.67	.36	.548
cvdem by cvproc by Spratio	.01	1	.01	.00	.965
cvdem by cvset by Spratio	.00	1	.00	.00	.975
cvhcost by cvproc by cvset	.59	1	.59	.13	.721
cvhcost by cvproc by spratio	.11	1	.11	.02	.878
cvhcost by cvset by spratio	.32	1	.32	.07	.792
cvproc by cvset by Spratio	.18	1	.18	.04	.846
caputil by cvdem by cvhcost by cvproc	6.73	1	6.73	1.46	.229
caputil by cvdem by cvhcost by cvset	.01	1	.01	.00	.956
caputil by cvdem by cvhcost by spratio	4.76	1	4.76	1.03	.312
caputil by cvdem by cvproc by cvset	.89	1	.89	.19	.661

Table : 6.1 Analysis of Variance for Cost Improvement (continued)

Source of Variation	SS	DF	MS	F	Sig of F
caputil by cvdem by cvproc by spratio	4.70	1	4.70	1.02	.314
caputil by cvdem by cvset by spratio	.07	1	.07	.02	.901
caputil by cvhcost by cvproc by cvset	.13	1	.13	.03	.869
caputil by cvhcost by cvproc by spratio	.18	1	.18	.04	.846
caputil by cvhcost by cvset by spratio	.31	1	.31	.07	.796
caputil by cvproc by cvset by spratio	.35	1	.35	.08	.783
cvdem by cvhcost by cvproc by cvset	.00	1	.00	.00	.975
cvdem by cvhcost by cvproc by spratio	.89	1	.89	.19	.661
cvdem by cvhcost by cvset by spratio	1.14	1	1.14	.25	.620
cvdem by cvproc by Cvset by spratio	.02	1	.02	.00	.951
cvhcost by cvproc by cvset by spratio	2.66	1	2.66	.58	.449
caputil by cvdem by cvhcost by cvproc by cvset	.19	1	.19	.04	.841

Table : 6.1 Analysis of Variance for Cost Improvement (continued)

Source of Variation	SS	DF	MS	F	Sig of F
caputil by cvdem by cvhcost by cvproc by spratio	.75	1	.75	.16	.687
caputil by cvdem by cvhcost by cvset by spratio	.78	1	.78	.17	.682
caputil by cvdem by cvproc by cvset by spratio	.75	1	.75	.16	.687
caputil by cvhcost by cvproc by cvset by spratio	.24	1	.24	.05	.819
cvdem by cvhcost by cvproc by cvset by spratio	.89	1	.89	.19	.661
caputil by cvdem by cvhcost by cvproc by cvset by spratio	.89	1	.89	.19	.661
(Model)	229.12	63	3.64	.79	.866
(Total)	1116.82	255	4.38		

R-Squared = .205

Table : 6.1 Analysis of Variance for Cost Improvement (continued)