The Pennsylvania State University

The Graduate School

Department of Industrial and Manufacturing Engineering

INTELLIGENT SENSE RESPONSE SYSTEM FOR SUPPLY CHAIN EVENT MANAGEMENT

A Thesis in

Industrial Engineering and Operations Research

by

Nazrul Islam Shaikh

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2006

The thesis of Nazrul Islam Shaikh was reviewed and approved* by the following:

Vittaldas V. Prabhu
Associate Professor
Department of Industrial & Manufacturing Engineering
Thesis Advisor and Chair of Committee

Dr. Richard. A Wysk
Professor and Leonhard Chair in Engineering
Department of Industrial & Manufacturing Engineering

Dr. Gautam Natarajan
Associate Professor
Department of Industrial and Systems Engineering
Texas A&M University

Dr. Dennis Lin
Distinguished Professor of Statistics and Supply Chain Management
Department of Supply Chain and Information Systems

Dr. Arvind Rangaswamy,
Anchel Professor of Marketing
Department of Marketing

Dr. Richard J. Koubek
Professor and Department Head
Department of Industrial & Manufacturing Engineering

*Signatures are on file in the Graduate School

Abstract

More often than not, there are discrepancies between what was planned and what is /can be, or should eventually be executed. For instance, production shortages, missed deadlines, delays in receiving and dispatch of shipments, and order changes are more or less ubiquitous. If service level requirements (SLR) are associated with the successful implementation of the plans, these discrepancies between the planned and the executed can be expensive. In the business world, SLRs such as fill rates, on time delivery, zero defects, are associated with the successful implementation of the plans-and the discrepancies are expensive. The causes of these discrepancies can be attributed to poor planning or execution, or the occurrence of unplanned "events"-event management (EM) attempts to limit the losses incurred due to some of these "events" categories. The intelligent sense-response approach for EM focuses on the control theoretic approach.

Concepts developed within the realms of process control have proven that closed loop systems have superior performance in terms of maintaining the SLR and rejecting disturbances than the corresponding open loop systems. It is an extension of this concept, i.e., tighter integration between planning and execution through both feed forward and feedback mechanisms that the intelligent sense-response approach presented in this dissertation aims to exploit for performance enhancement of supply chains. Responding to the identified discrepancies is an important component of the proposed approach, and a bulk of this dissertation focuses on algorithms for efficient response. The rest of the dissertation focuses on the integration of the sensing capabilities that the supply chains are currently equipped with, with the proposed response mechanisms.

The first part of the dissertation proposes the aggregate high fidelity modeling (AHFM) approach for planning and replanning that provides evaluated responses to identified discrepancies with the objective of minimizing the losses as well as impact of the event on the SLRs of the system. AHFM aggregates the flexibility/redundancies that are available at various stages of planning and optimally assigns the resource/time from the aggregate pool to compensate the impact of the disturbance and problems. A solution approach for the AHFM approach, the computational complexity of which scales linearly or is polynomial is also presented. The second section of the dissertation discusses the integration of the AHFM approach with the sensing capabilities of the supply chain.

The formulation and solution of AHFM has been explored in several domains: manufacturing, air travel and food processing have been considered. Early investigation into these problem domains has yielded remarkable improvements in performance (up to 30% improvement in a discrete manufacturing environment, 84% improvement in process industries and up to 25% improvement in fleet scheduling). For the case study taken up for the IBM supply chain, the indicated improvements are of 27%. While these improvement percentages cannot be generalized for evaluating gains from implementing AHFM, the examples illustrate that at least in some cases, the magnitude of potential profits cannot be ignored-specially when this improvement comes at little extra cost. In addition, preliminary results of AHFM indicate that the technique yields feasible solution to problems that were considered infeasible.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am deeply indebted to Dr. Vittal Prabhu, Associate Professor of Industrial and Manufacturing Engineering, Penn State University for his constant support, guidance and motivation throughout the research. He constantly instilled interest in me to explore new and innovative topics to increase my vision on the topic of my research.

I am also grateful to Dr. Richard Wysk for provided insights and guidance throughout the project, and Dr. Arvind Rangaswamy for his patients and feedbacks on the thesis.

Finally, I thank my family and lab members for all the help and support provided. Their comfort and encouragement was one of the prime drivers in my completion of my thesis.

Chapter 1

Intelligent sense-response for supply chain event management

1.1 Introduction

More often than not, there are discrepancies between what was planned and what is /can be, or should eventually be executed. For instance, production shortages, missed deadlines, delays in receiving and dispatch of shipments, and fluctuating prices are more or less ubiquitous. If service level requirements (SLR) are associated with the successful implementation of the plans, these discrepancies between the planned and the executed can be expensive. In the business world, SLRs such as fill rates, on time delivery, zero defects, are associated with the successful implementation of the plans-and the discrepancies are expensive. Taking the change in shareholders wealth as an indicator of the effect of the breach in the SLR by a company, Singhal and Hendricks (2003) found a staggering decrease of about $120mn or more per company per supply chain malfunction.

Tighter coupling between planning and execution (p&e) has the potential to decrease the impact that these discrepancies have. Concepts developed within the realms of process control have proved that close loop systems have superior performance in terms of maintaining the SLR and rejecting disturbances than the corresponding open loop systems (see for example Ogata, 2000, Stephanopoulos, 1996). By closing the loop between p&e, event management (EM) systems attempt to exploit precisely this property to achieve performance enhancement.

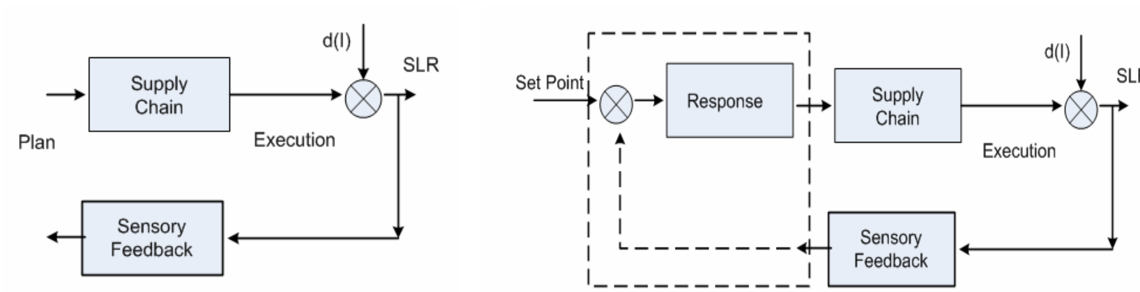
This idea of a closed loop p&e systems is gaining currency in the industrial settings. Market leaders in enterprise resource planning (ERP) and supply chain management (SCM) solutions providers, SAP, Viewlocity, Bearing Point, Highjump Software, etc made an early entry in the field as providers of the service. Some of the capabilities that these EM systems offer are [1-6]:

- Improved visibility across the supply chain and availability of desired statistics. For instance, improved information quality about product availability, orders, and fulfillment status.
- Real-time information on supply chain events that affect plans and schedules.
- Connections with logistics providers and alerts for the entire user network.
- Access to inventory levels, WIP, and order backlogs across the network.

These developments and features mostly address the visibility aspect of EM. The expectations from EM are two folds:

1. Prevention of the events from having an impact on the systems SLRs.
2. If prevention is not possible, then early response to limit the losses incurred due to the variance between the planned and the executed.

Responding to the identified discrepancies is an important component of EM. The focus of this dissertation is on defining and characterizing the EM problem and integrating the sensing capabilities with the response mechanisms that are currently lacking (shown in Figure 1-1). This will enable the road-mapping of activities for further/future development of EM systems.



(a) State of current practice (b) Focus of proposed research

Figure 1-1 Contribution of proposed research to EM

1.2 A perspective on the EM problem

The planning problem can be viewed from both a static as well as a dynamic perspective. The static perspective is based on snapshots of the estimated states of the system at various instances of time. It can, in general, be transformed into the standard linear programming (LP) form with time divided into discrete periods. This problem can be represented in a general form:

Maximize $z = Cx$, (1-1)

Subject to constraints of form $Ax = b$ and $x \geq 0$ (1-2)

The solution to the problem x^p is the plan that is in place and needs to be executed. The elements of the vectors *A*, *b*, and *C* are indicative of the state of the system (define the system) and are called the state parameters. Together, the state parameters and the plan that is generated characterize the system and its state. SLRs are incorporated as constraints.

An event is defined as the cause or a consequence of Δ change in one or more parameters/variables that are related to the initial state of the system. So if an event does occur, it will cause a Δ change in one or more parameters and variables that form a part of the planning problem (Equations (1-1) and (1-2)), i.e., will cause any of the transformations $A \to A', b \to b'$, or $C \to C'$.

This change in the parameter value of A, b, or C further implies that the operating point x^p, i.e., the solution to the planning problem, and/or the optimal value z^p changes and the original values may no longer be achievable or optimal. The new planning problem (replanning problem) can be formulated as:

Maximize $z' = C'y$, (1-3)

Subject to $A'y = b'$ and $y \geq 0$ (1-4)

The solution becomes the new operating plan y^P. The execution systems implement the transition from plan x^p to y^p.

The static perspective to the planning problem however gives only a partial view of the EM problem. Execution is an integral part of EM and the execution systems have dynamics associated with them. From the time persistent perspective, an “event” can be viewed as the cause or the consequence of a mean shift in one or more of the state variables. This generalized idea can be represented as

$$y_t = \mu + \frac{\theta(B)}{\phi(B)} a_{t0} + \gamma f_{E_i}(t) \qquad (1\text{-}5)$$

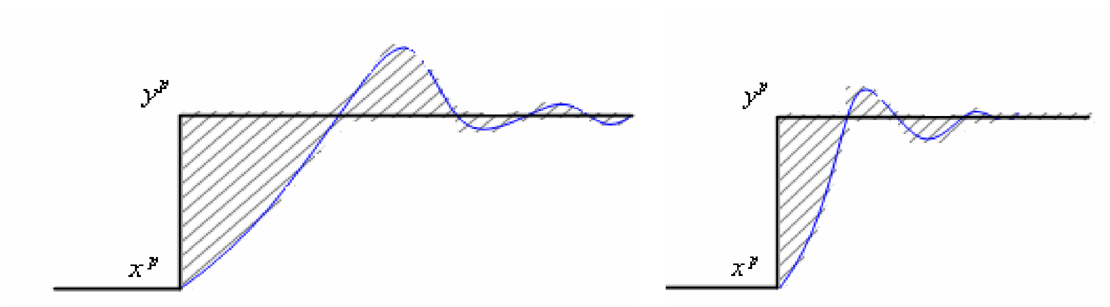
Where $f_{E_i}(t)$ represents the specific nature of the effect that the event E_i and has the form of a step function, the hump, the ramp or the spike functions. For example $f_{E_i}(t)$ will take a form similar to Equation (1-6) if in case a step function is used for its representation.

$$f_{E_i}(t) = \begin{cases} 0 & 0 < t < t_0 \\ 1 & t \geq t_0 \end{cases} \qquad (1\text{-}6)$$

Further, the mean shift in the state estimates cause the system parameters to change and EM has one or more of the following objectives:

1. Regulate a system to remain near a fixed nominal condition in spite of disturbances.
2. Follow a nominal path with minimum error, even though the system parameters have uncertain values.
3. Respond to arbitrary changes in the desired nominal condition with well behaved transients in state and control variables.

The concept of closed loop servo control between p&e for achieving the same has existed in literature since long. For instance, Herb Simon published a paper in 1960 discussing the applications of servo control in business system, and the dynamical analysis of business systems was initiated by Forrester in 1958. The emphasis of these studies and closed loop designs was however mainly on performance and stability analysis. The focus of EM is more on providing higher fidelity to trajectory tracking system.



(a) Low fidelity system (b) High fidelity system

Figure 1-2 State transition by a dynamic system

Of course the fidelity of the system can be improved by making the execution system agile, or by increasing the redundancy of the system. The challenge is to achieve the higher fidelity without making changes in the physical system, and achieving the same by improved algorithms and logistical reorganization.

1.3 Research methodology

Not making changes in the physical system automatically implies that the focus should be on the planning and the replanning system. An implicit assumption made in the aforementioned description of the EM problem is that the state transition that needs to be made, i.e., $x^p \rightarrow y^p$ when an event (or a potential occurrence of an event) is encountered, is known. This however may not always be the case. Planning in supply chains comprises of covering all or most of the activity circles (representative of these hierarchies and spanning over modules) presented in Figure 1-3. It is a complex problem and generally prohibitively expensive to model and solve both computationally as well as time wise. The solution y^p may therefore not be available in the required time, or may not be satisfactory.

Further, a property that is generally observed in these large scale planning systems is their hierarchical organization. The quality of the solution obtained from solving the hierarchical problem depends on the efficacy of the decomposition technique that is used (Yan, et al, 2004). The decomposition provides sub-optimal results, and therefore the scenario provides potential for improvement.

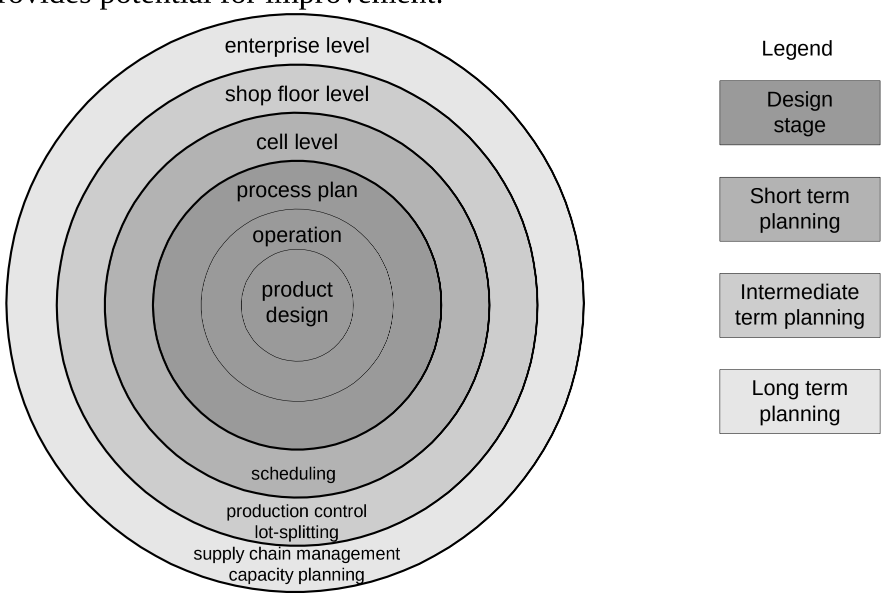


Figure 1-3 Modeling hierarchies as seen via concentric activity circles

The aggregate high fidelity modeling (AHFM) approach is proposed herein that addresses the abovementioned issues. AHFM aggregates the flexibility/redundancies that are available at the various planning levels and then optimally assigns the resource/time from the aggregate pool to deal with various disturbance and problems. It addresses the abovementioned issues by providing the following features:

1. The model spans over multiple levels of the activity circles, thereby does not compromise of the size of the solution space and hence the feasibility.

2. The computational complexity of the solution approach scales linearly or polynomial.

Further, AHFM relies on the parameter variation approach that looks for solutions in the neighborhood of the previous solution so that there is minimum amount of the so called system nervousness when a transition is made from one operating region to the other.

For this research, the formulation and solution of parameter variability has been explored in several domains: manufacturing, air travel and food processing has been considered. Early investigation into these problem domains has yielded remarkable improvements in performance (up to 59% improvement in a discrete manufacturing environment, 84% improvement in process industries and up to 25% improvement in fleet scheduling). For the case study taken up for the IBM supply chain, the indicated improvements are of 27%. While these improvement percentages cannot be generalized for evaluating gains from implementing AHFM, the three examples illustrate that at least in some cases the magnitude of potential profits cannot be ignored. In addition, the conversion of parameters to variables adds flexibility to the models. In fact preliminary results of AHFM to the airline fleet scheduling problem indicate that the technique yields a feasible solution to problems that were considered infeasible.

1.4 Roadmap

The rest of the dissertation is organized in 6 chapters. The AHFM approach is presented in Chapter 2 and the simulation-optimization based approach for solving the AHFM is presented in Chapter 3. Case studies exemplifying the application of the proposed approaches (AHFM) and its extended utilities are presented in Chapter 4. The requirement for cost reduction is also contingent on the early detection of the logistical anomaly. Chapter 5 is focused on addressing the early detection problem, and the integration of sense with response. The insights drawn and the conclusions are presented in Chapter 6. Based on the insights and conclusions, the EM problem characterization is presented in Chapter 7.

The AHFM approach is presented in Chapter 2. Chapter 3 presents a simulation optimization based approach for real time determination of the new operating schedule for the manufacturing unit. A hybrid GA-DATC based algorithm is developed for the further enhancement of the proposed simulation optimization algorithm, and is presented in the addendum section of the chapter. The application domain in the two chapters is discrete manufacturing.

Chapter 4 presents case studies exemplifying the applicability of the proposed approach (AHFM) and the extended utilities of the same. Two sample problems from the process industries and the transportation arena have been selected for demonstrating the applicability of the AHFM approach. The utility of the approach is demonstrated on the cost vs. due date assignment problem in a manufacturing setting.

Issues related to early detection and monitoring are presented in Chapter 5. It is required that the system states be monitored so as to detect any anomaly early. There is a cost associated with the delayed capture of information. There however is a cost associated with the wrong signal as well. The sensing problem is therefore modeled as the economic design of a control chart with the in control time exponentially distributed. The underlying principles behind the economic design of a control chart for a process wherein these two costs coexistent are presented in Chapter 5. Further, issues related to the integration of Sense-Response are also discussed in the chapter.

Conclusions have been presented in Chapter 6. In Chapter 7, events have been defined and classified. The impacts that the events potentially have are also classified. These classifications form the basis for the approaches that are taken for managing the events, and these have been presented in Chapter 7 as well. A direct application of the classification of the events and their impact is in the development of the priority chart used for prioritizing alerts. The development of the priority charts is presented in the addendum to Chapter 7.

1.5 Contributions

This work brings together concepts in SCM, control theory, optimization, and statistical analysis and estimation. The contribution of this dissertation can be divided into two categories: the contributions from the practitioners and the academicians perspectives. For the practitioner, the dissertation provides a framework that can be used to evaluate and select the EM system that he/she needs to install. ROI studies, the promises and the pitfalls can be evaluated before the investments.

The problem characterization provides the framework for theoretical contribution to specific problems picked up from the various sections and chapters. The academicians focused on specific problems can know where to concentrate, and how to do and how their work contributes in the overall picture.

As far as the technicalities of the contributing are concerned, the focus of this work has been on integrating operating parameter selection, maintenance management decisions, routing and production planning decisions to generate additional capacity if need for dampening the effect of parameter uncertainty. Two ideas are proposed:

1. The AHFM approach that aggregates the flexibility/redundancies that are available at the various planning levels and then optimally assigning the resource/time from the aggregate pool to deal with various problems.
2. The parameter variation approach that looks for solutions in the neighborhood of the previous solution so that there is minimum amount of the so called system nervousness when a transition is made from one operating region to the other.

However, the aggregation of the production planning models get more and more complex as we increase the number of hierarchies that the aggregate model covers. It is essential to limit the complexity of the problem with an emphasis on the tractability and computational complexity. One of the main contributions of this work is in the development of algorithms that have a relatively low degree of complexity, and can be solved sufficiently fast to obtain solutions to the AHFM. The algorithms scale more or less linearly, and the neighborhood seeking approach also ensures that the losses due to transience are minimized, and so is the system nervousness.

Both the planning and replanning processes implicitly assume that the information of the state variables and the transition in states are available. For decision making where human is involved, this assumption may be valid, but for automated planning systems, there is a need to set the stage so that this information is made available to the planning systems. “Sensing” ability provides this information and capability to the planning systems. An auxiliary contribution of this work is the economic design of control charts for monitoring and detecting the occurrence of events. Economic models of control charts have been formulated for data driven sensing capabilities in automated systems.

Chapter 2

Responding to an event in a hierarchical setting: the aggregate high fidelity modeling approach[1]

2.1 Introduction

Redundancy and flexibility are seen as the means of achieving reliability to satisfy the SLR in case of abnormal events and shocks. Amongst other forms, it is seen as system reserve capacity in terms of time (e.g., time buffers in scheduling), capacity (e.g., buffer capacity in flexible manufacturing systems), precision (e.g., accuracy in terms of quality), and safety factor (e.g., in design). The key idea is that by providing a margin between the design parameters and their limit states, the system reserve capacity can be used to control the capacity of the overall system to sustain the SLR under bounded uncertainty and disturbances.

From a control theoretic perspective, this approach of introducing reserve capacity for designing the so called robust and survivable systems falls under the broad category of robust control: the goal of which is to design systems that guarantee closed-loop stability under a wide range of uncertainty and disturbances. Traditional approaches to determining redundancy and system reserve capacity however fall under the broad category of "stability robustness". The stability robustness criteria go hand in hand with the hierarchical planning approach that is in vogue in several domains. The planning problem is decomposed, with each subset having a reserve capacity.

The implication of this is that the interaction between various levels of planning becomes less important and the sub-problems can be solved independently. By providing conditions that satisfy the stability robustness criterion at each level, the individual components at the different levels can be seen as functioning in a more or less independent manner. For instance, in manufacturing, typical product design, process planning and production planning and control are usually decomposed by the following functional hierarchy – based on customer specifications designers develop the product; process engineers choose the process plan considering only the local efficiency; based on the chosen process plan, company managers define appropriate capacity and supply chain coordination, plant managers handle customers' orders and choose the batch sizes and cell/department supervisors provide the detailed production schedule. Concurrent engineering (Clark and Wheelwright, 1993, Chang, et al, 1992) exploits the potential of multidisciplinary product design and process planning integration, but production planning, which is responsible for the global system performance, is often out of the

---

[1] Based on:

1. Masin, M., Shaikh, N.I., and Wysk, R.A., (2003), "Aggregative parameter variation in optimization modeling", IEEE Transactions on Automation and Robotics, Aug 2003.
2. Shaikh, N.I., Masin, M., and Wysk, R.A., (2003), "Shop floor planning using a parameter variation modeling approach", IIE Transactions on Design and Manufacturing (under review).

scope of the concurrent engineering process. Current planning models extend over two to three of these levels but a master-slave relationship is usually maintained between the levels, i.e., decisions at one activity level are taken as parameters for the next.

Aggregate High Fidelity Modeling (AHFM) proposed herein falls under the performance robustness criterion and suggests a different, profit-based, decomposition – the parameters at all levels can be modeled as "system variables" and examined for improving the global performance. AHFM integrates planning decisions starting from process planning to the detailed shop floor scheduling to enterprise level production planning. The approach essentially is a new partitioning approach that accentuates the flexibility that is available at each level of planning and integrates the different levels of planning such that the available flexibility is aggregated over various levels.

This approach is an extension of the High Fidelity Modeling (HFM) approach taken at Penn State (Steele, et al, 2001, Son and Wysk, 2001, Masin, et al. 2003, Shaikh, et al. 2004). HFM contains detailed simulations of work-floors, and, given the global system objectives, integrates and optimizes work-floor operations with lower level process control of resources and higher-level production and supply chain planning and control.

2.2 Literature Review

Aggregation rules have been used to integrate some of the hierarchy levels shown in Figure 1-1. Mehra et. al. (1995) integrated a two-level hierarchical problem and showed that the solutions to the monolith problem was within 1.5% of the detailed hierarchical model. They also found that the aggregate models allowed absorption of random events without the need for frequent recomputation. Blending these advantages of the aggregation rules with those offered through HFM for production planning and enterprise integration is the basic idea the AHFM approach. A brief introduction to the state of the art in production planning and aggregate modeling in manufacturing is presented here.

Efficiency production plans are critical for fulfilling product demand. By improving capacity utilization, greater demand fulfillment and higher levels of profits can be achieved at little or no extra cost. Sizing batches effectively, determining efficient operations routing, adjusting process parameters, and scheduling parts to inhibit starvation and blocking fall under the broad scope of production planning. It is a complex problem and generally prohibitively expensive to model and solve both computationally as well as time wise. Decomposition of the monolithic problem using the hierarchical production planning (HPP) approach has been widely used in the production planning literature (Davis and Thompson, 1993). The concept of HPP proposed by Hax and Meal (1975) decomposes the whole planning problem into tractable sub-problems that have been solved sequentially to obtain good overall solutions.

The quality of the solution obtained from solving the hierarchical problem depends on the efficacy of the decomposition technique that is used. For production planning problems, existing hierarchical approaches traditionally employ one of the four decomposition approaches explained by Yan, et al, 2004: (1) The product aggregation approach as demonstrated by Saad, 1988, Simpson, 1999, Simpson & Erenguc, 1998, and Yeh, Tarng,

& Chen, 1988 amongst numerous other, (2) the spatial-temporal decomposition approach which partitions the problem into space and time (Malakooti, 1989; Qiu & Burch, 1997; Nguyen & Dupont, 1993), (3) the process decomposition approach (Villa, 1989), and (4) event frequency decomposition (Akella, 1989; Gershwin, 1988; Kimemia & Gershwin, 1983). Other than the four broad categories mentioned by Yan, partitioning approaches have been developed in other areas, specifically those of Multidisciplinary Design Optimization (MDO). Sobieski and Haftka (1997) provide a comprehensive review of the work done in the area of multidisciplinary aerospace design optimization, some of which, including collaborative optimization (Kroo, et. al., 1994, 1995, Braun and Kroo, 1995), and simultaneous analysis and design (SAND) is highly relevant for production planning as well.

Quite often, the sub problems develop into a master and slave hierarchy where the solutions at one level form the starting point for the next level even when no such restrictions exist in reality. This provides sub-optimal results, and therefore the scenario provides potential for improvement. Pancucci (1999) described the needs for an integration of the shop floor to the management level. Process planning and production planning have been integrated to a certain extent by Kopac and Dolinsek (1992). Larsen (1993) identified concurrent process planning, schedule specific integration architecture and dynamic and distributed planning methodology as some of the techniques for integration. Akturk and Avci (1996) proposed a methodology to determine the optimal machining conditions and tool allocation simultaneously to minimize the production cost of multiple operations, where there are alternate tools for each operation. They concluded that exploiting the interaction between the tool allocations and machining condition optimization decisions could achieve improved solutions and eliminate infeasibilities occurring due to tool contention amongst operations. Further, Akturk and Ozkan (2001) developed a multistage algorithm for solving scheduling problems in a flexible manufacturing system while considering the interrelated sub-problems of processing time control, tool allocation, and machining conditions optimization. The approach recognized the tradeoff in automated manufacturing systems and they reported improved production efficiency and cost reductions.

2.3 The AHFM approach

The traditional approach to production planning applies functional decomposition to planning tasks: first, process engineers locally optimize all process parameters, including the selected process plan and cutting speeds, and, only then, production engineers optimize the production quantities and schedules in order to obtain the maximum profit when all process variables are considered as fixed parameters. The AHFM-based integrative models try to identify variables that may have significant impact on the company profitability and decompose the planning problem accordingly.

The existing HPP approach creates a master and slave hierarchy where the solutions at one level form the starting point for decisions to be made at the next level. In AHFM approach, only “hard” parameters, such as tools and material characteristics are fixed, while most of the other parameters such as cutting speeds, feed rates, and depth of cut are considered for being soft parameters, modeled as global variables in the aggregate model.

Handbook-based parameters are usually considered being “optimal”; however, they are usually based on local operational rules without a global view of long-term profitability.

In order to fully appreciate AHFM approach, there is a need to understand the entire process and identify the operational parameters that are in fact variables. Parameters that directly impact process speed, such as spindle speed or feed in machining, are an obvious choice. However, there are less obvious parameters that impact operational planning models. For example, different sample plans for quality assurance can provide the same level of quality; in semiconductor industry these sample plans have direct impact on the production flow of wafers and, consequentially, on the production control models. The cost versus throughput rate tradeoff of raw material selection in discrete manufacturing, concentration selection in process industries, or universal versus specialized operator selection in services, can be optimally solved only in operational integrative model and not during the design stage as it is usually done.

Local optimization models should also be exploited to set the variation methodology by reducing the set of variable parameters. For example, as will be shown in the next section the cost model for machining operations lead us to setting the feed parameter to its maximum subject to the surface quality constraint and optimizing the spindle speed only. Controlling the set of variable parameters helps to manage the tractability of the integrated optimization models. The following steps are proposed for the implementation of the AHFM approach.

1. **Traditional model**: Formulate the problem using existing modeling procedures for problem optimization.
2. **Visible and hidden parameters**: List all of the parameters used in the model, including hidden ones (e.g., chosen process plan/path for manufacturing each part in the scheduling problem).
3. **Potential for parameter variation**: Identify the “soft” parameters that were fixed in previous design steps. Reduce the set of variable parameters using local optimization models.
4. **Integrative model**: Incorporate all/most of the remaining “soft” parameters in the original model.
5. **Implementation**: Develop an optimization method / heuristic solutions to solve the extended problem.

2.4 Case study: AHFM in manufacturing environment

In this section, the AHFM methodology is illustrated in a manufacturing environment starting from an existing high-level plan, through low-level components, to the integrative high fidelity model. The model will consist of an aggregative production planner for manufacturing based on a real implementation in the FAME lab at the Pennsylvania State University. The layout of shop floor is shown in Figure 1. The system consists of two workstations, each of which consists of Haas CNC machines tended by ABB industrial robots. There are several kinds of Haas CNC machines, such as vertical milling centers (VF-0 and VF-3), a horizontal machining center (HS-1), and turning centers (SL-20 and SL-30), all of which have the capability to manufacture a variety of prismatic and rotational parts with reliability and accuracy. An Arena RT simulation based control system is utilized for computer-integrated control of the work cells. Further details are available at http://www.engr.psu.edu/cim/FAME/index.html.

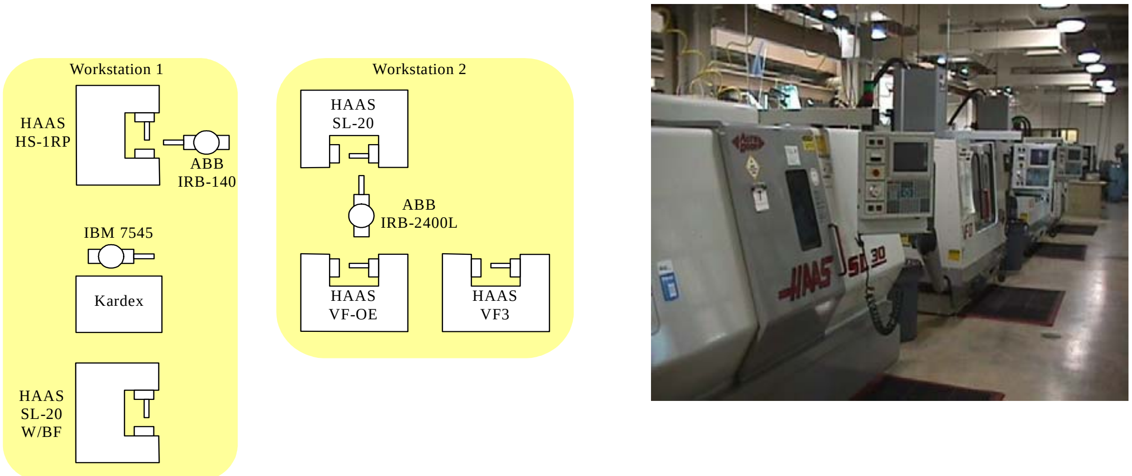


Figure 2-1 Example layout of Penn State CIM Lab

The main products are Penn State-based souvenirs. This work will be focused on one of the souvenirs – three parts from a Penn State chess set, shown in Figure 2-2.

Part KING Part BISHOP Part Pawn

Figure 2-2 Example parts (PSU chess set)

2.4.1 Traditional aggregative production planning

Traditionally, during the *process* planning stage, several process alternatives are checked and then one alternative, considered as the best one, is chosen. Consequentially, the *production* planning assumes a single process plan and known (fixed) processing times for all the parts that are to be manufactured. Let G be the set of part types; Jg be the set of process plans for part type g, $g \in G$; Ij be the set of process steps of process plan j; M be the set of available machines. The complete notation is given in Appendix A. We define t_{ijm} be the effective processing time of part type g, $g \in G$, using process plan j, $j \in J_g$, on machine m, $m \in M$, in the i^{th} processing step, $i \in I_j$; x_j be the number of parts made using process plan j, k_g be the synchronizing coefficient between part types that defines the required production volume ratios, e.g., for part "King" it is 1.0, for part "Bishop" it is 0.5, and for part "Pawn" it is 0.125; t_{jm} be machining time for part using process plan j on machine m. In traditional production planning, $\left|J_g\right| = 1$ for all part types $g \in G$, i.e., there is only a single process plan for each part. The P1 model maximizes the throughput of chess sets, which is limited by the capacity of the individual resources, K_m.

P1: Max $k_1 \sum_{j \in J_1} x_j$ (2-1)

Subject to

$$k_1 \sum_{j \in J_1} x_j = k_g \sum_{j \in J_g} x_j \qquad \forall g \in G \mid g > 1 \tag{2-2}$$

$$\sum_{i \in I_j} t_{ijm} = t_{jm} \qquad \forall g \in G, j \in J_g, m \in M \tag{2-3}$$

$$\sum_{g, j \in J_g} x_j t_{jm} \le K_m \qquad \forall m \in M \tag{2-4}$$

$$x_j \ge 0 \quad \forall g \in G, j \in J_g \tag{2-5}$$

Equation (2-1) defines the objective function of maximum productivity – production quantities of any part can be used since all parts are synchronized through Equation (2-2). Equation (2-2) ensures that while making a chess set, for every "king" produced during a given production schedule, two "bishops", and eight "pawns" are also produced. Equation (2-3) finds the total processing time of each process plan j on each machine m. Equation (2-4) ensures that the production capacities of the individual machines are not exceeded. A schedule to manufacture x_j parts is then generated.

2.4.2 Visible and hidden parameters

Machining time t_{ijm}, synchronizing coefficient k_g and resource capacity K_m are straightforward parameters used in the aggregative production-planning model P1. However, one of hidden assumptions of the model is a given process plan for each part. This process plan should be an additional parameter taken into account.

2.4.3 Potential for parameter variation

The synchronizing coefficient k_g is a part of the product (chess set) specification and, therefore, cannot be varied. Other parameters should be considered for parameter variation.

2.4.3.1 Machining time

Selection of appropriate machining parameters is critical as they determine the effective processing times of the product. These parameters are selected through rule-based, data driven or model-based approach. Model based approach helps determine the operating parameters that correspond to parameters that minimize the production cost per component or minimize the effective production time given in Equations (2-6) and (2-7).

$$C_{ijm} = \frac{C_b}{N_b} + C_m t_{ijm} + C_r \left(\frac{t_{ijm}^{(m)}}{T_r} \right) \tag{2-6}$$

$$t_{ijm} = t_{ijm}^{(m)} + t_{ijm}^{(h)} + t_r^{(t)} \left(\frac{t_{ijm}^{(m)}}{T_r} \right) \tag{2-7}$$

Where C_b is a batch setup cost, N_b refers to number of parts in the batch, C_m is a cost of machining operation on machine m, C_r is a cost of tool r, $t_{ijm}^{(m)}$ is the machining time for the i^{th} processing step of the j^{th} process plan on machine m, $t_{ijm}^{(h)}$ is the material handling time (usually a constant) for the i^{th} processing step of the j^{th} process plan on machine m, $t_r^{(t)}$ is the tool change time (usually a constant) for the r^{th} tool, T_r is the tool life of the r^{th} tool. In our production environment, the same set up time is required for each part in the batch, therefore, C_b is proportional to the batch size, and N_b should be set to one. For the general case, batch sizes should be part of the decision variables as widely discussed in lot-splitting/lot-streaming literature (Benjaafar, 1996, Johnson, 2003). Minimization of the operating costs, C_{ijm}, or effective processing time, t_{ijm}, is taken as the objective and is solved as a constrained minimization problem, where the typical constraints are of the form presented in (2-8)-(2-12) (Chang et al. 1992):

Constraints on the Spindle Speed (depends on the machine):

(a) $v_m^{min} < v_{ijm}$ (2-8)

(b) $v_{ijm} < v_m^{max}$

Feed constraint (depends on the machine):

(a) $f_m^{min} < f_{ijm}$ (2-9)

(b) $f_{ijm} < f_m^{max}$

Cutting force constraint:

$$K_F f_{ijm}^q d_{ijm}^w < F_{ijm}^{max} \tag{2-10}$$

Power constraint:

$$P_{ijm} = F_{ijm} v_{ijm} = K_F f_{ijm}^q d_{ijm}^w v_{ijm} < P_{ijm}^{max} \tag{2-11}$$

Surface finish constraint:

$$K_S f_{ijm}^h D_{ijm}^{-1} < R_{ijm}^{max} \tag{2-12}$$

Where, K_S and h are specific coefficient and exponents of surface roughness constraint; K_F, q, and w are specific coefficient and exponents of cutting force constraint; and D_{ijm} is the tool diameter used for the i^{th} operation of the j^{th} process plan on machine m. The tool diameter is usually given and set to the largest feasible size for the given operation, since, as follows from Equation (2-12), tools with higher values of D_{ijm} can have higher feed rates, f_{ijm}. The feed rate, f_{ijm}, cutting speed, v_{ijm} and the depth of cut, d_{ijm}, determine the machining time, $t_{ijm}^{(m)}$, and the tool life, T_r. The general relationship between the machining time and tool life with the machining conditions is presented in Equations (2-13) and (2-14).

$$t_{ijm}^{(m)} = \frac{K_{ij} L_{ijm} D_{ijm}}{n f_{ijm} v_{ijm}} \tag{2-13}$$

$$T_r = \frac{K_r}{v_{ijm}^{\alpha'} f_{ijm}^{\beta'} d_{ijm}^{\gamma'}} \tag{2-14}$$

where, L_{ijm} is the tool path length for the i^{th} operation of the j^{th} process plan on machine m, K_r is the Taylor's tool life constant for tool r; α', β', and γ' are speed, feed and depth of cut exponents for tool r for operation i, and n is the number of teeth. Solving a mathematical model using Equations (2-6) or (2-7) as the objective and Equations (2-8)-(2-14) as the domain constraints gives the values of the machining parameters v_{ijm}, f_{ijm}, and d_{ijm} for the i^{th} operation of the j^{th} process plan on machine m. In order to reduce the number of cuts the depth of cut d_{ijm} should be fixed at the maximum allowable limit dictated by the tool properties and material properties. Between the two remaining decision variables, the spindle speed, v_{ijm} and the feed parameter, f_{ijm}; the following property essentially reduces the decision space to the spindle speed only.

Property. If the objective is to minimize a function of production time and/or production cost described by Equations (2-6) and (2-7), then in the optimal solution at least one of constraints (2-8)a, (2-9)b, (2-10), and (2-12) is bounding, i.e., either the feed parameter f_{ijm} is set to its maximum value or the spindle speed v_{ijm} is set to its minimum value or both.

Proof. Let the optimal solution spindle speed and feed parameters be v_{ijm} and f_{ijm} respectively. If at least one of constraints (2-8)a, (2-9)b, (2-10), and (2-12) is bounding, the property holds. If none of these constrains is bounding we get a contradiction as follows. Let simultaneously decrease v_{ijm} by $(1+\varepsilon)$ to $v_{ijm}/(1+\varepsilon)$ and increase f_{ijm} by $(1+\varepsilon)$ to $f_{ijm}(1+\varepsilon)$, where ε is a small number. If constraints (2-8)a, (2-9)b, (2-10), and (2-12) are not binding they should continue to be satisfied with the new spindle speed and feed parameter. Constraint (2-11) holds too since $q \le 1$ (Chang et al., 1992). The exponents in the tool life Equation (2-14) have the following relation: $\beta > \alpha$, therefore the new tool life will be longer while, according to Equation (2-12), the machining time is unchanged. Consequentially, the new production time and production cost are better than the original ones – contradiction to the assumption that we have an optimal solution with the spindle speed and feed parameter equal to v_{ijm} and f_{ijm} respectively.

Using typical values for the tool life coefficients and other parameters, f_{ijm} usually is set to the upper bound value based on the surface finish constraint before v_{ijm} reaches to its lower bound. Consequentially, we can substitute Equations (2-13) and (2-14) into Equations (2-6) and (2-7) and obtain the following generic form:

$$C_{ijm} = \frac{a_{ijm}^{(C)}}{v_{ijm}} + b_{ijm}^{(C)} v_{ijm}^{\alpha_{ijm}} + c_{ijm}^{(C)}, \tag{2-15}$$

$$t_{ijm} = \frac{a_{ijm}^{(t)}}{v_{ijm}} + b_{ijm}^{(t)} v_{ijm}^{\alpha_{ijm}} + c_{ijm}^{(t)}, \qquad (2\text{-}16)$$

where, $a^{(C)}$, $b^{(C)}$, $c^{(C)}$, $a^{(t)}$, $b^{(t)}$, $c^{(t)}$ are coefficients defined by part and tool materials, manufacturing process, and machining and tool change costs; their values are found using manufacturing data handbooks and equipment costs. Constraints (2-8)-(2-12) can be reduced to Constraint (2-17):

$$v_{ijm}^{\min} \le v_{ijm} \le v_{ijm}^{\max} \qquad \forall i, j, m \qquad (2\text{-}17)$$

Where, $v_{ijm}^{\min}$ and $v_{ijm}^{\max}$ are the lower and the upper feasible bounds of v_{ijm} based on the above constraints.

2.4.3.2 Alternative process paths/plans

In the traditional approach, like in selecting the machine parameters, a process path/plan that minimizes either processing time or production cost per part is usually chosen. The operations routing summaries (ORS) are specific to the shop floor and take into consideration the machines and the tooling that is available. Multiple ORS are possible for the same part. However, in practice, usually one ORS (typically the one that is most economic) is selected and used to generate the process plan. If the machining time is made a decision variable, multiple ORS could be selected. An alternate ORS that can be used for machining the pieces in the chess set is presented in Masin et al, 2003, Figures B1-B4 and Tables B1-B3. The process plan contains the details of the machining parameters that should be used with the selected ORS.

Given that only Workstation 2 is available for manufacturing the chess set parts and given the available cutting tools, a set of feasible process plans for the three parts and their machining features has been developed (Figures B1-B4 and Tables B1-B3). The nodes at machine level include material handling and machine-level jobs. The OR junction presents alternative operations and the AND junction presents alternative sequences for the necessary operations. The best three process plans are taken into consideration.

For equipment level plans, the alternatives can be generated from different tooling and machining features. For example, as shown in Figure B4 in Appendix B, operations #15, #35, #45, and #55 are plan alternatives with a different tools and the operations #21, #31, #41 and #51 are plan alternatives with different machining features. For our example operation #15 is the alternative for operation #10, where a face cutting tool is replaced from face mill (55° MCLNR) to general the right hand taper (35° DVJNR). Furthermore, the operation #31 is the alternative of #30, where the grooving operation is appended to the right hand finish turning.

2.4.3.3 Machine capacity

In traditional aggregative planning, designers use conservative estimates of the nominal machine capacity based on historic performance. Detailed scheduling tries to make the actual plan as close (i.e., as feasible) as possible to the original aggregative plan. If there is not enough capacity, e.g., due to machines and tools breakdowns, finite buffers and/or inefficient schedule, the production plan is scaled down to meet the available capacity. Preventive maintenance, buffer design and production scheduling may directly affect the effective machine capacity. However, in the relatively simple manufacturing facility used in this example, the effective variability of processing times is the main reason why the nominal capacity may be overestimated.

2.4.4 Building integrative models

The main theme of the integrative models is putting together the profit objective and decision variables identified in the previous step.

2.4.4.1 Integration of machining time and alternative process paths/plans

HPP usually takes a bottom up approach for solving the planning problem, i.e., ORS are selected first, then the processing parameters are determined, production goal are established, and finally production schedules are generated. However, AHFM approach begins with the P1 model (top down) and integrates the ORS and process parameter selection models with it. The soft, i.e., variable, parameters and constraints in the P1 model are identified and replaced by their representative variables. The integrated model is then solved to generate optimal combination of ORS, process parameters, and process goals that maximize the profits. The traditional P1 model find the production quantities x_j when the processing times t_{ijm} and process plan j for part type g are fixed. Model P1 can be modified into Model P2 according to AHFM approach by combining Equations (2-15) and (2-16) with the P1 model:

P2. Max $pk_1 \sum_{j \in J_1} x_j - \sum_{g,\, j \in J_g,\, i \in I_j,\, m} x_j \left(\frac{a_{ijm}^{(C)}}{v_{ijm}} + b_{ijm}^{(C)} v_{ijm}^{\alpha_{ijm}} + c_{ijm}^{(C)} \right)$ (2-18)

Subject to

$$k_1 \sum_{j \in J_1} x_j = k_g \sum_{j \in J_g} x_j \qquad \forall\, g \in G \mid g > 1 \tag{2-19}$$

$$\sum_{g,\, j \in J_g,\, i \in I_j} x_j \left(\frac{a_{ijm}^{(t)}}{v_{ijm}} + b_{ijm}^{(t)} v_{ijm}^{\alpha_{ijm}} + c_{ijm}^{(t)} \right) \le K_m \qquad \forall\, m \in M \tag{2-20}$$

Constraint (17) defines the feasible range of spindle speed for each operation

$$x_j \ge 0 \quad \forall g \in G,\, j \in J_g \tag{2-21}$$

where, p is the selling price of the chess set. In AHFM, the cutting parameter selection is done during production planning providing an opportunity to improve the system's global performance.

2.4.4.2 Integration of machine capacity

We suggest an iterative procedure for capacity adjustment. In the iterative procedure, detailed simulation-based scheduling checks the feasibility of the solution of model P2 using traditional heuristics. If there is no feasible solution, the capacity is reduced and model P2 is solved again. In order to be effective, the resulting solution should differ from simply scaling down the number of units in the original production plan. In order to be efficient, the procedure should converge fast. It is found that the iteration procedure is efficient – just one iterations is sufficient, but not effective – the resulting production plan is completely equivalent to scaling down of the original one.

2.5 Implementation

The efficacy of the AHFM approach is displayed through a sample problem of producing a chess set containing three parts: King, Bishop and Pawn. For part King, the machining operations consist of facing, turning, grooving, cut-off, and milling. In particular, the part needs to be moved from the Haas SL-20 to VF-OE or VF-3 after the initial turning operation. During the milling operations of clock dimples and tower slots, fixture rotations are necessary to machine the four sides of the part. For part Bishop, most of the operations are similar to those of part King except there is a milling operation necessary to be performed for laces. For part Pawn, several milling operations are needed to machine the 5 paw dimples and a heel pad on the top.

Once the processing steps in the alternative routes in P2 have been determined, the *a, b* and *c* coefficients for Equations (2-15) and (2-16) are determined. The parameters that are used for determining *a, b* and *c* depend on the tool material, work piece material, manufacturing process and machining and tool change costs, and have been obtained from the data handbooks. The *a, b* and *c* parameters used for a set of process plans and machining costs are presented in Appendix B. Four cases were considered for analysis.

1. **Case 1 – Traditional**: When the production plan is determined based on machining parameters that are considered fixed and a single process plan per part is used. In this case, decisions made during the process planning stage are propagated to production planning stage as fixed parameters. Given the processing times and paths, the production planning module finds the maximum production quantities using P1 model. Given fixed sales prices and unconstrained demand, P1 model is equivalent to profit maximization.
2. **Case 2 – Variable machining time**: When the production plan allows variable machining times, but only one process plan per part is used. In this case, only process paths are fixed during the process planning stage while the cutting speeds (and resulting processing times and costs) are determined by the P2 model during the production planning stage.
3. **Case 3 – Variable process plan**: When the production plan allows multiple process plans per part, but machining parameters are considered as fixed. In this case, during the process planning stage, a set of possible process plans is developed with fixed cutting speeds in each. How much each process plan is used is determined by the P2 model during the production planning stage.
4. **Case 4 – Variable process plan and machining time**: When the production plan allows multiple process plans per part and allows variable machining times as well. In this case, during the process planning stage, a set of possible process plans is developed. All other process decisions, i.e., how much each process plan is used and what are the cutting speeds, are done by the P2 model during the production planning stage.

For determining the actual throughput and profitability, detailed schedules have to be generated and the actual throughput needs to be determined. Popular heuristics are incorporated into the simulation model to generate feasible schedules wherever possible.

The original production quantities were scaled down if they did not meet the available machines' capacity. The results from the simulation were used to generate the operating schedule.

2.6 Results and discussion

CONOPT with GAMS interface has been for optimization, Arena 7 for simulation and MINITAB 14 for statistical analysis (Brooke et al. 1998, Kelton et al. 2001, Carver 2003). The following results are obtained for each case.

Case 1: The maximum profits that can be generated when different handbook values of spindle speeds are selected after a single process plan has been selected a priori are presented in Figure 2-3(a). As can be seen, the profitability can considerably deteriorate if bad machining parameters are chosen. During the process planning stage, usually two spindle velocities are considered – the slow spindle speed with lowest cost per part and the fastest allowed spindle speed. The both reduce the potential profitability by tens of percent!

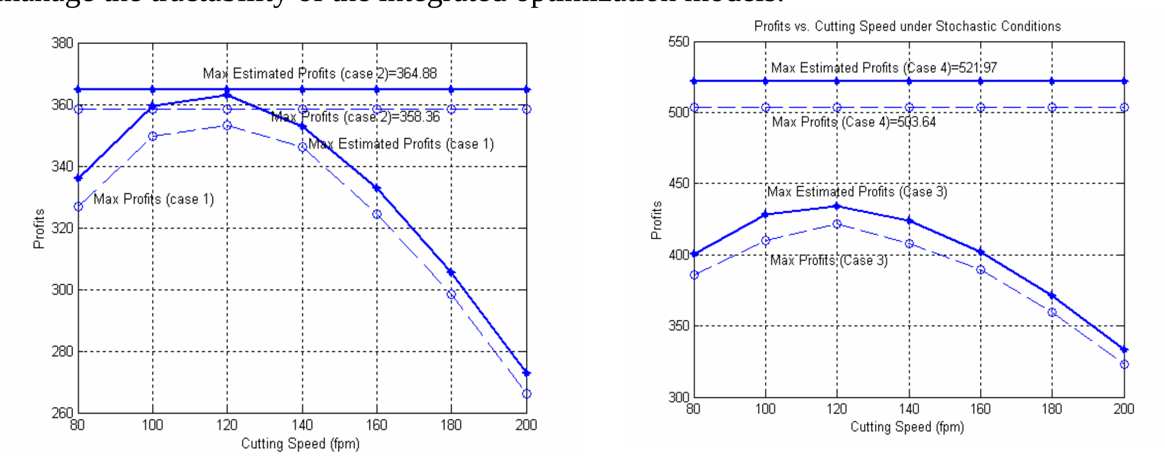
Case 2: When machining parameters are varied, the maximum achievable profit is $\hat{Z} = 364.85$ and $\hat{x}$ =94. Simulation was then used to determine the operating schedule and the actual machining capacity under shop floor conditions. The actual number of parts that can be completed is x =92, thereby generating actual profits of $Z = 358.36$. These have been presented in Figure 2-3(a) as a horizontal line. Even though the 1.5% difference between profit in this case and the best profit in the previous case is relatively small, moving the cutting speed decision into the production planning stage eliminates the possibility of choosing bad speed that can cost tens of percent.

Case 3: The maximum profits that can be generated when different handbook values of spindle speeds are selected (multiple process plans) are presented in Figure 2-3(b). The cutting speeds were fixed at 8 levels in the operating region to determine the profitability that can potentially be achieved for those fixed operating parameters. These estimates are made under conditions of zero machine starvation or blocking. Actual profits however are slightly lower than these estimated values and are determined by simulating the shop condition (see Figure 2-3(b)). As shown in Table 1, the profits in Case 3 dominate the profits in Case 1 by about 15% since a more balanced utilization of capacity is possible.

Case 4: When multiple process plans are allowed and the machining parameters are varied, $\hat{Z}$ =521.97 and $\hat{x}$ =133. The results of the simulation and the comparison of the actual profits with the estimated profits for both fixed and variable parameters are presented in Figure 2-3(b). The profit obtained in Case 4 is better than the highest profit obtained in traditional Case 1 by more than 30%, that is much higher that just sum of the improvement of Cases 2 and 3. We should notice that during the process planning stage, the cutting speed that optimizes the profitability in Case 1 is not known, so the actual improvement from Case 1 to Case 4 can be even higher. Variation of cutting speed becomes more important when production floor is more complex with more execution alternatives. This leads us to a conjecture that large-scale system would benefit more from the AHFM approach since (a) they have more opportunities for alternative process plans, and (b) local optimization in large-scale systems would lead to higher imbalance and loss of production capacity.

Overall, adjusting hierarchical decision-making regarding process plans and machining parameters has resulted in a 30% increase in the profitability of the shop floor operation chosen for the example across Case 1 to Case 4. Actual increase in profitability depends on actual parameters, but AHFM production plans *always* dominant the traditional ones and at least in *some* cases the improvement is measured in tens of percent without any investment required.

Local optimization models should be exploited to set the variation methodology by reducing the set of variable parameters without affecting the optimal cost or profit. For example, the cost model for machining operations lead us to setting the feed parameter to its maximum subject to the surface quality, cutting force and power constraints and optimizing the spindle speed only. Controlling the set of variable parameters helps to manage the tractability of the integrated optimization models.



(a) Case1 and Case 2 (b) Case 3 and Case 4

Figure 2-3 Profits vs. Cutting Speeds (a) For Case1 and Case 2, and (b) For Case 3 and Case 4

Table 2-1. Combined Results

Cutting Speed	Case 1			Case 3		
	Math Model Exp. Profits	Actual Capacity	Actual Profits	Math Model Exp. Profits	Actual Capacity	Actual Profits
80	335.93	0.9732	326.94	400.275	0.9636	385.72
100	359.31	0.9729	349.58	428.27	0.9574	410.03
120	363.04	0.9731	**353.28**	434.09	0.9719	**421.90**
140	352.89	0.9811	346.25	424.07	0.9608	408.06
160	332.75	0.9752	324.52	402.27	0.9677	389.29
180	305.34	0.9768	298.26	371.41	0.9687	359.80
200	272.64	0.9768	266.31	333.375	0.9694	323.19
	Case 2			Case 4		
Var	364.88	0.9821	**358.36**	521.97	0.9648	**503.64**

2.7 Conclusions

In this chapter, the AHFM approach has been presented that can be used for integrated planning in a variety of industries. The illustration is with a small set of problems, that the model formulation, although not generic, can be easily modified to accommodate a variety of system specifics. The problems that have been illustrated reflect the actual characteristic of such systems, but the size of the problems was reduced for illustration. Scaling the technique will undoubtedly hit an impasse where optimal problem solvers will no longer function adequately. A major part of the focus of this chapter was to uncover the potential savings that may be gleaned from such a technique. Because the opportunity (increased profitability from 25 – 85% for realistic small scale problems), there is a strong indication that heuristics for large-scale problems should be developed as the scale exceeds the capacity of today's solvers. The potential savings for integrating the planning domain are too large to ignore this critical aspect of the problem. The results indicate that local optimum can produce bad solution to the larger problem and that the integrative planning is essential for efficient systems operation.

Chapter 3

The simulation optimization approach to solving the aggregate high fidelity model[2]

3.1 Introduction

Combining the benefits of a global view and high fidelity modeling can significantly improve the overall system performance. Higher fidelity lends flexibility to the models, yielding feasible solutions to some problems that were initially considered infeasible by hierarchal planning procedure. However, the complexity of the problem and the solution strategy increases as the fidelity of the modeling approach increases. It is important that the responses be evaluated fast so that the cost incurred when the system is out-of control is limited.

In this chapter, the structure of AHFM is studied and a spatio-temporal partition heuristic is proposed wherein mathematical modeling is used for taking global decisions, selecting process plans and process parameters, and generating strict bounds on the solution. Simulation is used for developing detailed model of the shop floor and its control, taking time dependent decisions, determining actual machine capacities, and generating schedules. The motivation behind such a partitioning is to ensure that appropriate tools are used to tackle appropriate aspects of the production planning problem. Mathematical modeling is efficient for static scenarios and simulation-based approaches appropriate to study dynamic or time dependent issues and interactions. The hybrid model calls the mathematical model and the simulation model alternately, the output of one taken as the input for the other till there is a convergence, or the desired level of productivity has been reached.

The work is organized as follows. The next section describes the recent works regarding aggregate modeling in production planning, hybrid modeling focusing on simulation optimization, and simulation based scheduling. Section 3.3, reviews the composite modeling approach and the structure of the problem. Section 3.4 shows the proposed implementation strategy. Section 3.5 gives a case study. The scalability of the proposed AHFM approach and the simulation-optimization based solution strategy is studied in Section 3.6. Results are presented in Section 3.7.

---

[2] Based on:

1. Shaikh, N.I., Masin, M., and Wysk, R.A., (2003), “Shop floor planning using a parameter variation modeling approach”, IIE Transactions on Design and Manufacturing (under review).
2. Shaikh, N.I. and Prabhu, V.V., Reed, P. (2003), “Hybridized Arrival Time Control Approach to Job-Shop Scheduling”, in the Genetic and Evolutionary Computation Conference (GECCO) Late Breaking Papers, Erick Cantu-Paz (ed.), Chicago, 2003.

3.2 Literature Review

Scheduling involves allotment of resources to competing entities over time, so as to satisfy one or more objectives. Within the specific domain of manufacturing management, scheduling involves assigning jobs to machines or vice versa and timetabling the operations dictated by the process plans (Rodammer and White, 1989). Over the last thirty-five years, researchers and practitioners have approached the scheduling problems that come up, specifically when resources are scarce. Combinatorial optimization techniques (e.g., Baker, 1974; Pinedo, M., 1995; Morton and Pentico, 1993; etc), control theoretic approaches (e.g., Gershwin, 1989; Prabhu, V., 1998; Kogan and Khmelnitsky, 2001; etc), and artificial intelligence (AI) based approaches (Smith, S.F., 1991; Brown, D.E., and Scherer, W.T., 1995; Bagchi, T.P., 1999; etc) have been proposed. The focus of most of the researchers has been on determining the complexity of the problems, development of efficient solution strategies for special cases, and the development and analysis of heuristic methods for the solution of more intractable cases (Ovacik, I.M., and Uzsoy, R, 1997).

Implementation of these scheduling approaches in industry is, however, not that widespread. Theoretical approaches focus on complexity analysis as well as rigorous analysis of exact procedures or heuristics for mathematically tractable special cases of the real problems (Ovacik and Uzsoy, 1997). Most large-scale, real-time optimal scheduling problems do not fall into these special categories. Large and complex search spaces for solutions, dynamic shop floor conditions, and a multitude of shop floor conditions are the factors that set these problems apart. These cannot, in practice, be solved by traditional optimization techniques such as dynamic and integer programming, AI techniques such as genetic algorithms, or control based techniques relying on feedback (Montana, et at, 1998).

In recent years, simulation tools have also proven valuable for the prediction of machining state variables over a wide range of operating parameters. Such simulation packages, however, are seldom an integral part of machining parameter optimization modules. Stori (1999), proposed a methodology for incorporating simulation feedback to fine-tune analytic models during the optimization process. Thus, through a limited number of calls to the computationally expensive simulation tools, process parameters may be generated that satisfy the design constraints within the accuracy of the simulation predictions, while providing an efficient balance among parameters arising from the functional form of the optimization model. The formulation of realistic models and their analytical evaluation (Stori, et al, 1999) is now more or less a reality. Simulation models provide the desired realism and can be evaluated numerically over a time period of interest (Sims, M.J, 1997). Simulation based scheduling is fast emerging as a powerful technique for timetabling events under complex shop floor scenarios.

3.3 Composite Modeling

The composite model used for determining resource utilization and production levels that would maximize the profitability is (P3). The objective is to maximize the profits at the enterprise level.

$$\text{(P3): Max Z} = \sum_{g=1}^{G}(C_g \sum_{i=1}^{I}\delta_{gi}) - \sum_{g=1}^{G}\sum_{i=1}^{I}\delta_{gi}(\sum_{j=1}^{J}C_{gij}) \tag{3-1}$$

$$S_{g1}\sum_{i=1}^{I}\delta_{(g1)i} = S_{g2}\sum_{i=1}^{I}\delta_{(g2)i} \qquad \forall\ g1 \neq g2 \tag{3-2}$$

$$v_{gj}^{\min} < v_{gj} < v_{gj}^{\max} \qquad \forall\ g, i, j \tag{3-3}$$

$$C_{gij} = \frac{a_{gj}^{(C)}}{v_{gj}} + b_{gj}^{(C)} v_{gj}^{\alpha_{ijm}} + c_{gj}^{(C)} \qquad \forall\ g, i, j \tag{3-4}$$

$$P_{gij} = \frac{a_{gj}^{(t)}}{v_{gj}} + b_{gj}^{(t)} v_{gj}^{\alpha_{ijm}} + c_{gj}^{(t)} \qquad \forall\ g, i, j \tag{3-5}$$

$$\sum_{k=1}^{TC} X_{gijk} = \delta_{gi}\alpha_{gj} \qquad \forall\ g, i, j \tag{3-6}$$

$$\sum_{k=1}^{TC} kX_{gijk} + P_{gi(j+1)} - (1-\delta_{gi}\alpha_{g(j+1)})M \leq \sum_{k=1}^{TC} kX_{gi(j+1)k}$$
$$\forall\ \text{g, i and j} = 1,\ldots,\text{J - 1} \tag{3-7}$$

$$\sum_{g=1}^{G}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{gijk}\delta_{gi}\alpha_{gj}Y_{gjm} + \sum_{g=1}^{G}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k'=k+1}^{\min(TC,k+P^{\max}-1)} X_{gijk'}\delta_{gi}\alpha_{gj}Y_{gjm}\xi_{gj(k'-k)} \leq 1$$
$$\forall\ \text{m, k} \tag{3-8}$$

make the production planning model combinatorial and NP hard. It is difficult to solve a small size problem in reasonable time frame using this approach. Researchers have proposed relaxation and formulation of the Lagrangian dual for models similar the P3. (P3) can therefore be reformulated as (P4)

$$\text{(P2)}: \text{Max Z} = \sum_{g=1}^{G}\text{C}_\text{g}(\sum_{i=1}^{I}\delta_{gi}) - \sum_{g=1}^{G}\sum_{i=1}^{I}\delta_{gi}(\sum_{j=1}^{J}C_{gij}) -$$
$$\sum_{k=1}^{TC}\sum_{m=1}^{Mc}\lambda_{km}\left(\sum_{g=1}^{G}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{gijk}\delta_{gi}\alpha_{gj}Y_{gjm} + \sum_{g=1}^{G}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k'=k+1}^{\min(TC,k+P^{\max}-1)} X_{gijk'}\delta_{gi}\alpha_{gj}Y_{gjm}\xi_{gj(k'-k)} - 1\right) \tag{3-9}$$
$$\forall\ \text{m, k}$$

Subject to the (3-2), (3-3), (3-4), (3-5), and (3-6)

The dual problem is solved using sub-gradient optimization algorithm in a smaller time frame and the solutions are within 4% for small to medium size problems for scheduling problems (Ventura and Radhakrishnan, 2002).

3.4 Spatio-Temporal Partitioning and Hybrid Algorithm

Mathematical modeling for scheduling involves discretisation of time into small steps and making the processing times integral multiple of the time step. This causes a large increase in the number of variables. If search enumeration techniques are used, it becomes necessary to use heuristics that restrict the search space. Lagrangian relaxation techniques are usually used to generate tight upper bounds in the search space and sub-gradient optimization techniques are applied to come up with a good solution. Sub-gradient optimization technique is an iterative procedure wherein the relaxed problem is solved for fixed values of the Lagrangian multiplier to generate a convex hull for maximization problems. The lower bound is updated after every iteration. However, even these techniques become difficult to apply when the number of constraints increase, or when the problem cannot be partitioned efficiently, or when the size of the problem increases. Besides this, the solutions that are generated using Lagrangian relaxation techniques are infeasible unless optimal solution is reached. Summarizing, the problems encountered when scheduling is incorporate in the composite models are:

- Discretization of time and variables.
- Large increase in number of variables and the search space.
- Problem becomes combinatorial if search enumeration techniques are used.
- Solution strategies have to rely on techniques such as Lagrangian relaxation, which generate tighter bounds for limiting search space, however critical constraints (machine capacity) are relaxed.
- Feasible solution is not guaranteed in the neighborhood of the solution generated by Lagrangian relaxation technique.

The hybrid algorithm proposed in this work aims at solving some of the abovementioned problems through a spatio-temporal partition of the composite model as against structural partitioning involving Lagrangian relaxation, etc. The model is explicitly partitioned into static and a dynamic part. The static part can be seen as the relaxed model that is used to generate the process parameters and determine lot sizes constrained by the processing capabilities generated by the dynamic model. The proposed algorithm for the hybrid model is presented in Figure 3-1 and the algorithm stages and steps are explained next. This technique generates a feasible solution after each iteration, and the process is stopped when the goals of the decision-maker (DM) are satisfied.

3.4.1 Stage 1: Capacity adjustment

The overall objective of maximizing the profitability depends on Equations (3-3), (3-4), and (3-5) for determining the process plans and the process parameters. The contribution Equations (3-6), (3-7), and (3-8) make in this objective is to enforce sequencing and limit lot sizes to within the machining capacity. The static model can therefore be formulated as (P5). (P5) is time independent and determines the process plans, processing time, and machining parameters that maximize the global objective. All these parameters do not depend on the interaction between parts.

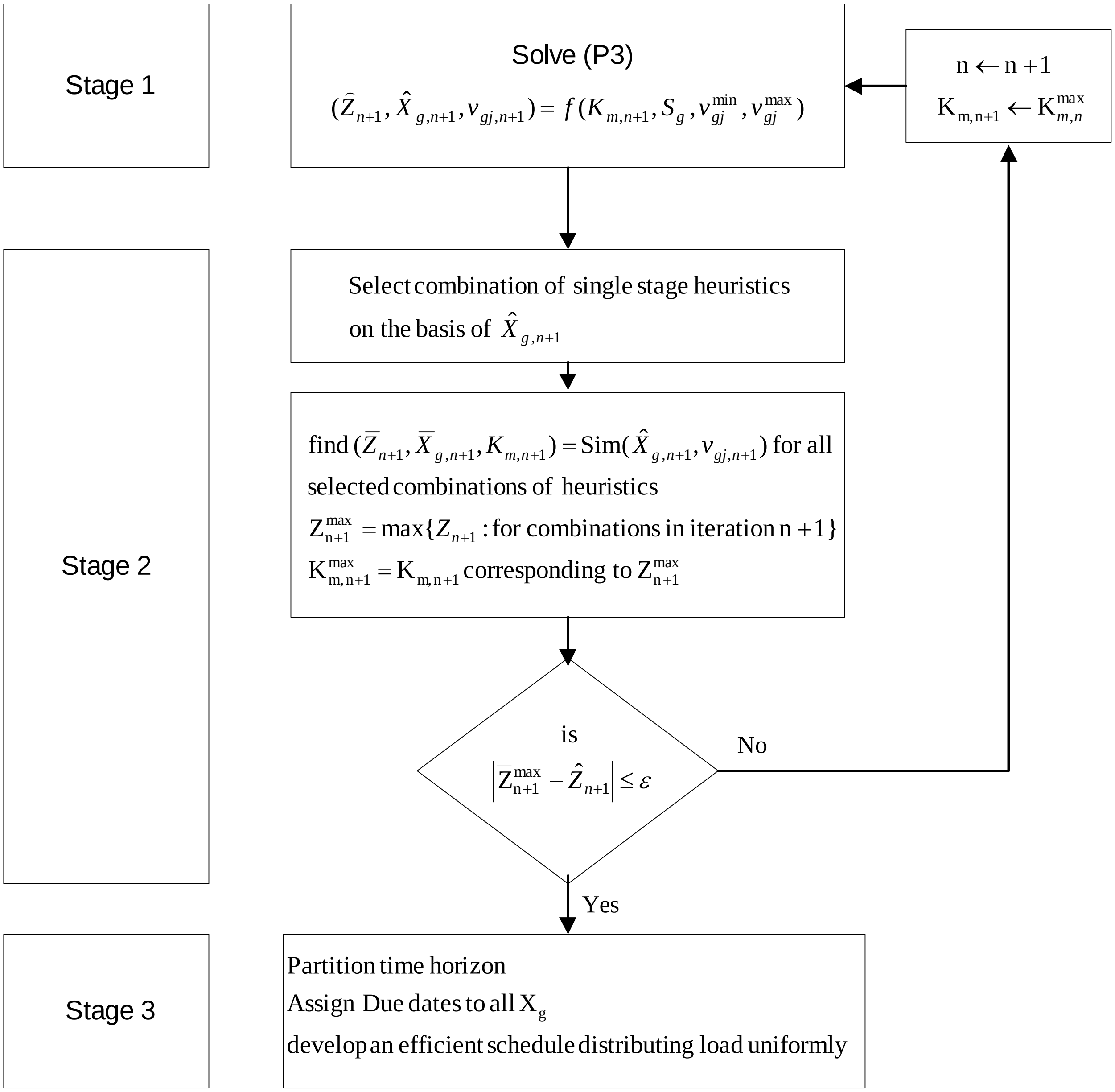


Figure 3-1 Spatio-temporal partition algorithm for composite high fidelity production planning

$$\text{(P5): Max } C_g \sum_g x_g - \sum_{g,j} x_g \left(\frac{a_{gj}^{(C)}}{v_{gj}} + b_{gj}^{(C)} v_{gj}^{\alpha_{ijm}} + c_{gj}^{(C)} \right) \qquad (3\text{-}10)$$

$$\sum_g x_g \left(\sum_{j=1}^{J} Y_{gjm} \left(\frac{a_{gj}^{(t)}}{v_{gj}} + b_{gj}^{(t)} v_{gj}^{\alpha_{ijm}} + c_{gj}^{(t)} \right) \right) \le K_m \qquad \forall\, m \qquad (3\text{-}11)$$

$$S_{g1} x_{g1} = S_{g2} x_{g2} \qquad \forall\, g1 \ne g2 \qquad (3\text{-}12)$$

The solution to (P5) is independent of sequencing constraints, and the solution therefore is the upper bound on the profitability of the shop floor for the machining capacity. Mathematical modeling is used in this case as the constraints and the objective function are quasi-convex and non-linear programming tools can converge to a good solution quickly.

3.4.2 Stage 2: Detailed Scheduling

Simulation is used to generate a dynamic model that takes as input the process plans, the processing times, and the lot sizes from the static model and uses the information to determine the actual processing capability of the shop floor. A detailed shop floor model is created to represent the actual shop floor, and the constraints therein. The parameters generated by (P5) model are taken as inputs and the simulation is executed to determine the minimum shop floor capacity that is required to complete the lots request. If the required shop floor capacity for satisfying the demands set by (P5) is higher than the capacity that is available (due to interactions, starvation at bottleneck, constraints due to material handler, etc), the machining capacity in the (P5) model is reduced, and a new request is generated by the (P5) model. The process is repeated till the actual shop floor capacity is greater than or equal (within an epsilon) the shop floor capacity in the (P5) model and all the lots can be completed within the required time span.

Simulation is selected as the tool for the dynamic modeling because of the following advantages:

1. It is an event driven system. The accuracy of the model does not depend on the size of the time step that is selected, and the processing times need not be integral multiples of the smallest time step.
2. Equations (3-6), (3-7), (3-8), that add complexity to the mathematical models are implicitly assumed to hold true in the simulation models. Equation (3-8) constraints the machine utilization to be below 1. It is either 0 or 1 in simulation. Similarly, sequences are implicitly followed if defined in the simulation model.
3. Simulation allows building of complex models that are representative of the actual conditions on the shop floor. An advantage of the simulation approach is that it can model the effects of such factors as policy changes, interaction between machines and transporters, human handlers, etc, which cannot be accounted for in analytical model (Werner, et al, 2001; Sims, M.J., 1997; Sun, Q., 2000; Lee, Y.H., and Kim, S.H., 2000; etc).
4. It provides an ideal platform for implementing single stage heuristics that are known to be optimal for special cases. It also allows combination of different heuristics to make complex combinations to represent different scenarios, i.e., a “what if” analysis (Fu, M.C., 2000; Law, A.M., and McComas, M. G., 2000; Ferris, et al, 2000; Oey, K., and Mason, S.J, 2001; Boesel, et al, 2001; etc).
5. Allows incorporation of the local knowledge of the shop floor manager with ease. It can also provide the user with an opportunity to perform exploratory tests upon the schedules being produced, can review the status of the shop floor model that is dynamically displayed, and can use his knowledge and practical experience to stop,

interact with the model, and try alternate scheduling approaches (Biswas, S., and Merchawi, S, 2000).

Stage 2 uses heuristics based sequencing and scheduling using simulations to assign priorities to jobs entering the system. As these dispatching rules are the only smarts that are introduced into the simulation model along with the details of the shop floor, Sims (1997) argues that defining the rules that are used to assign work to the resources in the model are critical. Once the heuristics are defined, the paths taken by the parts, their time in system, etc, is fixed (for deterministic simulation). Selecting the optimum combination of heuristics is not the focus of this work. It is assumes that a few heuristics can be used, and a combination of these is considered.

3.4.3 Stage 3: Generation of Efficient Plans

At the end of Stage 2, a feasible production plan is available. The simulation model keeps track of the entity movement and records when a part undergoes what operation. However, this schedule is not efficient. The WIP is high, time that the part spends in the system is high, and if there is any failure in the system, there will be a large number of half-finished parts. Efficient schedules are generated in Stage 3. In Stage 3, the single stage heuristics at each machine are fixed. This implies that when a part enters the system, the parts that are already in the system determine its time in system and sequence. Control based architecture is then used to determine the arrival time that would make the schedule efficient, i.e., lower WIP, time in system, and spread the work load uniformly over the time horizon while keeping the throughput at the level determined at the end of Stage 2.

The underlying concept behind control theory based simulation is its loop structure and the feedback from within, i.e., the notation of circular causality. Figure 3-2 displays the concept, the details of which are explained later. The feedback loops have either positive or negative polarities, indicating whether the loop has the tendency to reinforce or to counterbalance a change in one or more of its loop elements (Scholl, H.J., 1999).

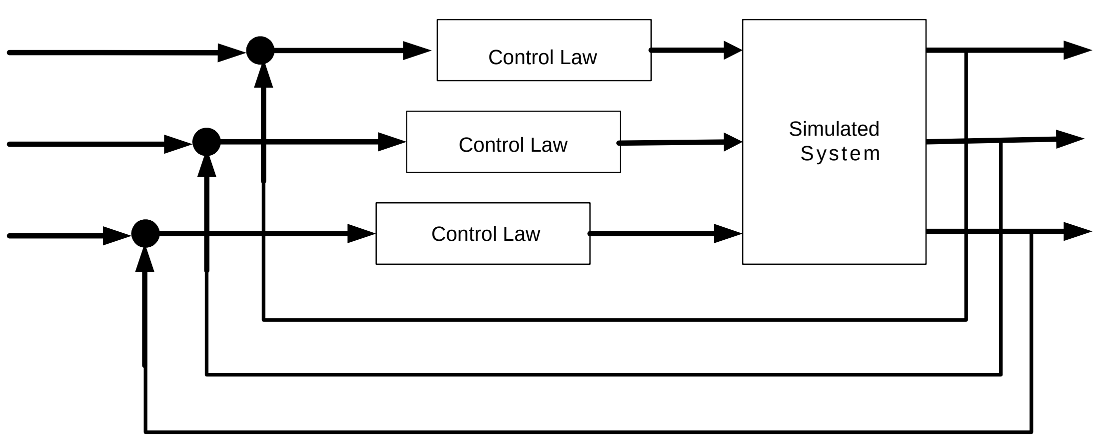


Figure 3-2. Arrival Time Controller for CTBSS (from Shaikh, N.I., 2001)

Arrival time control (ATC) is an algorithm that treats the timing of discrete events as a continuous variable and adjusts the arrival times using feedback from simulation. Since a set point is required for a controller to operate, and time is the continuous variable, due date based events are usually scheduled by ATC. A deterministic simulation of the real processing system is used to obtain the predicted completion time. Using the deviation of the predicted completion time from the due date, the arrival time of the entity into the system is adjusted to reduce the deviation. If there are n-orders, the resulting system will be an n-dimensional multivariable control system where due date, completion time, due date deviation, and arrival times are n-dimensional vectors which act as the command, output, error and manipulated vectors respectively (Shaikh, N.I.,2001). The schedule with the best performance discovered in these iterations is saved and used to schedule events

It has been shown that when the integral control law is used for the arrival time controller, then the dynamics of the system in terms of arrival time convergence is predictable and bounded. It has also been shown that when ATC is applied to a broad class of processing systems, it out-performs common dispatching rules (Prabhu, V.V., and Duffie, N.,1996a; Prabhu, V.V., and Duffie, N.,1996b; Prabhu, V.V., 2000)

3.5 Case Study

In this section we illustrate the methodology of composite production planning for manufacturing based on a real implementation in the FAME lab at the Pennsylvania State University.

3.5.1 Implementation Details

Some of the implementation issues encountered are as follows:

3.5.1.1 Stage 1

P5 model is developed and solved using GAMS. The model, though non-linear, is convex or quasi-convex at least for some ranges of cutting speeds. Standard packages such as CONOPT, MINOS, and SNOPT can be applied for non-linear optimization. It takes as input the sequences that are possible for a given part, machine availability and capacity, and the parameters for determining the process parameters (a, b, and c). The outputs of the model are the lot sizes that need to be made by the various process plans and the processing times for the parts in the machines that maximize the profits at the enterprise level.

3.5.1.2 Stage 2

A detailed simulation model was developed for a cluster of three machines: SL20, VF3, and VFOE. It was assumed that the shop floor is a job shop and any sequence involving the three machines is allowable. Some of the other assumptions that are made in the model are as follows:

- The setup times are included in the processing times.
- A part recaptures a machine immediately after a processing step if it requires the same machine. This assumption allows grouping of processing times at a single machine for operations that are in sequence. Because of this assumption, the processing times of individual steps in not required for scheduling and the sum of

all processing times in a group is directly taken as the output from the GAMS model.
- Transportation time between machines was assumed to be 0.
- The system was modeled as a deterministic system and different heuristics were introduced as queue selection rules.

Initially, to determine the machining capacity, it is required that the throughput be maximized. SPT queue selection rule was used at each machine so as to ensure this. Different queue selection rules were also tried, however the completion time with other rules was much higher. Once the parts are created, they are sent to the machine according to the sequence they are supposed to follow (each part has a predefined sequence). The block diagram of the logic at each machine is presented in.

After the required number of parts have been created and sent to the required sequence, the parts either get processed, or wait in the system. At this stage, there are interactions between parts, and it is this interaction that causes actual machining capacity to be below the theoretical values. Machines may be blocked or starved at particular time instances due to the processing sequences.

The machine capacity in this model is in terms of time for which it is available. For a planning horizon for 8 hrs, a machine capacity of 480 in the (P5) model implies that it is available for 480 minutes in 8 hrs, i.e., for 100%. If there is starvation in the system and the schedule required to complete the lot size determined by the (P5) model is tight, the simulation run length will exceed the maximum allowable time span (in this case, 480).

If the simulations run length exceeds the time span, the iterative procedure mentioned in Figure 3-1 is then employed to decrease the machining capacity in the (P5) model. The (P5) model is executed again and a new set of optimal processing parameters, lot sizes and sequences are derived for the decreased machining parameters. "Results.txt" file is updated and the simulation model is rerun. This process is repeated till the simulation run time is smaller than or equal to the time span of the planning horizon.

At the end of the iterative procedure, the file "Start.txt" contains a feasible schedule. The file "Out.txt" has data of the profitability that can be achieved using the above procedure. At this point, it can be inferered that the actual machine availability of the machines are only for approximately 82% of the time. For the rest of the time, they are starving.

It may be noted that the bottle necking machine, M1 is now utilized for approximately 83% of time. The parts are sent to machines M2 or M3 after processing at M1. For the rest of the time span, it is starved. This is necessary to ensure that all the parts are completed within the specified time span. This interpretation can only be made when we know the interactions of the parts within the system, through the simulation. Another crucial feature to note is that though the schedule is feasible, it is not efficient. The WIP is high and so is the waiting time for the parts in the system. Therefore, there is a need to refine the production plan that is generated at the end of Stage 2. The production plan (the schedule in particular) at the end Stage 2 is therefore refined in Stage 3 to generate efficient plans.

3.5.1.3 Stage 3

In stage 3, the single stage heuristics at each machine are fixed. This implies that when a part enters the system, its time in system and sequence are fixed. An ATC is used to determine the arrival time that would make the schedule efficient, i.e., loads uniformly spread over the time horizon, lower work in process (WIP) and lower time in the system. The simulation model with the ATC is executed for multiple replicates. The output of one replication is used to modify the inputs for the next replicate, the modifications being determined by the integral ATC.

Since the ATC manipulates the arrival time of the part on the basis of the due date, a hypothetical due date is assigned to each part. The due date is so assigned that the system is forced to try and complete entire sets together, i.e., try and make 8 pawns for every king made in a given time interval.

The time horizon is divided in $2\mathrm{N}$ equal time intervals where $\mathrm{N} = \min\{\sum_{g \in G} X_g\}$. In the example, N corresponds to the number of kings that are to be made. Based on the synchronization coefficients, due dates are assigned to all parts. For the model chess set, 1 king, 2 bishops, and 8 pawns are assigned the due date equal to $\frac{TC}{2N}$, where TC is the total planning horizon. The parts in the second set are assigned a due date equal to $\frac{3TC}{2N}$, the third set, $\frac{5TC}{2N}$. This is a heuristic that is employed for due date assignment and the logic behind it is that the jobs will be symmetrically distributed in a V shape about the common due date, and if the earliness and tardiness are equal, each set will occupy 2 time intervals on the bottlenecking machines.

Once the part has been created, it is sent to the starting station used in previous models where task assignment takes place. Like in the simulation model for stage 2, the start time of each operation on the part is noted and recorded. At the end of the sequence, the part goes to a write block wherein that information of its completion time, and process start time is recorded. However, before exiting the system, the part goes to a ATC and based on the difference between the due date/time and the completion time, the arrival time of the entity is modified. By repeated application of the procedure, there is a gradual decrease in the deviation from the due date/time, decrease in WIP and time in system. When the desired reduction in the deviation has been achieved, the process is stopped and the desired schedule is made available. The decreasing in time in system as the number of replications progress is presented in Figure 3-3.

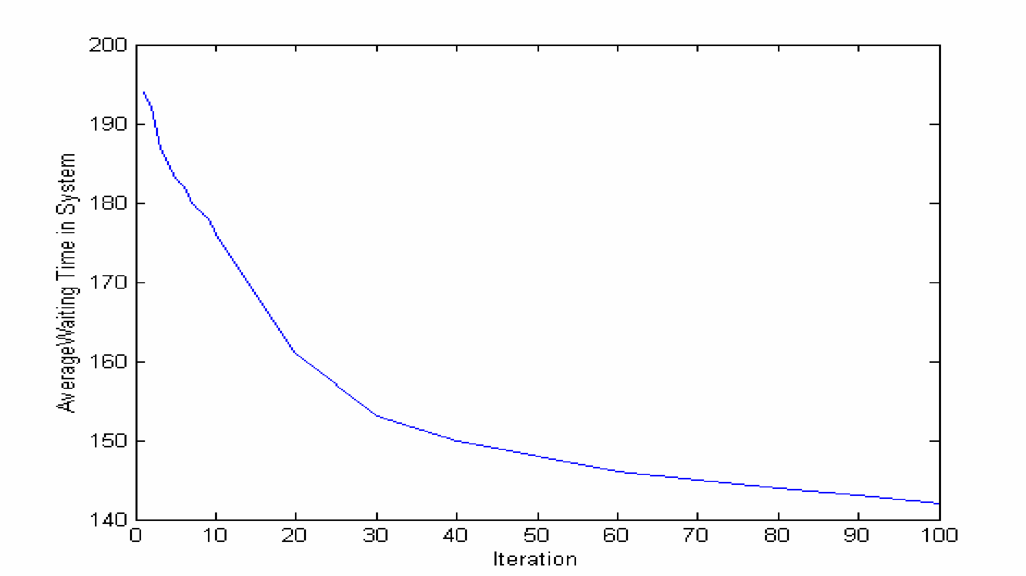


Figure 3-3 Variation of average time in system with replication

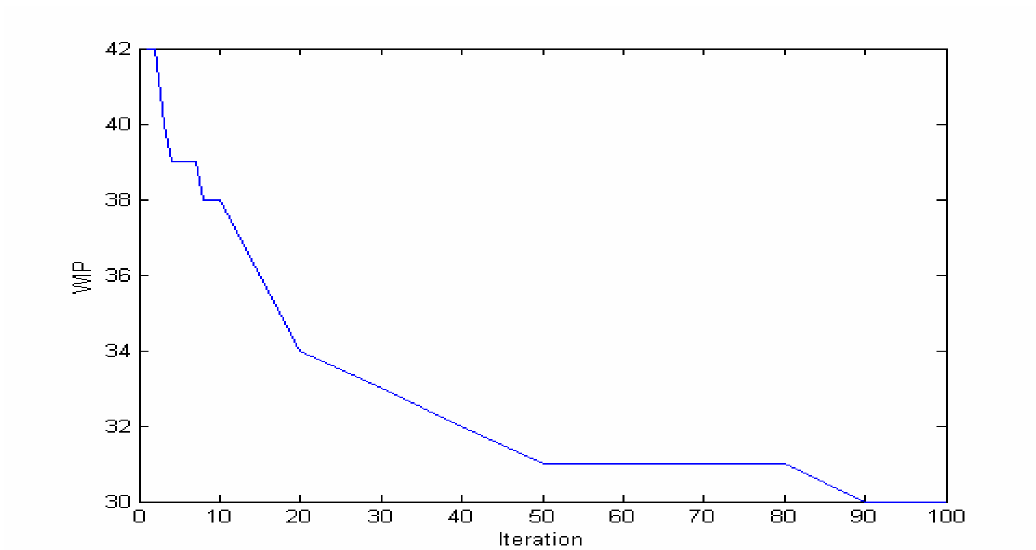


Figure 3-4 Variation of average WIP with replication

An integral controller has been attached to each processing unit. Its function is to minimize the deviation of the completion time of an order from due date. As explained before, the integral controller brings a gradual decrease in the error.

3.6 Scalability of AHFM

Scalability of the P2 model can be determined in terms of the computational time and effort required to solve the P2 model as the problem size increases. The methodology used here comprises of two steps. The first step comprises of identifying the factors that can influence the size of the P2 solution space, their impact on the computational effort, and their interaction with each other. The second step comprises of building empirical equations for predicting the computational time and effort required for solving the P2 model for a given problem.

Step 1: A designed experiment has been conducted to determine the factors that contribute to the size and complexity of the model, and the interactions of these factors with each other. The number of part types, number of processing steps per part (assuming that all the processing is conducted on unique machines), the number of process plans that are available per part type (assuming that all the process plans for a single scenario are unique and non overlapping), the range for the machining parameters, and the number of machines, are the factors identified as contributors to the problem size. The factors and their levels are presented in Table 3-1. It may be noted that four of the five factors are independent: the number of machines (E) is basically the product of factors B and C. The response that is being studied is the computational effort that is required to solve the P2 model. The relationship between the computational effort and the computation time is linear and while the computational time is dependent upon the system configuration, the computational effort is independent of it. In Figure 3-5(a) we plot the relationship between the numbers of iterations and computing time for a Pentium III, 512 MB RAM, 1.1Ghz computer used for this work.

Table 3-1 The process parameters (Treatment factors) that are of interest and their level

Ser.	Factor	Range for DOE (-/+)		Range for regression		No of Levels
		Min	Max	Low	High	
1	Number of part types (A)	3	11	10	100	6
2	Number of operations per part type (B)	3	7	3	20	5
3	Number of process plans per part type (C)	1	3	1	5	5
4	Percentage variability allowed in cutting speed (D)	0	100	0	100	7
5	Number of Machines (E=BC)	N/A	N/A	5	40	5

A 2^{4-1} experiment with two replications was conducted using MINITAB with D=ABC as the design generator for the 4 factors (A, B, C, and D) and the ranges presented in Table 3-1. The half factorial experiment with two replicates at each of the 8 combinations indicate that the factors D, A, and C, and to a certain extent BC (indicator of the number of machines) have a statistically significant effect on the computational effort. The interactions (other than BC which is statistically significant at 0.1) are statistically insignificant. The effect of process plan complexity is also statistically insignificant as far as computational effort is concerned. The Pareto chart indicating the relative contribution of the factors and their interactions is presented in Figure 3-5(b).

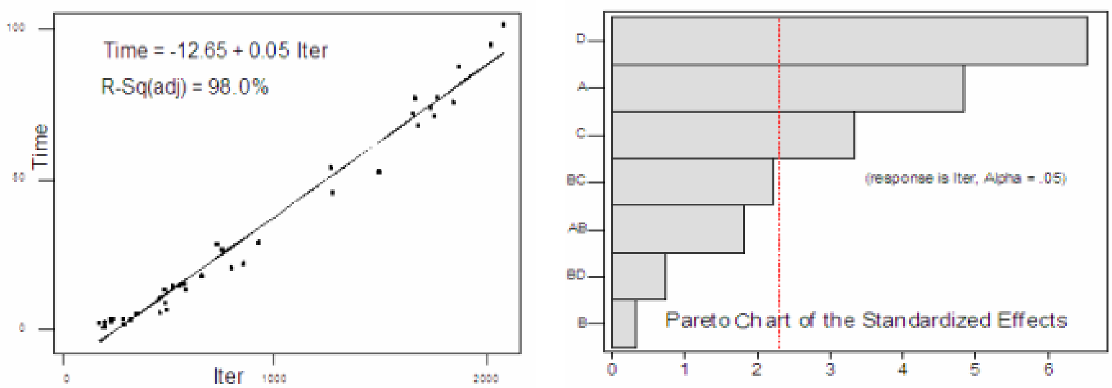
Step 2: To determine the effect that each of these factors, i.e., D, A, C, and BC have on the computational effort, regression over a wide range of values that covers most reasonable size problems was used. The ranges that were considered are provided in Table 3-1. The relationship between the factors D, A, C, and BC and the computational effort required for solving the P2 model (in terms of number of iterations) are presented in Figure 3-6(a-d).

Based on the relationships that have been determined in step 2, it can be concluded that the computational effort is a function of A^2, C, C^2, D and E. The approximate relationship between the computational effort and the statistically significant factors can be determined by fitting a regression equation between the identified factors and their impact and the computation effort. In this integrative regression, C becomes not significant. Using D, A^2, E, and C^2 as the predictors, the regression equation that is obtained is as follows:

$$\text{Iter} = -2408 + 16.5\text{D} + 1.03\text{A}^2 + 102\text{E} - 55.6\text{C}^2$$

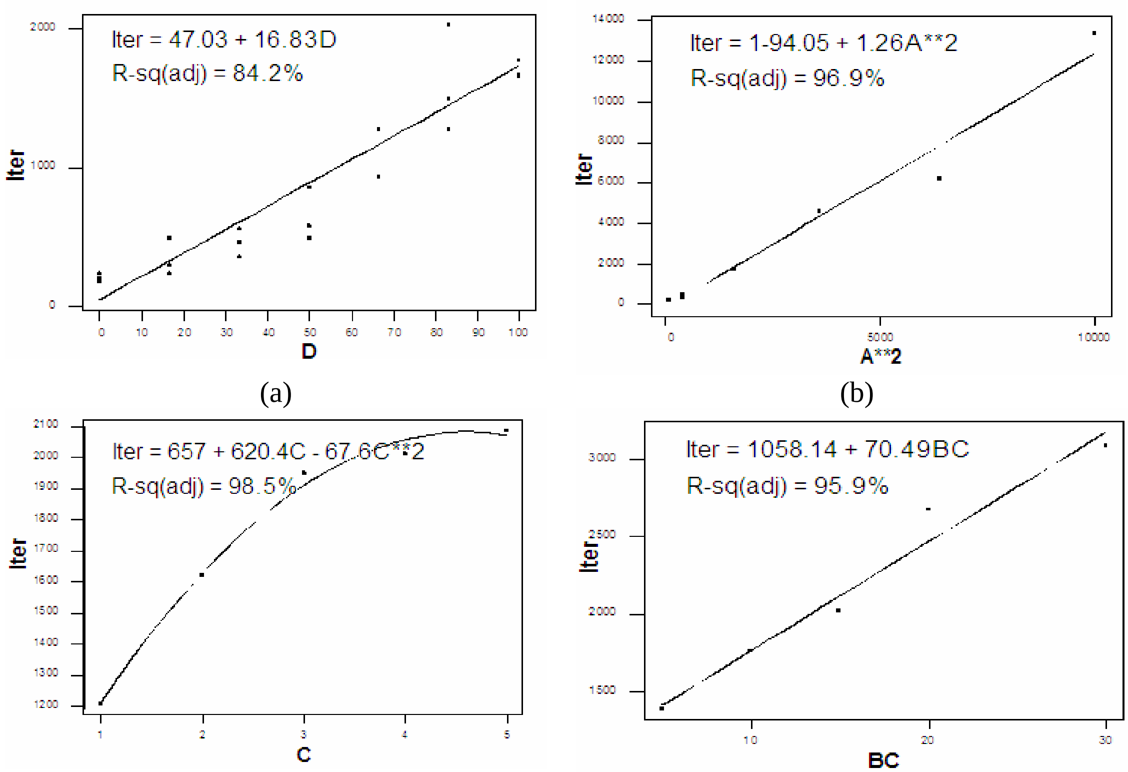
All the factors D, A^2, E, and C^2 are statistically significant with P-values equal to 0.000, the residuals are normally distributed (indicated by the normality plot in Figure 3-7(a)) and the fit is representative of the actual value (the fitted line of slope 1.00 with 0.0 intersect in Figure 3-7(b)). The adjusted R^2 value, indicative of the quality of the regression estimate is 93.3%.

The regression equation indicates that the computational effort scales as a function of $O(\text{A}^2\text{+D+E})$ and the number of alternative paths, C, reduces the complexity. The computational complexity of Model P2 enables us to solve industrial size problems with up to five hundred part types, one hundred of machines and 100% variability range in less than 250,000 iteration that is approximately four hours on Pentium III, 1.1 GHz with 512M RAM.



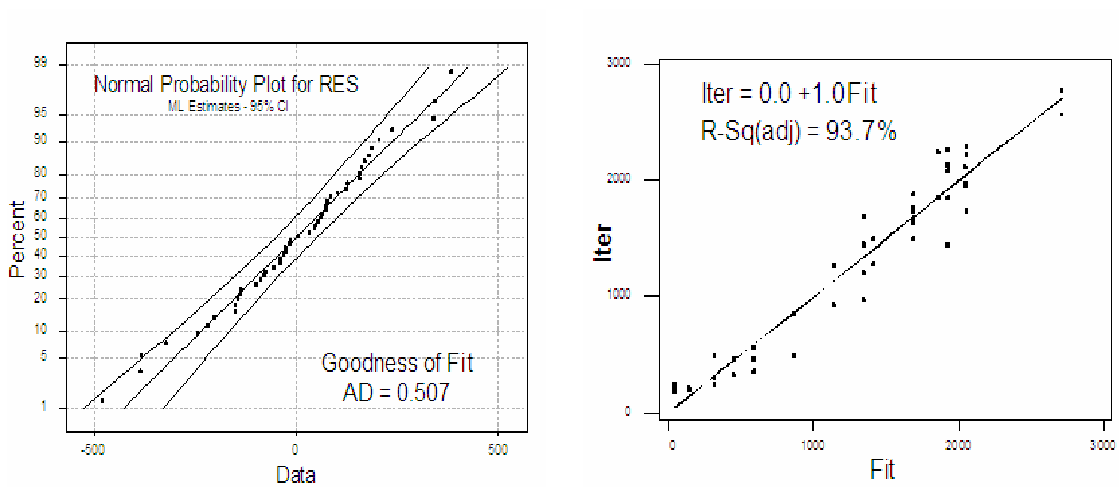
(a) (b)

Figure 3-5 (a) The relationship between computational effort and computational time, and (b) Pareto chart for effects of the four factors and their Interactions on the computational effort to solve P2.



(c) (d)

Figure 3-6 Scaling of Computational Effort with (a) The percentage variability allowed (D), (b) The number of part types (A), (c) The number of available process plans per part type (C), and (d) The number of available machines (BC=E)



(a) (b)

Figure 3-7 (a) Normal probability plot of residuals for the regression equation, and (b) the relationship between the actual values and the fitted values for computation effort by the scaling equation.

3.7 Discussion

The aggregation of the production planning models get more and more complex as we increase the number of hierarchies that the aggregate model covers. It is essential to limit the complexity of the problem with an emphasis on the tractability and computational complexity. Analysis of special properties of the models, such as unimodal regions, is required. A factorial experimental design could be used to determine potential gain of parameter variation in different operational levels and implementation scale of the standard optimization packages such as CONOPT, MINOS, and SNOPT for non-linear optimization, OSL and CPLEX for mixed integer linear models, and DICOPT for mixed integer non-linear models.

Implementation of the developed concept in real-life environments may require development of heuristic algorithms that can deal with large-scale models. Based on our small but realistic examples the opportunity gap is large enough to justify “optimal versus heuristic” losses. The gap may even increase with problem size: the balancing effect of parameter variation is more important in complex manufacturing environments with multiple products, alternative process plans and shifting bottlenecks.

At the end of Section 3.6, it is shown that the computational time results for systems with increased size (number of parts, number of alternative plans, variability of the range of parameters and the number of machines). As expected, when the size of the problem increased so did the solution time. Solution time for increases in size appears to grow linearly as the problem increases in the number of machines, and the number of operations in a part plan. The computational time increased quadratically as the number of parts increased. Solutions for problems of a large size (20 machines and 100 parts types) were obtained in less than an hour using a personal computer. It appears that this procedure can easily be scaled to solve real manufacturing problems.

Addendum Chapter 3:

Enhancing convergence properties of simulation optimization by hybridized ATC[3]

3.8 Introduction

There is a need for highly flexible, low complexity systems that can handle the large-scale real time scheduling problems. In this work, a multiple population genetic algorithm (GA) and control theoretic based arrival time control (ATC) have been combined to achieve the objective of just in time (JIT) scheduling in a job-shop. The GA and ATC compliment each other and the hybrid system, relying on its massively parallel architecture can handle large-scale JIT scheduling problems efficiently. The hybrid approach has the potential to drastically reduce the computational complexity of the problem. Initial results show a 21% improvement in solution quality over simple ATC and 0.5% improvement over heurisitcal solutions for a sample problem considered here.

Scheduling for just in time (JIT) delivery is fast becoming the de facto requirement in many small to medium scale manufacturing shop floors. Over the last few years, researchers and practitioners have approached the problems that arise in job shop scheduling when the objectives are earliness-tardiness related. Combinatorial optimization techniques (e.g., Ventura, 1995, Pinedo, 1995, Baker, 1974), control theoretic approaches (e.g., Prabhu, 2003, Kogan and Khmelnitsky, 2001), and artificial intelligence (AI) based approaches (Scholl et al, 1995) have been proposed. The implementations of these scheduling approaches, however, are not that widespread in industry.

Researchers focus on complexity analysis, rigorous analysis of exact procedures, or heuristics for mathematically tractable special cases of the real problems. Most large-scale, real-time scheduling problems do not fall into these special categories. Large and complex search spaces for solutions, dynamic shop floor conditions, and a multitude of shop floor conditions are the factors that set these problems apart. There is a need for highly flexible systems with low complexity that can handle such problems. It is also essential that the system be robust enough to handle variations in shop floor conditions and scalable to handle real size problems. The focus of this work is on combining features of evolutionary algorithms and control theoretic approaches and developing a hybrid approach for JIT scheduling. The goal is to develop a scalable, robust, and low complexity approach that generates efficient schedules.

---

[3] Based on:
Shaikh, N.I. and Prabhu, V.V., Reed, P. (2003), "Hybridized Arrival Time Control Approach to Job-Shop Scheduling", in the Genetic and Evolutionary Computation Conference (GECCO) Late Breaking Papers, Erick Cantu-Paz (ed.), Chicago, 2003.

The proposed hybrid approach relies on problem partitioning and local search for developing schedules. Arrival time control (ATC) formulation (Prabhu, 2003) transforms the scheduling problem into a continuous variable control problem and eliminates combinatorial complexity. Though ATC has an exponential convergence rate towards local attractors, it cannot guarantee optimality for most scheduling problems. In this work, ATC has been used to partition the problem and generate initial populations near local attractors. A multiple population genetic algorithm (GA) has been applied within these partitions on the populations of solutions to seek superior solutions adaptively. Offline analysis is conducted and the best solution is archived for generating the final schedule. The rest of the chapter is organized as follows: The job-shop scheduling problem with earliness-tardiness penalties has been described in Section 2. The ATC algorithm has been described in Section 3 and the GAs in Section 4. The proposed hybrid algorithm is presented in Section 5. Results are presented in Section 6 and future research in Section 7.

3.9 Scheduling and complexity

Within the domain of production control, scheduling refers to the specific activity of timetabling the operations dictated by the process plans so as to achieve desired levels of performance at the shop floor level (Rodammer and White, 1988). The job-shop scheduling problem is that of scheduling a set of n jobs that have to be processed on m machines, where the processing of each job consists of m operations performed on these machines in a specified sequence. The operation of job i has to be performed on machine j with deterministic processing time t_{ij}. A machine can process only one job at a time, and an operation cannot be preempted. The objectives can be one or more, including minimization of makespan, tardiness, or maximize machine utilization, customer satisfaction, etc. The focus of this work is on earliness-tardiness related objectives in JIT systems, i.e., minimize D, the mean squared deviation (Equation (3-13), subject to Equations (3-14), (3-15), (3-16), and (3-17).

$$Min \quad D = \sum_{i=1} \sum_{j=1} (d_{ij} - c_{ij})^2 \tag{3-13}$$

Subject to:

$$\sum_{k=1}^{TC} X_{ijk} = 1 \qquad \forall \text{ i, j} \tag{3-14}$$

$$\sum_{k=1}^{TC} kX_{ijk} + t_{i(j+1)} \le \sum_{k=1}^{TC} kX_{i(j+1)k} \quad \forall \text{ i and } j = 1,..., J - 1 \tag{3-15}$$

$$\sum_{i=1}^{I} \sum_{j=1}^{J} X_{ijk} Y_{ijm} + \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k'=k+1}^{\min(TC, k+t^{\max}-1)} X_{ijk'} Y_{ijm} \xi_{ij(k'-k)} \le 1 \qquad \forall \text{ m, k} \tag{3-16}$$

$$\xi_{ijk} M \ge t_{ij} - k \qquad \forall \text{ i, j, and k} \tag{3-17}$$

Equation (3-14) ensures that all the operations for a part are completed; Equation (3-15) ensures that operations are performed in the right order, and Equations (3-16) and (3-17) ensure that the machining constraints are not violated.

It has been shown that even the simplest models that deal with earliness tardiness penalties are NP-Complete (Cheng and Gupta, 1989). The number of possible schedules exceeds $O((n!)^m)$. Most of the solution techniques for JIT scheduling problems rely on relaxation techniques and heuristic methods for this very reason. The primary goal of this work is minimize the complexity of finding high quality scheduling solutions.

3.10 ATC in scheduling

ATC is a simulation-based algorithm that transforms the scheduling problem into a continuous variable control problem. As the name suggests, the algorithm calculates the schedule for all the parts by manipulating their arrival times. Given the expected processing time and the due date for each part, the algorithm tries to minimize the deviation of completion time from the due date by adjusting the part arrival time in the next simulation run. Essentially, for a *n* part *m* machine problem, there will be *m* resulting systems (1 for each machine), and each comprising of n-dimensional multivariable control systems, where due date, completion time, due date deviation, and arrival times are n-dimensional vectors acting as the command, output, error and manipulated vectors respectively (Prabhu, 2003). The logic is graphically illustrated in Figure 3-8.

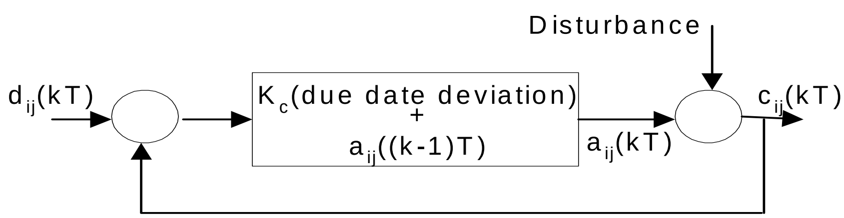


Figure 3-8 Feedback Control for ATC

The algorithm calculates the completion time of each part (which is nothing but the sum of arrival time of the part and the processing times of each of the parts that arrived before it and the queuing time). Based on the total time for the completion of all the parts, deviation from the actual due-date is calculated and the arrival times for the parts at next iteration are changed according to the deviations. Equation (3-18) represents the mathematical expression for the ATC controller.

$$a_{ij}(t) = K_c \int (d_{ij} - c_{ij}(\tau))d\tau + a_{ij}(0) \qquad (3\text{-}18)$$

Where, a_{ij} is the arrival time, K_c is the control system gain. On further simplification the above equation transforms into Equation (3-19) where T is the discrete time step. The schedule with the best performance discovered in these iterations is saved and used to schedule events.

$$a_{ij}(kT) = K_c(d_{ij} - c(k-1)T) + a_{ij}((k-1)T) \qquad (3\text{-}19)$$

It may be noted that queuing introduces discontinuity in the system. Parts are processed in the order in which they arrive and if a part arrives when another part is being processed at the machine, it incurs queuing time. If a part arrives when the machine is idle, this queuing time is zero. All the parts that are processed by a machine between two consecutive machine idle times form a fragmented queue and are part of an "active sub-problem". The simulation identifies these fragmented queues as the arrival times for the parts in the active sub-problem converge to steady-state values. The relative order of processing within each queue fragment however continues to change as the ATC algorithm keeps searching for better schedules based on the processing sequence change. This phenomenon known as chattering (illustrated in Figure 3-9) occurs as ATC is oblivious to optimal.

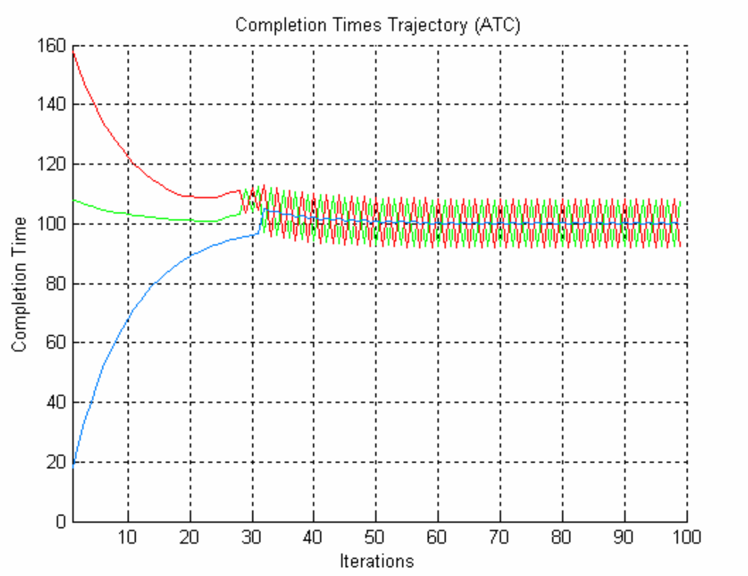


Figure 3-9. Chattering in ATC

After the convergence of the arrival times and formation of active sub-problems, the n part m machine job-shop scheduling problem can be seen as $\alpha_i \ \ell_{\alpha i}$ part 1 machine sequencing problems, where there are $\ell_{\alpha i}$ parts in the α_i^{th} sub-problem. Recent work has focused on developing techniques to design systems to improve scheduling performance and predictability (Cho and Prabhu, 2002). Attempts are being made to incorporate a global system view with the ATC algorithm (Hong, et al, 2003). This work relies on the GAs for introducing a global perspective into ATC based scheduling.

3.11 GA for job-shop scheduling

Some of the aspects that make the GA a favorable tool for combining with ATC are:

Flexibility: The GAs can effectively handle problems that many traditional optimization algorithms cannot include: (1) discrete spaces, (2) non-linear, discontinuous evaluation functions, and (3) nonlinear discontinuous constraints. GA has the ability to extract an initial population and display rapid convergence even in discontinuous regions.

Scalability: The favorable scaling of evolutionary algorithms as a function of the dimension of the search space makes them particularly effective in comparison with other search techniques for large search spaces.

Robustness: The GAs use a population rather that a single point in the search space for evolution. This incorporates robustness into the search procedure as well as fault tolerance when inter-process communication is used.

Besides the fact that a GA is efficient in the discontinuous region, the approach blends well with the distributed architecture that ATC maintains. Local populations can be generated for each individual zone of discontinuity. The details have been presented in Section 5.

To use GA, however, one must first represent the problem in the right structure. It may be noted that the ATC is a simulation-based approach and the machine capacity constraints (Equations (3-16) and (3-17)) are implicitly defined in the simulations. Equations (3-16) and (3-17) are satisfied in the partitions that are created by ATC. Constraints (3-14) and (3-15) however, have to be included in the representation that is selected. The procedure then applies selection and variation operators to these individuals in the population to generate new individuals. GA use the selection operator to pick high quality individuals for mating, producing superior solutions by combining parts of the parent solutions intelligently. The objective function of the problem being solved determines show good each individual is. In this work, we use the following specifications for the GA.

3.11.1 Representation

Several chromosome-encoding schemes have been discussed in literature for sequencing and scheduling of operations. These include operation-based, job-based, preference list based, machine based, or random keys based representations (Pongcharoen et al, 2003). These representations tend to become complicated as the problem size and the complexity increases. In this work, a variation of the alphanumeric representation has been used.

In the proposed hybrid approach, the partitioning stage generates sub-problems that involve sequencing of parts on a single machine. Since each sub-problem is specific to that machine, ordinal values of the entity types in a sub-problem are mapped into each chromosome of the sub-population. Essentially, a sub-population will be comprised of chromosomes of a given length having unique integer numbers from 1 to the ordinal number representative of the length. For example, if there are 9 parts in an active schedule for a machine, a chromosome for that sub-population can be represented as: 1 2 3 4 5 6 7 8 9, where “1” is the ordinal value of an entity in that sub-problem.

3.11.2 Initialization

The chromosomes in the initial sub-populations are generated randomly. Heuristic rules such as SPT (shortest processing time first) or FIFO (first in first out) could be applied for initial population generation, but these heuristics perform well only in the common due date problems and so cannot be generalized for a sub-population. The initialization process for sub-problem α_i has two stages:

1. Strings of length $\ell_{\alpha i}$, having integer numbers sequences from 1 to $\ell_{\alpha i}$ are generated. Essentially, if *N* strings have to be created for a sub-population with 9 entities, *N* strings of characters 1 2 3 4 5 6 7 8 9 will be generated

2. The numbers in each chromosome string are then mixed randomly within the string itself, and new sequences containing all the numbers in a random order are generated.

These are then used as inputs as new members of the population. For example, 2 8 4 7 1 3 9 6 5 is a new chromosome.

3.11.3 Selection

A modified "*N* best" reproduction scheme (Cheng, et al, 1996) has been used for generating the children population and determining the population for the next generation. In this scheme, η off springs are produced from the previous population of size *N*. The *N* best chromosomes out of the $N+\eta$ old ones constitute the population for the next generation.

3.11.4 Diversity Operators

3.11.4.1 Crossover

Each chromosome in has length $\ell_{\alpha i}$ and comprises of $\ell_{\alpha i}$ unique integer numbers ranging from 1 to $\ell_{\alpha i}$. This property should be maintained in the child population that is generated after the crossover and mutation operations. The "uniform order based crossover" is considered to be a good fit to this kind of constraints (Lee and Choi, 1995). This crossover operator considers both the absolute and relative positions of genes in the parent chromosomes for generating the children. The crossover operation is conducted in four steps:

1. Pick two parent chromosomes. The parents can be picked based on their fitness function values or randomly. In this work, they are randomly picked.
2. Generate a binary string of length $\ell_{\alpha i}$, the length of the chromosomes in the population (9 in this case). A binary template could be as follows: 1 1 0 0 0 0 1 0 1
3. Fill in the positions in Child 1 (or Child 2) by copying them from Parent 1 (or Parent 2) whenever the bit strings contain a 1 (or 0). For example, suppose the following parents are selected for crossover:
   Parent 1: 1 2 3 4 5 6 7 8 9 and Parent 2: 2 3 5 4 1 8 9 7 6

   And the Binary template is 1 1 0 0 0 0 1 0 1, then the resultant intermediate children chromosomes are:

   Child 1: 1 2 _ _ _ _ 7 _ 9 and Child 2: _ _ 5 4 1 8 _ 7 _ respectively
4. List the genes from parent 1 (or parent 2) associated with a 0 (or 1) in the binary template and permute these genes so that they appear in the same order as they appear in on parent 2 (or parent 1). Thus, the crossover yields the following final children:
   Child 1: 1 2 3 5 4 8 7 6 9 and Child 2: 2 3 5 4 1 8 6 7 9

3.11.4.2 Mutation

Intra-string mutation has been used for introducing variability in the system. The mutation operator chooses two jobs in a chromosome at random and exchanges their positions. For example, if the parent is: Parent 1: 1 2 3 4 5 6 7 8 9 and the probability of mutation is 0.1, the child can be: Child 1: 1 9 3 4 5 6 7 8 2

3.11.5 Time Continuity

Time continuity has been proposed as an added tool for introducing diversity and variability (Goldberg, 2002). A fixed number of randomly generated chromosomes λ are also injected into the population after every κ generation to add diversity and search pressure (Reed, 2000, 2002). Therefore, $N \rightarrow N + \lambda$ after every κ generations.

3.11.6 Swapping

Before the objective function value associated with a chromosome can be evaluated, it is required that the local feasible of the sequence be tested. Although a $\ell_{\alpha i}$ part 1 machine problem is being solved in each active sub-problem, for re-entrant lines, etc, there is possibility that there are some precedence constraints that need to be adhered to. The swapping operator swaps the two genes if it is violated. For example if in parent 1, the ordinal values 2 and 9 represent the same part, but different processing steps, and the operation indicated by ordinal value 2 has to be performed before the operation indicated by 9, and the position of the two genes are interchanged in the final chromosome.

Child initial: 1 9 3 4 5 6 7 8 2

Child final: 1 2 3 4 5 6 7 8 9

It may be noted that the final child and the parent chromosome represent the same sequence. This is quite common and can lead to pre-convergence. Steps such as population injection are therefore recommended.

3.11.7 Fitness evaluation

Each chromosome in a sub-population represents an active schedule for the machine for a given time frame. The arrival time of the first part is determined from the ATC partition and since the schedule is active, there is no inserted idle time. The j^{th} part in a given sequence incurs a waiting time of w_{ij} after that the machine processes it for duration of t_{ij} and then completes it at c_{ij}. The completion time for the j^{th} part in the sequence can therefore be expressed as

$$c_{ij}(t) = a_{i1}(t) + w_{ij}(t) + t_{ij}(t) \tag{3-20}$$

And because of zero inserted idle time,

$$w_{ij}(t) = a_{i1}(t) + t_{i1}(t) + t_{i2}(t) + ... + t_{i(j-1)}(t) \tag{3-21}$$

From Equations (8) and (9), completion time can be expressed as

$$c_{ij}(t) = a_{i1}(t) + t_{i1}(t) + t_{i2}(t) + ... + t_{ij}(t) \tag{3-22}$$

The fitness criteria, the due date deviation is then calculated as

$$D = \sum_{j=1}^{\ell_{\alpha i}} (d_{ij}(t) - c_{ij}(t))^2 \tag{3-23}$$

3.12 The Hybrid Approach

The proposed approach can be represented as shown in Figure 3-10. The approach involves six basic steps:

3.12.1 Data Coding

This step involves converting the problem from a n part, m machine, m processing step into a $n \times m$ part m machine problem, that is suitable for the proposed algorithm. Unique part ids and process ids are assigned to each of the $n \times m$ entities. This forms the primary key that uniquely identifies each part and the processing step that it needs to undergo. Alphanumeric coding scheme suggested by (Pongcharoen et al, 2003) has been used in this work as it can be extended easily and interpretation is straightforward. It may be noted that each of the initial n parts can have unique process plans with any number of processing steps. The assumption of m processing steps per part has been taken for simplicity and clarity in expression and demonstration.

3.12.2 Due Date Assignment

This step involves assignment of due dates to each of the $n \times m$ entities. Due dates of the m^{th} processing steps for all the n parts are fixed (or predetermined). Due dates for the remaining parts can be assigned using different heuristics depending upon the goals and constraints of the shop floor; the due dates can be uniformly spaced in the planning horizon, or assigned based on the allowed makespan for the part. It is essential for the due date assignment rule to not contradict the precedence constraints that exist in the process plans or because of other shop floor constraints such as setup. In this work, a common intermediate due date has been assigned to all the parts for illustration of the overall algorithm.

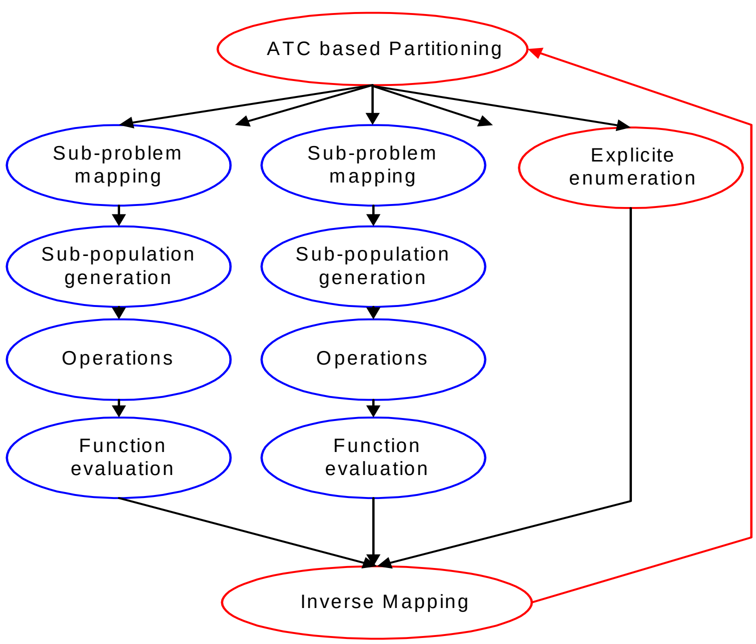


Figure 3-10. Hybrid Algorithm

3.12.3 ATC based Partitioning

Given the $n \times m$ entities, their processing times and due dates, and the shop floor conditions (constraints and machine states), the next step involves executing the ATC algorithm to partition the problem into active sub-problems. From Figure 3-9, it can be seen that the ATC converges in approximately 45 iterations. We fix $I = 50$ for this work. In general, we can take $I \approx 10T$, where T is the discrete time step used in Equation (7).

3.12.4 Sub-Problem Mapping

This step involves identification of the partitions and the sub-problems. Entities separated by an idle time of τ or greater duration in a machine schedule are assumed to be in different partitions. The value of τ has to be less than or equal to the smallest processing time of an entity in the active schedules separated by that τ. Once the partitions have been determined, mapping tables as shown in Tables 1 and 2 can be developed.

3.12.5 GA based Sequencing

This step involves creation of a sub-population of GA for the active sub-problems that have been created. To improve efficiency, sub-populations are not associated with all the active sub-problems. For a sub-problem α_i with $\ell_{\alpha i}$ entities, explicit enumeration involves $\ell_{\alpha i}!$ function evaluations. For the same size problem, GA based approach will require $N \times g$ function evaluations, where N is the population size and g is the number of generations. Both N and g are dependent on $\ell_{\alpha i}$. Making conservative estimates and using heuristics mentioned in competent GA literature (Reed, 2000, 2002, Goldberg, 2002), the number of function evaluations can be approximated to $2\ell_{\alpha i} \times 2.8\ell_{\alpha i}$. Sub-populations are created only when the number of function evaluations using explicit enumeration far exceeds the number of function evaluations required by a GA, i.e., $2\ell_{\alpha i} \times 2.8\ell_{\alpha i} \times 10 \le \ell_{\alpha i}!$. This happens when $\ell_{\alpha i} \ge 7$. It the number of entities in an active sub-problem is greater than 7, sub-populations are created. The details of this step have been described in detail in Section 4.

Table 3-2 Mapping tables for active schedules 1 and 2

Table 1 Mapping Table for an active schedule 1				Table 2 Mapping Table for an active schedule 2			
Serial	Part id	t_{ij}	d_{ij}	Serial	Part id	t_{ij}	d_{ij}
1	14.2	9	320	1	2.1	5	150
2	6.2	8	320	2	3.1	4	150
3	1.2	7	320	3	31.1	3	150
4	4.2	6	320	4	4.1	2	150
5	3.2	5	320	5	7.1	1	150
6	2.2	6	320	6	11.1	2	150
7	11.2	7	320	7	14.1	3	150
8	7.2	8	320	8	6.1	4	150
9	31.2	9	320	9	1.1	5	150

3.12.6 Inverse Mapping

After a fixed number of iterations/generations/time, the best solutions in each of the sub-population is taken and reinserted into the ATC model. Infeasibilities, if any, are eliminated.

3.13 Results

3.13.1 Problem Definition

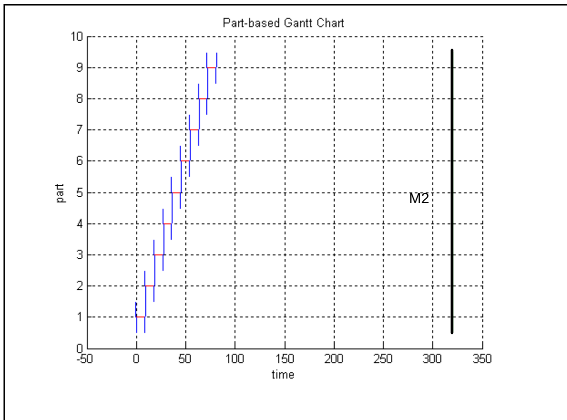
The problem taken up to illustrate the application of the hybrid approach comprises of 9 parts with 2 processing steps each, and to be processed on 2 machines. The common due date for all the parts is 320 time units. The processing times and the Part ids (primary key) are presented in Table 3-3. The desired schedule should minimize the deviations from the due dates. The common due date problem has been selected because the solution lies in the discontinuous region and ATC is inefficient in converging to a solution.

Table 3-3 Problem definition

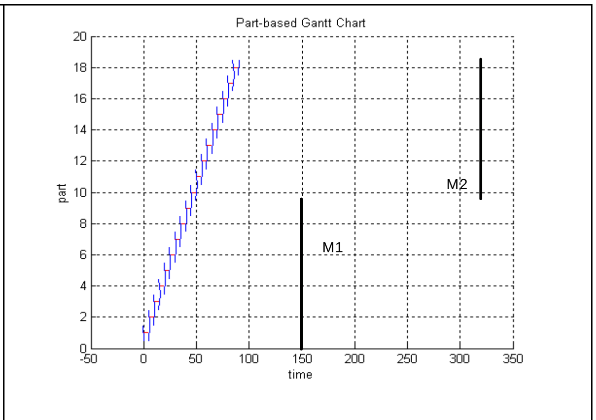
Serial	1	2	3	4	5	6	7	8	9
Id	1	2	3	4	6	7	11	14	31
t_{ij}	5	5	4	2	4	1	2	3	3
t_{ij}	7	6	5	6	8	8	7	9	9
d_{ij}	320	320	320	320	320	320	320	320	320

3.13.2 ATC based Partitioning

The initial 9 parts, 2 machines, 2 processing step problem (Gantt Chart shown in Figure 3-11(a)) was converted to a 18 part 2 machine problem. The due dates and part Id. assigned to the intermediate steps have been presented in Table 3-2. The Gantt chart of the modified problem is presented in Figure 3-11 (b). ATC was then used to partition the solution space. The mean squared deviation (MSD) performance of the algorithm is presented in Figure 3-12(a). The solution space partitions are illustrated in Figure 3-12(b).

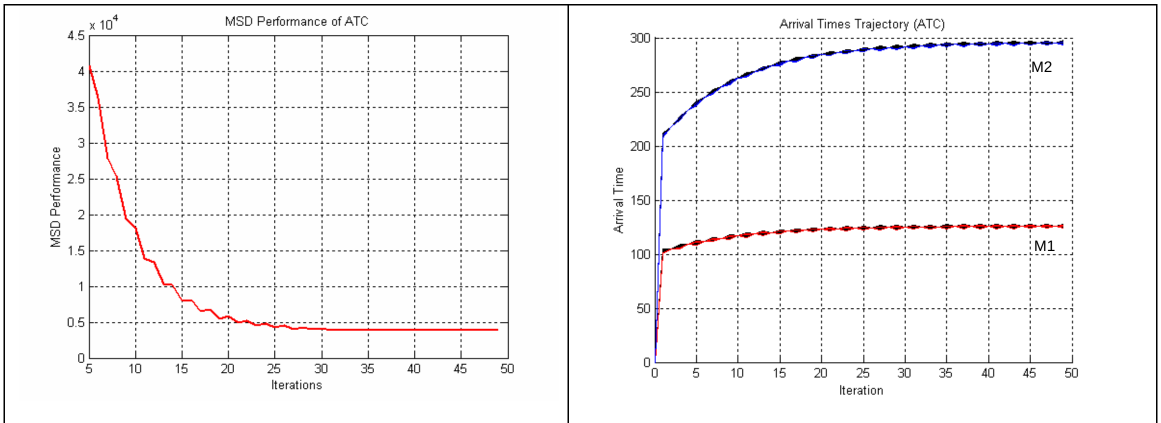


(a) Original Problem



(b) Modified Problem

Figure 3-11. Initial Gantt Charts for original and the modified problem



(a)	(b)

Figure 3-12. (a) MSD performance of ATC and (b) Solution Space Partitioning by ATC

The Gantt Charts created by the ATC algorithms are presented in Figure 3-13. The problem comprises of two local attractors and can be partitioned into two sub-problems. A GA is then used in the sub-populations to optimize the sequencing of the arrival times of the parts in each local sub-population.

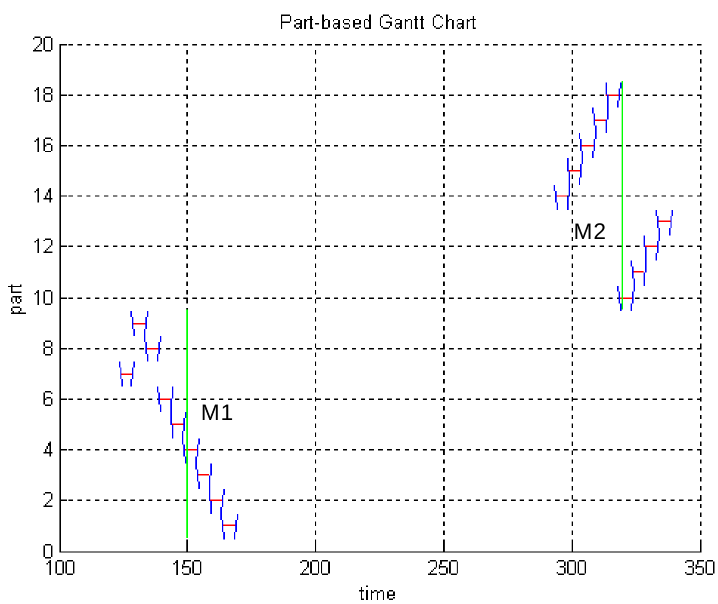


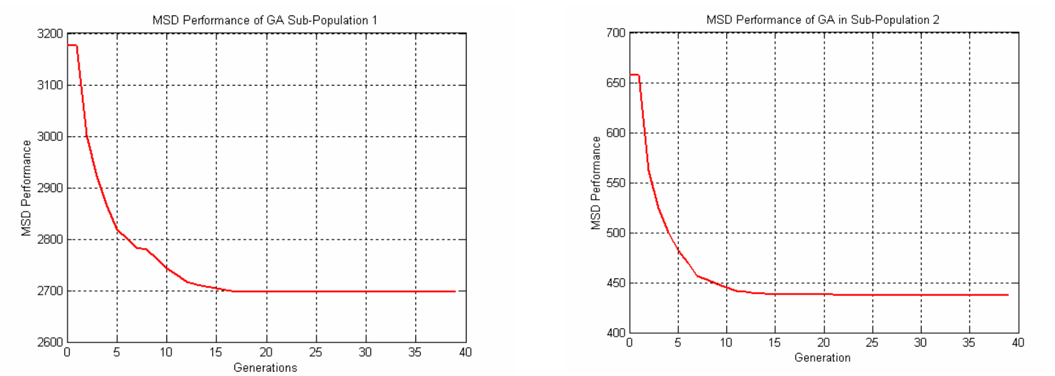
Figure 3-13. Final Gantt chart for Modified Problem

3.13.3 GA based Sequencing

The mapping tables have been presented in Table 3-2. These are used as inputs for the GA sub-populations. The representation and operators presented in Section 5 were used for each sub-population thus created (in this case, 2). The parameters used in the sub-population are presented in Table 3-4. The MSD minimization performance in the sub-populations has been presented in Figure 3-14 (a) and (b).

Table 3-4 GA Parameters

Id.	N	ℓ	$\eta_{init} \approx \ell$	$\eta_f \approx 2\ell$	λ	κ $(2\eta_f / \ell\lambda)$	Gen $\approx 2\eta_f$	p_m
1	9	9	10	20	2	8	40	0.1
2	9	9	10	20	2	8	40	0.1



(a) Sub-Population 1 (b) Sub-Population 2

Figure 3-14. MSD performance of GA in Sub-Populations 1 and 2

3.13.4 Comparison and verification

Initial results for the proposed hybrid approach have been presented in Table 3-5. There is an approximately 21% improvement from the solution that is given by using ATC alone, and 0.5% improvements from solution given by the heuristics for earliness-tardiness minimization heuristics (Ventura and Weng, 1995). Gantt chart displayed in Figure 3-15 shows the V-shaped alignment of the parts around the common due date as predicted by the heuristic developed by Ventura, 1995.

Table 3-5 MSD Performance of Algorithms

Algorithm	MSD
ATC (100 Iterations)	3840
V-Heuristic	3144
Hybrid	3005

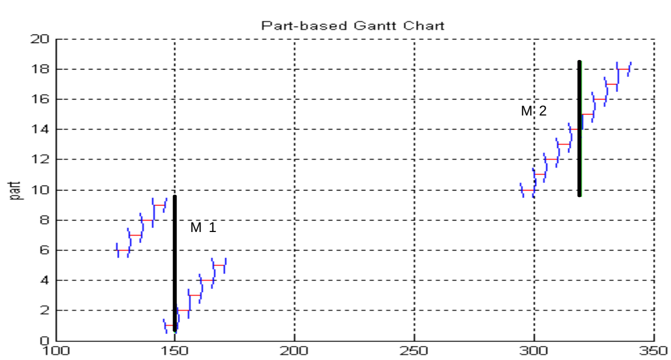


Figure 3-15. Hybrid Algorithm

3.14 Conclusions and future work

The ATC algorithm has an exponential convergence rate towards local basins of attraction in due date based scheduling problems. The proposed hybrid algorithm introduces a global system view in the ATC algorithm by using GAs. ATC is used to partition the problem and generate initial populations near the local attractors. GAs are then applied within these partitions on the populations of solutions to seek superior solutions adaptively.

A GA is efficient in the discontinuous region and blends well with the distributed architecture that ATC maintains. Local populations of chromosomes have been generated for each individual zone of discontinuity, thereby limiting the length of the chromosome. Offline analysis has been conducted and the best solution is archived for generating the final schedule.

Other advantages of the proposed hybrid approach are:

- Scalability: The GAs are efficient in the discontinuous region and blends well with the distributed architecture that ATC maintains. The computations for ATC can be executed on a separate processor in parallel, making them inherently suitable for massively parallel/distributed computing.
- Robustness and Fault tolerance: This feature is inherited from the GAs, where the operations are on populations of points rather than at single points in the search space.

Some aspects that are currently being explored for enhancing the hybrid approach are:

- A due date assignment strategy for the intermediate stages of a part. This will probably depend on the secondary objectives and constraints related to makespan. The assigned due dates for the intermediate stages dictates the distance between the partitions and the number of sub-problems.
- Time continuum, choice between the Baldwinian and the Lamarckian evolution, etc needs to be considered in details.
- Improved synchronization strategies for ATC and GA communication. This may be related to the selected evolution strategy.

Chapter 4

On the applicability and utility of the AHFM approach[4]

4.1 Introduction

The applicability of the AHFM approach has been demonstrated for the hierarchically organized planning systems in a manufacturing unit (Chapters 3 and 4). The scalability of the simulation-optimization based implementation strategy has also been tested for the same case. A major part of the focus of this previous chapter was on uncovering the potential savings that may be gleaned from such a technique. The indicated savings are quite significant and available by only changing the control of an operational system without any capital investment. It is also found that combined effect of alternative process plans and variable machining parameters is much higher than their individual improvements: while machining parameter variation and alternative paths add to the profit 1.5% and 15%, respectively, their combination improves the result by 30%. From the example, there is a strong indication that heuristics for integrative large-scale problems can get better results than optimally solved sequential models in the traditional hierarchical planning. The drawback is that implementing the AHFM approach is case specific.

The focus of this chapter is study the application of AHFM in two other domains, namely the continuous manufacturing, and transportation. Two case studies have been taken up and solutions are presented that illustrate the efficacy of the approach in these domains as well.

Further, the AHFM approach is extended to exemplify the utility of the AHFM approach in dealing with events. The case of addition of new orders has been taken up for illustration. Company managers often have to answer questions regarding price and due date for new customer's orders. While customer's price can be determined by many business parameters, the decision maker needs a clear view of the cost-due date tradeoff in order to answer the above question. If additional production capacity is available with the shop floor, the newly introduced task is incorporated in the plan without any difficulty. However, more often than not, the schedule for the shop floor is tight. Shop floors resort to different strategies to deal with the introduction of unplanned activities in the production plan in this scenario. Depending upon the backorder cost of the parts in the original production plan, the planner either drops some part of the previous production plan in order to accommodate the additional demand on resources, or, if the backorder cost is too high, the additional order is not accommodated in that time frame. The AHFM approach explores another possibility of dealing with the event, i.e., the

[4] Based on:

1. Shaikh, N.I., Masin, M., and Wysk, R.A., (2003), "Estimation of the cost of customized goods for a shop floor", Penn State University IME Department Working Paper.
2. Masin, M., Shaikh, N.I., and Wysk, R.A., (2003), "Aggregative parameter variation in optimization modeling", IEEE Transactions on Automation and Robotics, Aug 2003.

previous plans can be perturbed: the processing times assigned to the scheduled production can be changed to accommodate the additional load on the machining resources.

4.2 Background

Aggregation rules have been used to integrate some of the hierarchy levels shown in Figure 1-1. The AHFM approach indicates improvements in performance by up to 30% in manufacturing environment. The percentage increase is too high to neglect as this increase in profitability comes without any additional capital investment. Two other cases are included. Improvement are expected, but the question that needs to be answered is how much.

4.2.1 Process industries

A processing plant consists of a large number of equipment operated using different parameters under different conditions. Buskies (1997) worked on devising a rational method of selecting process parameters, such as throughputs, pressures, temperatures, concentrations, etc. under different operating conditions. A whole series of process parameters, which can be varied either not at all or within only very narrow limits were identified and taken into account in the planning and optimization of process plants. However, it was pointed that the problem of plant optimization only occurs when there are design parameters that can be selected at will or can, at least, be varied within certain limits. This is however, virtually always the case.

Horishi (1996) proposed a systematic procedure comprising of a two stage automated hierarchical planning procedure, consisting of both a coordinating function between production periods and an efficient job lot sequencing function. The problem of such production system management is to determine the product mix to be produced in the current planning period based on term-wise forecasts of demand quantities, constraints of plant capability, etc. Most of the current procedures operated in actual factories are empirical heuristic approach and are relevant under stable demand structure and a virtually static environment.

The optimization of process systems with complex investment cost functions, defined over several intervals of equipment sizes, operating pressures and temperatures has been addressed by Turkay and Grossmann (1998). Process synthesis models incorporate the selection of process units and operating conditions of these units. The discontinuities with respect to variables were modeled with disjunctions that were convened into tight mixed-integer constraints with the convex hull formulation for each disjunction. There is therefore a need for an optimization strategy in the form of an objective function, which is either reduced to a minimum, where lowest possible production costs are the targets, for instance, or advanced to a maximum where high profitability is the aim (Buskies, 1997).

4.2.2 Fleet scheduling

Fleet assignments determine the type of aircraft to operate for flights in a given schedule subject to a variety of constraints. Airlines usually determine the timings for their flights to respond to time-dependent demand and the requirements of frequency plans, available fleets, and aircraft routings. Nevertheless, delays and rerouting are unavoidable. Adjustments have to be made by altering some flights from their so called optimal times. Scarce runway time slots represent a resource whose value can be determined from the impact of such re-scheduling on the objective function of the original schedule. Cao and Abid (2000) analyzed the relationship between re-scheduling of flights and airline profit.

As we have seen recently, poor scheduling of flights resulting from temporary closings of airports causes substantial loss of profit and decreased levels of service for airline carriers. It is therefore imperative to have the facility to handle perturbations in the schedule well. Yan and Tu (1997) developed a framework in order to help carriers handle schedule perturbations resulting from the temporary closure of airports. Teodorovic and Stojkovic (1995), developed a model to reduce airline schedule disturbances. The heuristic model that was developed could facilitate the work of the dispatcher when making decisions regarding traffic management.

Brain et al. (2000) developed a simple variant of the basic fleet assignment models that assigned a time window to each flight and then discretized each window, allowing flight departure times to be optimized. They recognized that variability in scheduled flight departure times could result in improved flight connection opportunities and a more cost effective fleet assignment. They presented a generalized fleet assignment model for simultaneously assigning aircraft types to flights and scheduling flight departures. Bard et al. (2001) presented the time-band optimization model for reconstructing aircraft routings in response to groundings and delays experienced over the course of the day.

4.3 Applicability of AHFM in food industry

The example taken up in this section is in a totally different production environment – food, i.e., process, industry, where the focus will be on ice cream production. The basic steps in the production of ice cream are generally as follows: (1) blending of the mix ingredients, (2) pasteurization, (3) homogenization, (4) aging the mix, (5) freezing, (6) packaging, and (7) hardening as shown in Figure 4-1.

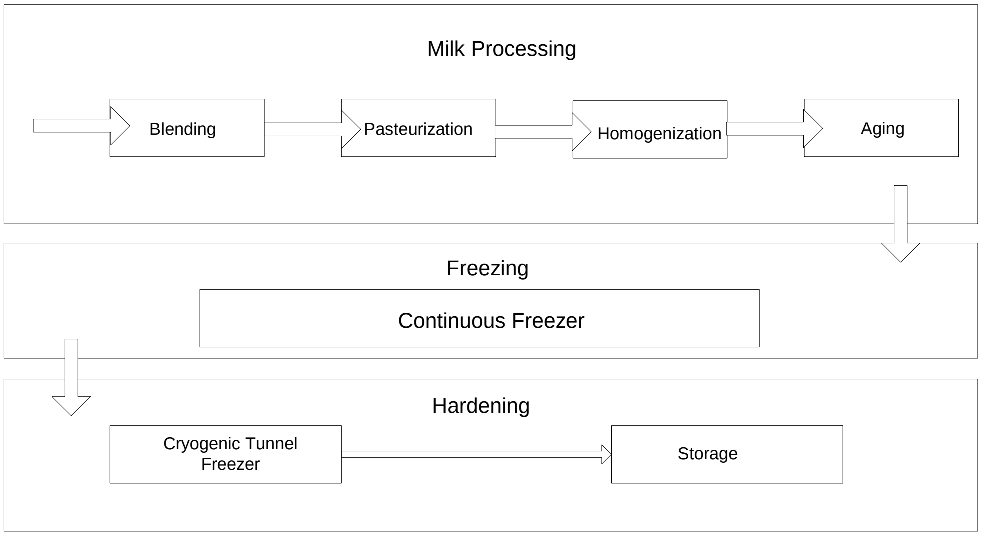


Figure 4-1 Commercial Ice-Cream Manufacturing Process

Figure 4-2 Process Flow Sheet for ice cream production

The production process can be represented by three steps as shown in Figure 4-2. The processing times and costs each of the blocks can be manipulated by varying

- The rate of pasteurization for the milk processing block
- The refrigerant flow rate in the freezer
- The dwell time and the temperature gradient in the hardening tunnel

In a typical industrial setup, there are several of these processing and freezing units running in parallel. Production planning in the process industry involves selection of the process units as well as the operating conditions of these units. The operating parameters of these processing units are again locally optimal and are usually determined by values supplied by the manufacturers of the unit. For example, a homogenizer in a dairy may be operated at 1000 l/hr as it is the recommended capacity of the unit for maximum efficiency. Most process industries are designed for a particular production capacity and all units are selected to give optimal performance at that production level.

Similar to the approach used in the P1 model, a P6 model can be formulated for determining the production capability of the process industry. The processing time required at different stage of production of a product can be used as parameters for the P3 model. Solution to the P6 model determines the optimal utilization of all resources in order to maximize the throughput rate of the production.

P6. Max $k_j \sum_i x_{ij}$ (4-1)

Subject to

$$k_j \sum x_{ij} = k_r \sum x_{ir} \quad (4\text{-}2)$$

$$\sum_i x_{ij} = \sum_l y_{lj} \quad (4\text{-}3)$$

$$\sum_i x_{ij} = \sum_k z_{kj} \quad (4\text{-}4)$$

$$\sum_j x_{ij} t_{ij} \le K_i \text{ , for all } I \quad (4\text{-}5)$$

$$\sum_j y_{lj} t_{lj} \le K_l \text{ , for all } l \quad (4\text{-}6)$$

$$\sum_j z_{kj} t_{kj} \le K_k \text{ , for all } k \quad (4\text{-}7)$$

$$x_{ij}\,,\; y_{lj}\,,\; z_{kj} \ge 0 \quad (4\text{-}8)$$

where x_{ij} is no. of gallons of ice-cream of type *j* processed at processing unit *i*, y_{lj} is number of gallons of ice-cream of type *j* processed at freezing unit *l*, z_{kj} is no. of gallons of ice-cream of type *j* hardened in the hardening tunnel section *i*, k_j is synchronizing coefficient between the ice-creams. K_i is capacity of processing unit *i*, K_l is capacity of freezing unit *l*, K_k is capacity of hardening tunnel section *k*.

The objective of the P6 model is to maximize the production of ice creams, subject to constraints of the process shop floor. Constraint (4-2) ensures that the production of the ice creams is synchronized. Constraints (4-3) and (4-4) ensure that material balance is maintained during various processing stages. Constraints(4-5), (4-6), and (4-7) ensure that the production capabilities of the units are not exceeded.

However, it is not unusual to hear that a process industry is operating at 130% or more of its design capacity. This is because most of the units can operate at higher production capacities than the book value supplied by the manufacturer. The strategy employed for increasing the capacity is similar to that on the manufacturing shop floor. The bottleneck unit is determined and it is operated at lower efficiency, but higher productivity, or is replaced by a unit of higher capacity. The productivity of the other units is adjusted accordingly. This leads to a shift in the bottleneck to a new unit. The procedure is repeated for continuous improvements.

Most of the industries take 72^{o} C and 15 second hold time as a fixed parameter for pasteurization. However detailed study shows that there is in fact a whole range of temperatures and holdup time we can select from. Similarly, the flow rates of the coolants in the freezer units (mechanical freezing systems) and hence the retention times are also variable. Besides this, the hardening tunnel based on cryogenic freezing technology too have a variable throughput rate that can be achieved by varying the belt loading and the cryogen flow rates. Therefore, in the ice-cream industry where pasteurization/aging/mixing form one set of aggregate processing, freezing and hardening the other two, we can identify the flow rate of the ice cream as the global system variable and write the time and the cost equations for processing units as follows:

$$t_{ijm} = \frac{a_{ijm}^{(t)}}{v^{\beta_{ijm}}{}_{ijm}} + b_{ijm}^{(t)} v_{ijm}^{\alpha_{ijm}} + c_{ijm}^{(t)} v_{ijm} \qquad m = i, l, k \tag{4-9}$$

$$C_{ijm} = \frac{a_{ijm}^{(C)}}{v^{\beta_{ijm}}{}_{ijm}} + b_{ijm}^{(C)} v_{ijm}^{\alpha_{ijm}} + c_{ijm}^{(C)} \tag{4-10}$$

Model P6 can be therefore be modified according to AHFM approach by combining substituting equations (4-9) in equations (4-5), (4-6), and (4-7) and converting the objective from maximizing throughput to that of maximizing profits. The modified model can therefore be written as

P7. Max

$$\begin{aligned} & pk_j \sum_i x_{ij} - \sum_{i,j} x_{ij} \left(\frac{a_{ij}^{(C)}}{v^{\beta_{ij}}{}_{ij}} + b_{ijm}^{(C)} v_{ij}^{\alpha_{ijm}} + c_{ij}^{(C)} \right) - \sum_{l,j} y_{lj} \left(\frac{a_{lj}^{(C)}}{v^{\beta_{lj}}{}_{lj}} + b_{li}^{(C)} v_{lj}^{\alpha_{lij}} + c_{li}^{(C)} \right) \\ & - \sum_{k,j} z_{kj} \left(\frac{a_{kj}^{(C)}}{v^{\beta_{kj}}{}_{kj}} + b_{kj}^{(C)} v_{kj}^{\alpha_{ikj}} + c_{kj}^{(C)} \right) \end{aligned} \tag{4-11}$$

Subject to constraints (4-2), (4-3), (4-4), and

$$\sum_{i,j} x_{ij} \left(\frac{a_{ij}^{(t)}}{v^{\beta_{ij}}{}_{ij}} + b_{ij}^{(t)} v_{ij}^{\alpha_{ij}} + c_{ij}^{(t)} \right) \le K_i \tag{4-12}$$

$$\sum_{l,j} y_{lj} \left(\frac{a_{lj}^{(t)}}{v^{\beta_{lj}}{}_{lj}} + b_{li}^{(t)} v_{lj}^{\alpha_{lij}} + c_{li}^{(t)} \right) \le K_l \tag{4-13}$$

$$\sum_{k,j} z_{kj} \left(\frac{a_{kj}^{(t)}}{v^{\beta_{kj}}{}_{kj}} + b_{kj}^{(t)} v_{kj}^{\alpha_{ikj}} + c_{kj}^{(t)} \right) \le K_k \tag{4-14}$$

Based on the data that is provided, P6 and P7 problems were formulated and solved using GAMS. Optimal solutions were obtained for both the problems. The book values were then inserted in P6 problem formulation to determine the optimal utilization of all resources in order to maximize the throughput rate of production and the corresponding profitability. P7, formulated on AHFM principle, was also solved and the results have been tabulated in Table 4-1.

Table 4-1 Resultant profits in the process industry example

P6 with handbook value (the lowest cost per product)	P6 with handbook upper bound	P7 (variable process parameters
\$255,794 (100.0%)	\$466,997 (182.6%)	\$471,030 (184.1%)

As can be seen from the results, there has been 84% increase in the profitability of the shop floor providing a huge opportunity for system optimization. The parameters that can be optimized using the integrative models are processing times, raw material/concentration, flow rates, etc. Parameters that directly impact process speed, such as flow-rates, conveyor speeds, etc are the obvious choice here.

4.4 Airplane speed variation in fleet scheduling

The AHFM principles are applied to a transportation problem in this section. In most of aircraft fleet scheduling models, the aircraft speed is assumed to be a given fixed parameter. This is usually the most economic speed, i.e., that which reduces fuel consumption per mile. However, most airplanes can fly at higher speeds (at least 5% to 25%) depending of the type of aircraft. Of course, higher speed results in higher costs, sometimes significantly higher, up to 50% from costs of the most economical speed. For the preliminary check of the concept we extended a basic daily fleet scheduling problem that can be formulated as follows: given (a) the size of fleet of each aircraft type, (b) daily schedule of flights between airports, (c) costs associated with an assignment of particular aircraft type to particular flight, (d) penalty costs of unscheduled flights, and (e) maintenance time for each aircraft type in each airport, what is the optimal assignment of airplanes to flights that minimizes the total cost. When the schedule is feasible, the penalty cost should be high enough to ensure assignment of all flights. However, when the schedule is infeasible for given fleets, the penalty cost can directly affect the optimal solution. The model is solved as a minimum cost network flow problem where nodes represent airports in each time interval and arcs represent scheduled flights and dwells (where airplanes are on the ground and remain at the same airport until the next time interval) as shown in Figure 4-3 (a). There are costs associated with each flight by each fleet type and penalties for unscheduled flights. In addition there is maintenance "turn" time for preparing airplanes to the next flight. The P8 problem can be formulated as:

P8 Min

$$\sum_{\{c,c_1,t,t_1,f\}\in A} C(c,c_1,t,t_1,f)\cdot x(c,c_1,t,t_1,f) + \sum_{\{c,c_1,t,t_1\}\in F} P(c,c_1,t,t_1)\cdot\left(1-\sum_{f\mid\{c,c_1,t,t_1,f\}\in A} x(c,c_1,t,t_1,f)\right) \quad (4\text{-}15)$$

Subject to

$$\sum_{f\mid\{c,c_1,t,t_1,f\}\in A} x(c,c_1,t,t_1,f)\le 1 \qquad \forall\{c,c_1,t,t_1\}\in F \qquad (4\text{-}16)$$

$$x(c,c_1,t,t_1,f)=x(c_2,c_3,t_2,t_3,f) \qquad \forall f,\{c,c_1,t,t_1,c_2,c_3,t_2,t_3\}\in H \qquad (4\text{-}17)$$

$$y(f,c,t)=y(f,c,t-1) \qquad \forall\{f,c,t\}\notin G \qquad (4\text{-}18)$$

$$\begin{aligned} &y(f,c,t-1)+\sum_{c_1,t_1\mid\{c_1,c,t_1,t-M(f,c),f\}\in A} x(c_1,c,t_1,t-M(f,c),f)= \\ &y(f,c,t)+\sum_{c_1,t_1\mid\{c,c_1,t,t_1,f\}\in A} x(c,c_1,t,t_1,f) \qquad \forall\{f,c,t\}\in G \end{aligned} \qquad (4\text{-}19)$$

$$\sum_{c,c_1,t,t_1\mid\{c,c_1,t,t_1,f\}\in O} x(c,c_1,t,t_1,f)+\sum_{c\in C} y(f,c,\text{"23:59"})\le N(f) \quad \forall f \qquad (4\text{-}20)$$

$$x(c,c_1,t,t_1,f)\in\{0,1\} \quad \forall\{c,c_1,t,t_1,f\}\in A \qquad (4\text{-}21)$$

$$y(f,c,t)\in \text{int} \quad \forall f,c,t \qquad (4\text{-}22)$$

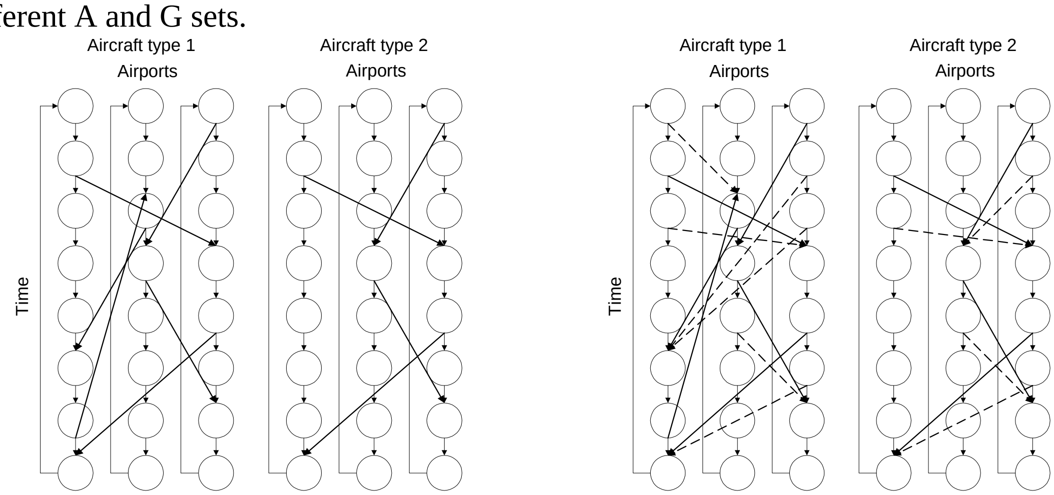
where f is aircraft type; c, c_1, c_2, c_3 are airports; t, t_1, t_2, t_3 are times; A is the set of all (eligible) flight-aircraft arcs (c, c_1, t, t_1, f) for flights that leave city c at t and arrives in c_1 at t_1 using aircraft of type f; G is the set of all nodes with arcs from set A, F is the set of all flights; H is the set of all flights that must use the same plane; $M(f, c)$ is the maintenance time for aircraft of type f in c; O is the set of overnight flights, i.e., flights (c, c_1, t, t_1) such that $t > t_1 + M(f, c)$; $N(f)$ is the size of the fleet of aircrafts of type f; $C(c, c_1, t, t_1, f)$ is the cost of flight (c, c_1, t, t_1) using aircraft of type f; $P(c, c_1, t, t_1)$ is penalty cost for unscheduled flight (c, c_1, t, t_1); $x(c, c_1, t, t_1, f)$ is a binary variable that is equal to one if flight (c,c_1,t,t_1) is scheduled with aircraft of type f; $y(f, c, t)$ is an integer variable that is equal to the number of planes of type f to be on the ground in airport c in time period [t, t + 1]. The objective function (4-15) summarizes the costs. Constraint (4-16) ensures that each flight is assigned at most once, constraint (4-17) ensures assignment of the same aircraft for continuation flights, constraint (4-18) balances nodes with dwell arcs only, and constrain (4-19) balances other nodes with additional flight arcs. Constraint (4-20) is the capacity constraint – it ensure that in midnight the total number of aircraft of each type is less or equal to the fleet size; combined with balancing constraints it guarantees appropriate number of aircraft of each type. Constraints (4-21) and (4-22) are binary and integer constraints on decision variables.

Faster flights mean more "flight" arcs with higher costs and additional constraint that only one flight should be done for each scheduled flight as shown in Figure 4-3 and presented in model P6. Note, sets A and G are larger than in model P9.

P9 Equation (4-15)
Subject to
Constraints (4-16), (4-18), (4-19), (4-20), (4-21), (4-22)

$$\sum_{t_4 \in S(c,c_1,t,t_1,f)} x(c, c_1, t_4, t_1, f) = \sum_{t_5 \in S(c_2,c_3,t_2,t_3,f)} x(c_2, c_3, t_5, t_3, f) \quad \forall f, \{c, c_1, t, t_1, c_2, c_3, t_2, t_3\} \in H \tag{4-23}$$

where $S(c,c_1,t,t_1,f)$ is set of feasible taking off times when flight speed is increased. Equation (4-23) replaces Equation (4-17), all other constraints remain the same but with different A and G sets.



(a) Basic problem (b) Flight speeding problem

Figure 4-3 Time-space network for fleet scheduling problem

A small-scale model of three cities and three fleets was taken from the AMPL test problem site at http://www.ampl.com/cm/cs/what/ampl/NEW/ILOG/ (Airline fleet assignment problem, files fleet2i.mod and fleet2.dat). It was solved using GAMS interface and OSL solver. In this problem the flight schedule is feasible, i.e., all flights can be scheduled. In the parameter variation model we assume that the flight times can be reduced by 10 minutes with costs increased by 15% or reduced by 20 minutes with costs increased by 50%. As shown in
Table 4-2, the total cost was reduced by 25% by allowing speeding the flights. If we reduce the fleet size making the flight schedule infeasible, e.g., due to technical failures, flight speeding can help keeping the schedule or, at least, reduce the costs as shown in Table 4-2.

Table 4-2 Results for the fleet assignment problem

	Fleet sizes	Traditional approach	Parameter variation
Original problem	72S-6, 73S-6, L10-2	2068.00 (100%)	1576.00 (76%)
Infeasible problem 1	72S-3, 73S-3, L10-1	2353.00 (100%)	1613.65 (69%)
Infeasible problem 2	72S-1, 73S-1, L10-1	3038.00 (100%)	2236.85 (74%)

4.5 Discussion

To summarize, process speeds were optimized in three totally different areas. In machining example we get improvement up to 59% compared to models where “book” values used to set parameters for process times. In food industry the gains are up to 84% where compared to the standard pasteurization parameters. In fleet scheduling the potential gain of speeding the airplanes can reach 25%. The conversion of parameters to variables adds flexibility to the models. In fact, preliminary results of the application of high fidelity modeling to the airline fleet scheduling problem indicate that the technique yields feasible schedule to problems that were initially considered infeasible by hierarchal planning procedure.

However, the aggregation of the production planning models get more and more complex as we increase the number of hierarchies that the aggregate model covers. It is essential to limit the complexity of the problem with an emphasis on the tractability and computational complexity. Analysis of special properties of the models, such as unimodal regions, is required. A factorial experimental design could be used to determine potential gain of parameter variation in different operational levels and implementation scale of the standard optimization packages such as CONOPT, MINOS, and SNOPT for non-linear optimization, OSL and CPLEX for mixed integer linear models, and DICOPT for mixed integer non-linear models.

Implementation of the developed concept in real-life environments may require development of heuristic algorithms that can deal with large-scale models. Based on our small but realistic examples the opportunity gap is large enough to justify “optimal versus heuristic” losses. The gap may even increase with problem size: the balancing effect of parameter variation is more important in complex manufacturing environments with multiple products, alternative process plans and shifting bottlenecks.

4.6 Conclusions

In this chapter, the extension of the AHFM approach for integrated planning in a variety of industries was explored. Using a small set of problems, it has been demonstrated that the model formulation, although not generic, can be easily modified to accommodate a variety of system specifics, i.e., manufacturing, processing and travel. The problems that have been illustrated reflect the actual characteristic of such systems, but the size of the problems was reduced for illustration. A major part of the focus of this chapter was to uncover the potential savings that may be gleaned from such a technique. Because the opportunity (increased profitability from 25 – 85% for realistic small scale problems), there is a strong indication that heuristics for large-scale problems should be developed as the scale exceeds the capacity of today's solvers. The potential savings for integrating the planning domain are too large to ignore this critical aspect of the problem.

Chapter 5

Sensing the delta deviation in logistic performance and the integrated sense-response system

5.1 Introduction

Planning requires estimates of A, b, and C, while replanning is based on information about A', b', and C'. Both the planning and replanning processes however implicitly assume that the information of the state variables and the transition in states is available. For decision making where human is involved, this assumption may be valid, but for automated planning systems, there is a need to set the stage so that this information is made available to the planning systems. “Sensing” ability provides this information and capability to the planning systems. The role that sense plays is presented in Figure 5-1.

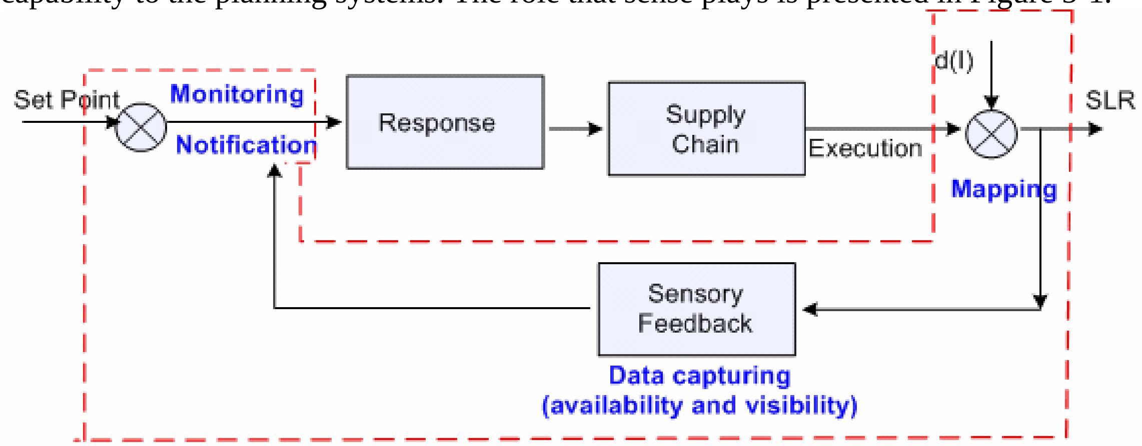


Figure 5-1 Role of Sense in EM

There usually are four main components in sense:

1. Mapping
2. Data capturing
3. Monitoring
4. Notification

Mapping provides the relationship between the SLR and the state variables A, b, and C. Data capturing refers to the physical data capturing and data transformation aspects of sensing. The other two components involve monitoring of the deviation between the planed and the executed, and notifying the response system of the deviation, if in case there is any.

From a time persistent perspective, the monitored variable (deviational variable) should be indicative of the occurrence of the step change that the event causes. The transition from A, b, and C, to A', b', and C' can be represented in the general form

$$y_t = \mu + \frac{\theta(B)}{\phi(B)} a_{t0} + \gamma f_{E_i}(t) \tag{5-1}$$

Where y_t is the observed state variable (deviational variable) and $f_{E_i}(t)$ represents the specific nature of the effect that the event E_i and has on the observed variable y_t. $f_{E_i}(t)$ can be of the form of a step function or ramp, or take any other form such as the hump and the spike functions.

$$f_{E_i}(t) = \begin{cases} 0 & 0 < t < t_0 \\ 1 & t \geq t_0 \end{cases} \tag{5-2}$$

$$f_{E_i}(t) = \begin{cases} 0 & 0 < t < t_0 \\ kt & t \geq t_0 \end{cases} \tag{5-3}$$

The transition in case of the step function is sudden while it is gradual in case of a ramp function. The objective of implementing the sensory system is early detection of the transition (onset of t_0) and notifying the response system.

This chapter proposes the following:

1. Transformation tables that related the SLR to the state variables
2. Control charts for monitoring the deviational variables

The use of the control charts is proposed for monitoring the step change as this is more or less the industry norm. However, the economic perspective to the design is taken as opposed to the more standard statistical perspective. An economic model is formulated for a Shewhart chart, with most of the assumptions that are used for the economic design (for instance, assumptions related to the characteristics of the process and the events) being more or less the same as available in literature (process is normally distributed; the time between events is exponentially distributed, and so on). Modifications are however made to include the special characteristics of impact that the events have in the business systems: namely that there is a cost associated with the detection delay and the cost is monotonically increasing. Any detected exception triggers the response mechanism.

However, when the transition is gradual, the monitored variable displays a trend when the deterioration sets in. To monitor a trended process, control charts with traditional interpretations are not suitable, since the basic i.i.d. assumption required in traditional chart design are violated. Any control chart that aims to respond to changes of process mean becomes inappropriate. Not only that, an additional estimate of the trend components is also required. In this chapter, the two factors are integrated into the economic design developed for the Acceptance control chart (ACC). Duncan (1986) introduced the ACC which responses only to crossover of predetermined control limits due to linear trend.

When the ACC is used, the decision factor is when the response should be initiated. The integration of the response mechanism to the detected anomaly in case of the ACC is also discussed in this chapter.

5.2 Background

Based on the description of the planning problem and the kind of impact that the event has on the planning problem and the parameter/variable that they affects, the events can be classified into five categories:

1. Those that change in objective coefficient (c_k) of a basic variable
2. Those that change the right hand side (RHS) coefficient (b_j)
3. Those that change the constraint coefficient (a_{ij})
4. Those that add a new constraint.
5. Those that add a new variable in the system.

The occurrence of the event is detected by monitoring some indicator variable. The indicator variables can therefore be classified into the corresponding 5 types. Monitoring the indicator variables, early detection of the mean shift, notification of the occurrence of the shift are critical issues for the implementation of the EM system.

Several issues come up when we discuss monitoring, some of the critical ones being:

1. What to monitor, i.e., is the data series that we are monitoring an indicator of the mean shift that the event causes.
2. Is the monitored variable a good indicator?

There is a lot of literature and statistical analysis that provides the answer for these questions. Since, we define an event as the cause/consequence in the mean shift, and model it that way, the time series should show a detectable mean shift when the event occurs. A good indicator provides the opportunity (if it exists) to predict the event, or detect the occurrence early. We will assume for the rest of this thesis that the variable that we are monitoring can be used to detect the mean shift, and is a good indicator.

Given that we know what to monitor, the next question is whether we are trying to compress data, (wherein we are combining the analysis of several different time series we with the idea of saving on the cost of the data collection and analysis but potentially loosing some predictability and detectability properties) or are we doing a data expansion (wherein we are using the same data, and looking into different aspects of it, like the integral, the derivative, and so on with the idea of improving the predictability and detectability of the time series with context to a specific event). The selected approach depends on the severity of the event and the cost of data collection. For the first case (data compression), there is a lot of work done in the area of multivariate statistics and multivariate control charts with PCA, FA, DA, etc. For the second (data expansion), we can get loads of information about projections and norms, and similar concepts from EPC literature. For the rest of the thesis, we will assume that we are operating on the transformed data. Planning requires estimates of A, b, and C, while replanning is based on information about A', b', and C'. These may or may not be directly related to the SLR that the company is interested in and is monitoring. Mapping functions are therefore required. The mappings may be linear or non-linear. It is desirable that the mapping be reliable.

The actual implementation of the sensing approaches is based on the perspective that we take on the EM problem. Here we present two such perspectives and the corresponding definitions. One perspective on EM focuses on addressing and eliminating the "cause" of the mismatch between the planned and the executed. This perspective is inspired from the SPC and quality control (QC) concepts, where the emphasis is on monitoring performance indicators (PIs) and the common cause variance therein, and successively eliminating the assignable cause variances. Further, as far as this approach is concerned, the root cause or consequence of a mean shift in one or more of the monitored parameters and variable is termed as an "event". The second perspective aims at addressing and eliminating the "effect" that the mismatches and events have on the PIs. This approach is analogous to enabling a tracking controller (which falls within the realms of the so called engineering process control (EPC) approach) with disturbance rejection and uncertainty handling capabilities. The cause or the consequence of coloring of the disturbances can be termed as "events". Whatever be the perspective that we take, the effect that the event has is more or less the same. It causes a Δ change in one or more parameters/variables. It is the scientific way of handling this Δ change that EM is all about. The focus of this chapter is on the statistical interpretation and use the statistical control charting based approach for monitoring, detection and notification of events and their occurrence.

Control charts are widely used for controlling industrial processes. It is an on-line statistical process monitoring tool which has been widely used to monitor the process mean and analyze the process capability. There are several types of control charts and a good discussion of the types and characteristics can be obtained from Montgomery (2000). One of the more popularly used control charts, the Shewhart or the x-bar chart is generally established under the assumption of normality are placed at $\pm\partial\sigma$ away from the process mean μ_0 ; where $\partial > 0$ and σ is the standard deviation of the process. After the process is run, samples of size n are taken from the output at an interval of every h time units. The values of from these samples are then sequentially plotted on the chart in order to show whether they fall within the control limits. If the sample average is outside the control limits or if the data exhibit a suspicious pattern, the process is regarded as out of control. The management then searches for the assignable cause and takes corrective actions to restore the process to the in-control state. However, the process can continue or can be stopped during the search for the assignable cause. A detailed study of the interpretation of Shewhart charts was carried out by Nelson (1985). It is particularly important that the use of a control chart requires choosing a sample size (n), a sampling interval (h), and the control limits (in terms of a multiple of the standard deviation of the sample mean, k, that is, $k = \delta\sqrt{n}$) for the chart. Selection of these parameters is usually called the design of the control chart.

However, traditional applications of SPC are based on the assumption of process stability that is violated in many cases. When the transition from an in-control state to an out of control state is gradual, as is the case when $f_{E_i}(t)$ is a ramp function, control charts with traditional interpretations are not suitable, since the basic i.i.d. assumption required in traditional chart design has been violated. Any control chart that aims to response to changes of process mean becomes inappropriate. Many manufacturing processes exhibits certain trend patterns that can be modeled adequately by a regression or time series model together with independent noise terms. A typical example is the tool-wear problem, as addressed by Quensenberry (1985). Duncan (1985) introduced the ACC, which responses only to crossover of predetermined control limits-and acceptance control charts will be used here. After Duncans work in 1985, many researchers, such as Wu (1998) studied application of acceptance control chart in various scenarios, and suggested that acceptance control chart is useful especially for the gradual deterioration case. Under the statistical process monitoring, the deteriorating factors, such as tool-wear and material consumption, act as common causes instead of special causes since the process will develop in a predictable manner. For trended and regularly adjustment process, which is referred to as trend-adjusted process, an important decision variable is the adjustment frequency, or the average adjustment interval (AAI). AAI is estimated together with the mean square deviation.

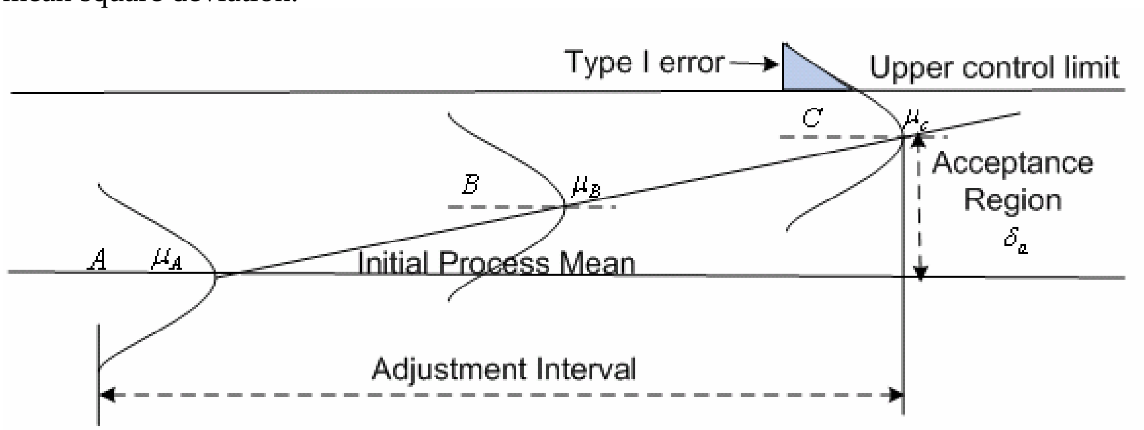


Figure 5-2 Setup for an acceptance control chart

The type I and type II error for such acceptance control chart are given by

$$\alpha_i = \Pr\left(Y_i > UCL \mid \mu_{yi} > \delta_a\right) = \phi\left(\frac{\hat{Y}_i - UCL}{\sigma}\right) \tag{5-4}$$

And

$$\beta_i = \Pr\left(Y_i < UCL \mid \mu_{yi} > \delta_r\right) = \phi\left(\frac{UCL - \hat{Y}_i - \delta_r + \delta_a}{\sigma}\right) \tag{5-5}$$

where

$$\hat{Y}_i = T + f(i) \tag{5-6}$$

is the predicted outcome of Yi at time instance i. In Tang et al. (2004), modified Type-I and Type-II errors for trended process within a renewal cycle I are proposed. They are based on the averaged alarm probabilities, as each alarm probability is dependent on the number of units since the latest adjustment. These average alarm probabilities will be used in the economical optimization of adjustment interval. They are computed as

$$\alpha_a = \sum_{i=1}^{I} \phi \frac{\left(\hat{Y}_i - UCL\right)/\sigma}{I} \tag{5-7}$$

$$\beta_a = \sum_{i=1}^{I} \phi \frac{\left(UCL - \hat{Y}_i - \delta_r + \delta_a\right)/\sigma}{I} \qquad (5\text{-}8)$$

Simple linear regression model is commonly used to model the trend in the trended process, such as the tool wear in Quesenberry (1988). The model that is considered here can be represented as:

$$Y_i = T + f(i) + \varepsilon_i \qquad (5\text{-}9)$$

where Yi represents the mean value of observed process output characteristic at time index i, T is the target value of process output, i is a sequence of normally and independently distributed quantity with mean μ and standard deviation σ , and f(i) denotes any type of monotonous and predictable deterioration, such as a liner drift.

The problem with the commonly used statistical approach to control chart design is that it is used in almost all processes as the standard procedure for implementing control charts, without regard to the cost consequences of the design. In EM, the main thing that is of interest is the economics of the whole issue. In order to overcome similar shortcoming, a number of researchers have proposed economic models for the design of control charts. In this chapter, an economic design of control charts approach is used to determine an optimal value of AAI, while the process quality is monitoring with a statistical control chart.

5.3 Economic design of control charts

The early research work on the economic design of control charts was initiated by Duncan (1956). Later, a large number of subsequent articles which extended Duncan's model were fruitful in the field of the economic design of Shewhart control charts. Lorenzen and Vance (1986) and Woodall (1986) discussed the advantages and disadvantages of using the economic design of control charts. Vance (1983) presented a bibliography of control chart techniques for the years 1970-1980. A literature review of control charts from 1981 to 1991 can be found in Ho and Case (1994).

The basic idea behind the economic design is to look at the costs incurred due to monitoring, sampling, diagnosing and dealing with the type I and type II errors while taking decisions regarding the choice of sample size (n), sampling interval (h), and the control limits ($k = \delta\sqrt{n}$) for the chart. If the control limit is too high, the probability of type I error is high as there is a higher probability of failure in detection of small changes. If in case it is too small, there is a higher probability of occurrence of type II error and therefore false alarms; in automated systems, false alarms are expensive as well. There therefore exits a tradeoff between the cost incurred in dealing with the false alarm, and the time delay observed in detecting the true alarm. Other costs and tradeoffs also exist. There are four basic components in the cost function.

1. The cost when the process is in control
2. The cost when the process is out of control (delay in detection and correction)
3. The cost for data capturing and processing
4. The cost for a false alarm

The objective of the economic designs is to determine the design parameters (i.e. n, h, and k) that minimize the expected cost or maximize the expected net-income per unit time.

To derive the expected net-income or cost functions, the following assumptions generic to most applications of Shewhart chart are made and have been included in this work as well:

A1. The distribution of the quality characteristic of the process output is normal. Applying the central limit theorem, this result is still approximately correct for the Shewhart chart even if the underlying distribution is non-normal.

A2. The Shewhart control chart is used to maintain current surveillance of a process mean which is subjected to an assignable cause.

A3. The process is started in the in-control state with mean μ_0 and standard deviation σ. The occurrence of the assignable cause results in a shift in the process mean from μ_0 to $\mu_0 + \delta\sigma$; where $\partial > 0$. When the shift of the process mean is identified by the ~ control chart, it is then restored to the "in-control" state by repairing and eliminating the cause.

A4. The occurrence time of an assignable cause follows an exponential distribution with parameter $\lambda > 0$ (thus, $1/\lambda$ is the mean time that the process remains in the in-control state).

A5. The process is monitored by drawing random samples of size n at times h, $2h$, $3h$, and so forth. The control limits of the Shewhart chart are set at $\mu_0 \pm k\sigma/\sqrt{n}$, where k is a multiple of the SD of the sample mean. In contrast to k, δ in assumption (A3) is the multiple of the SD of the process, where $\partial = k/\sqrt{n}$.

A6. When a single point falls outside a control limit, it is an indication that the process is out of control, and then a search for the assignable cause is made.

In general, the sooner the occurrence of an event is detected, the lower is the cost of rectifying it. On the other hand, the longer the detection delays, the more are the costs. Also, it has been shown that as time progresses, the complexity of the EM problem increases. Therefore it is assumed that the cost of rectification is dependent on detection delay. These assumptions are critical and need to be added to the existing assumptions while determining the parameters for the control chart.

5.3.1 Economic design of Shewhart chart

According to assumptions (A1), (A3), and (A5), the probability of a point falling outside a control limit when the process remains in the in-control state (type II error) is

$$\alpha = 2\phi(-k) \tag{5-10}$$

Similarly, when an assignable cause has occurred, the probability that it will be detected on one of the subsequent samplings is.

$$1-\beta = \phi\left(\partial\sqrt{n} - k\right) + \phi\left(-\partial\sqrt{n} - k\right) \tag{5-11}$$

From assumption (A4), the expected time of occurrence of the shift between the j^{th} and $\left(j+1\right)^{th}$ samples is

$$\tau = [1-(1+\lambda h)\exp(-\lambda t)]/[\lambda\{1-\exp(-\lambda t)\}] \qquad (5\text{-}12)$$

The expected number of false alarms that will occur before a shift is α times the expected number of samples taken before the shift. That is

$$\alpha \sum_{j=0}^{\infty} j \int_{jh}^{(j+1)h} \lambda\, exp(-\lambda t)dt = \frac{\alpha\, exp(-\lambda h)}{1-exp(-\lambda h)} = \frac{\alpha}{exp(\lambda h)-1} \qquad (5\text{-}13)$$

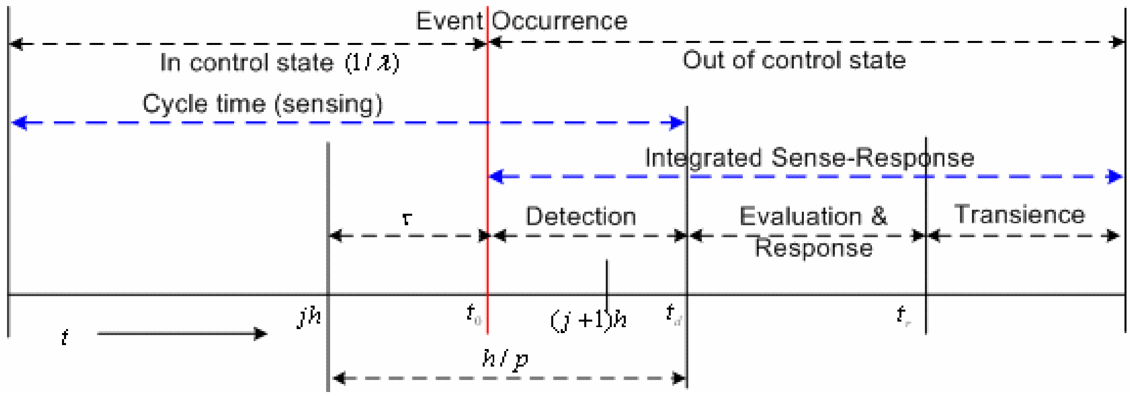


Figure 5-3 Detection delay of an event and the manufacturing cycle

The production cycle lengths for the models can be derived from the following time periods:

1. The expected time during which the process will be in control is $1/\lambda$. Note that this expected time includes τ.
2. The expected time from the occurrence of the assignable cause to the detection of a signal is $(h/p)-\tau$ Since the number of samples taken before the shift in the process is caught is a geometric random variable with mean $1/(1-\beta)$ and the length of time between samples is h, the expected time during which the process will be out of control before a signal is detected is $(h/p)-\tau$.

Thus, the expected cycle time for the models is:

$$E(t) = (1/\lambda) + [(h/p)-\tau] \qquad (5\text{-}14)$$

An income function can be regarded as a negative cost function. Therefore the net hourly income while the process is out of control can be set as a quadratic (loss) cost function. The expected net incomes of a cycle for the model can be obtained from the following four terms:

The cost when the process is in control

$$(V_0/\lambda) + \sum_{j=0}^{\infty} \int_{jh}^{(j+1)h} V(T-t)[(T-t)\lambda\, exp(-\lambda t)]dt \qquad (5\text{-}15)$$

Where t is the point of time at which the process changes its state from the in-control state to the out of control state since the last sampling time, jh, for $j = 1,2,...$ Note that the constant V_0 is the net hourly income which the process is in control, and $V(T-t)$, on the other hand is the net hourly income while the process is out of control, as a function of detection delay $(T-t)$.

The cost for data capturing and processing

$$bE(T')/h + n.e \text{ where } E(T') = (1/\lambda) + [(h/p) - \tau] \tag{5-16}$$

The cost of examining false alarms per cycle

$$\frac{\alpha Y}{\exp(\lambda h) - 1} \tag{5-17}$$

The cost of rectifying an assignable cause per cycle

$$\sum_{j=0}^{\infty} \int_{jh}^{(j+1)h} Z(T_2 - t)\lambda \exp(-\lambda t) dt \tag{5-18}$$

Similar to $V(T-t), Z(T-t)$ is the cost of rectification as a function of detection delay $(T-t)$. Combining the cost equations, we get the net income as:

$$\begin{aligned} E(NI) &= (V_0/\lambda) + \sum_{j=0}^{\infty} \int_{jh}^{(j+1)h} V(T-t)[(T-t)\lambda \exp(-\lambda t)]dt \\ &- \frac{\alpha Y}{\exp(\lambda h) - 1} - bE(T')/h - \sum_{j=0}^{\infty} \int_{jh}^{(j+1)h} Z(T_2 - t)\lambda \exp(-\lambda t) dt \end{aligned} \tag{5-19}$$

wherein

$$\tau = [1 - (1 + \lambda h)\exp(-\lambda t)]/[\lambda\{1 - \exp(-\lambda t)\}] \tag{5-20}$$

5.3.2 Economic design of acceptance charts

The main design factor that needs to be considered in the economic study is the average length of adjustment interval. Let n_1 and n_2 be defined as the average number of units produced during 'in-control' state and 'out-of-control' state for each cycle, respectively, and n_f the average number of false alarms per cycle. Since the 'in control' time is exponentially distributed with parameter, the expected number of n_1, n_2, and n_f can be derived as below:

$$E(n_i) = \sum_{i=0}^{\infty} i(\exp(-\lambda i/v) - \exp(-\lambda(i+1)/v)) = 1/(\exp(\lambda/v) - 1) \tag{5-21}$$

And the average number of false alarms per cycle is

$$E(n_f) = \alpha_a/(\exp(\lambda/v) - 1) \tag{5-22}$$

For the out-of-control state, it is assumed sample size equals to one, it is clear that number of units produced during this state equals to average run length (ARL) of the supervising acceptance control chart, which is directly relevant to Type II error:

$$E(n_2) = 1/(1 - \beta_a) \tag{5-23}$$

$$E(N) = E(n_1 + n_2) = \frac{1}{\exp(\lambda/v) - 1} + \frac{1}{1 - \beta_a} = \frac{\exp(\lambda/v) - \beta_a}{(\exp(\lambda/v) - 1)(1 - \beta_a)} \tag{5-24}$$

Summation of the above equations forms the overall process expected unit cost. The optimal adjustment interval can then be obtained by minimizing E(C), and the optimization formula for the expected total cost is

$$E(C) = \frac{C_0/(\exp(\lambda/v)-1) + C_1/(1-\beta_a) + Y\alpha_a/(\exp(\lambda/v)-1) + W}{\exp(\lambda/v) - \beta_a/(\exp(\lambda/v)-1)(1-\beta_a)} + a + \frac{C_a}{I} \quad \text{(5-25)}$$

$$= \frac{[(C_0 - C_1) + Y\alpha_a + W(\exp(\lambda/v)-1)](1-\beta_a)}{\exp(\lambda/v) - \beta_a} + a + C_1 + \frac{C_a}{I} \quad \text{(5-26)}$$

For the acceptance control chart, Type-I and Type-II error can be calculated. Other parameters in this optimization equation should be inherent to specific process at hand and estimated via past data. $\hat{Y}_i$ appears to be precondition for calculation of Type-I and Type-II error, previous knowledge on the process or some on-line forecast thus becomes necessary for the modeling. The procedure of deriving the cost-optimized for a trend-adjusted process can thus be summarized as follows:

1. Process modeling or forecasting of $\hat{Y}_{i+1}$ at instance i
2. Parameter estimation based on past data
3. Calculation of Type I and Type II error.
4. Calculation of the expected cost E(C) for instance i
5. If E(C) at time point I starts to exceed E(C) at time point i-1, adjustment is called for, and I=i

Because the process model is monotone, values of modified Type-I and Type-II error will also be monotonous. Thus a 'U' shaped E(C) curve is guaranteed and a minimum exists.

5.4 Implementation details: Case Study

Implementation of the EM approaches can be done at the strategic, tactical, and operational levels. The implementation at the tactical levels deal with the inclusion of risk analysis concepts in strategic plans. This is done during the technology road-mapping, and budgeting phases. A sample approach is presented for TPS in Section 6.2. At the tactical level, EM is integrated with the P&E systems. This can then be integrated for applications at the strategic level, or differentiated for applications at the operational level. A detailed analysis and application of the EM approach-Marketing Engineering, is presented in Section 6.3. Finally, the conclusions and general observations are presented in Section 6.4.

5.4.1 The IBM I-Series Supply chain: Problem setting

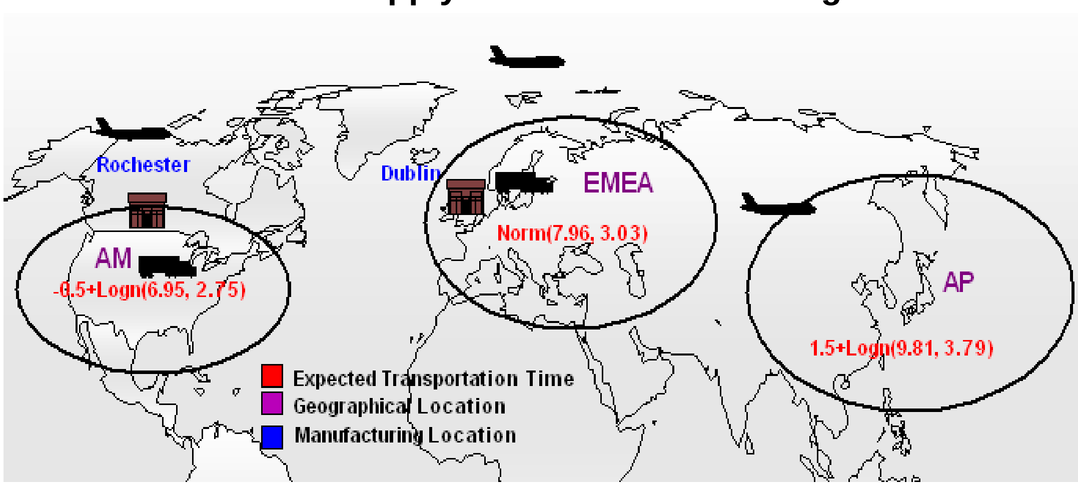


Figure 5-4 Structure of I-Series Supply Chain

The IBM I Series supply chain is structured as follows: There are two manufacturing units and three markets that are represented by AM, AP, and EWMA. These are geographically distributed as shown in **Error! Reference source not found.**. Fifteen model features are required to be made available at the three markets. The historic demand distribution for each of these models in the markets is available. The distribution of the transportation lead times to the markets from the two manufacturing units is also known, and is provided in the tables that follow. The objective of the planning and the execution systems is to maximize profits while satisfying the market demand.

The SLRs that IBM is currently focusing on are related to:

1. The average periodic inventory level (PIL)
2. The average work in process (WIP)
3. Stock out fraction (SOF)
4. Average backorder level (ABL)
5. Average backorder size per customer (ABS)
6. Probability of backorder per customer arrival (BCA)
7. Backorder fraction (BF)
8. Average transportation cost (ATC)

The state variables that have been selected to (on the basis of I Series data) are

1. Supplier Lead Time (SLT)
2. Demand distribution (DD)
3. Transportation time (TT)
4. Production lead time (PLT)
5. Planning discrepancy (PD)

5.4.1.1 Simulation Model

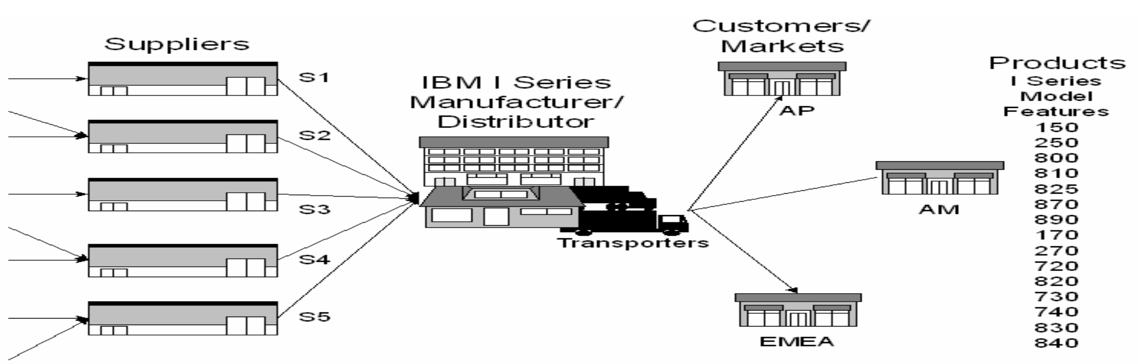


Figure 5-5 Simulated structure of ISeries supply chain

The demand distribution taken for analysis (based on data provided by IBM):

Table 5-1 Analysis of demand distribution using historical I Series performance data

			Demand

Serial No	Market Id	Product Id	Demand			
			Mean	Variance	Fitted Distribution	Std. Error
1	AM	250	5.69	1.44	Unif(4,8)	0.0189
2	AM	270	79.5	38.1	49+85*Beta(0.0517,0.092)	0.0771
3	AM	810	58.2	29.2	35+Weib(12.7,0.442)	0.0989
4	AM	830	18.6	6.49	Tria(4,24.8,27)	0.0247
5	AP	250	2.62	0.768	2+Gamm(2.68,0.23)	0.1114
6	AP	270	18.8	12.3	10+Weib(4.16, 0.438)	0.0964
7	AP	810	21.8	9.53	Tria(6,14.7,35	0.0994
8	AP	830	5.92	2.18	4+expo(1.92)	0.1240
9	EMEA	250	93.5	50.1	53+113*Beta(0.048,0.092)	0.0609
10	EMEA	270	77.3	38.4	49+86*Beta(0.035,0.0723)	0.0315
11	EMEA	810	55.6	28.4	33+64*Beta(0.0573,0.105)	0.0942
12	EMEA	830	13.1	2.36	Unif(10,16)	0.1372

The transportation details are as follows:

Table 5-2 Analysis of Transportation delays and lags using historical I Series performance data

Serial No	Market Id	Product Id	Transportation			
			Mean	Variance	Fitted Distribution	Std. Error
1	AM	250	6.76	4.27	3.5+Weib(3.04,0.887)	0.1009
2	AM	270	5.92	1.09	2.5+Erla(0.263,13)	0.3726
3	AM	810	7.06	4.77	-0.5+Logn(7.54,3.38)	0.2232
4	AM	830	6	1.21	Tria(3.5,6,8.5)	0.2242
5	AP	250	13	0.0	Unif(13,13)	0.0000
6	AP	270	11.5	5.11	Pois(11.5)	0.1877
7	AP	810	11.1	1.52	Norm(11.1,1.51)	0.0615
8	AP	830	11.0	0.0	Unif(11,11)	0.0000
9	EMEA	250	8.88	1.91	1.5 + Weib(7.54,6.13)	0.1848
10	EMEA	270	7.75	4.72	2.5+Logn(5.39,5.63)	0.1303
11	EMEA	810	7.22	3.36	Pois(7.22)	0.9422

A variance of 10%, 20%, and 50% was superimposed on the data and distribution

5.4.1.2 Mapping function

For determining the relationship between the SLRs and the observed variables, two mapping functions are determined: Linear regression and the non linear SVM based mapping. The mapping derived from the non-linear SVM is better , therefore the accuracy is higher.

Table 5-3 Analysis of Demand Distribution using historical I Series performance data

Serial No.	Metrics	Std. Error Support Vector Regression	Std. Error Linear Regression	Percentage Difference
1	PIL	0.021	0.026	23.80952
2	WIP	0.122	0.171	40.16393
3	SOF	0.113	0.114	0.884956
4	ABL	0.119	0.118	-0.84034
5	ABS	0.011	0.121	1000
6	BCS	0.071	0.080	12.67606
7	BF	0.322	0.322	0
8	ATC	0.044	0.080	81.81818

5.4.1.3 Deviational variables

Based on the simulation model, the status of the following variables can be monitored at discrete steps of time: SLT, DD, TT, PLT, and PD. The following deviational variables are conceivable:

- Q1. Actual demand observed in the k^{th} week (corresponding to the request column in data set for the first 13 weeks).
- Q2. Anticipated demand in the k^{th} week (corresponding to the plan column in data set for the first 13 weeks).
- Q3. Finished parts buildup/backlog at the assigned/scheduled dispatch time assigned for the products in the k^{th} week (corresponding to the "RqdShipDt" column in the dataset)
- Q4. Finished parts buildup/backlog at the assigned/scheduled customer request date assigned for the products in the k^{th} week (corresponding to the "CustRqstDt" column in the dataset).

Further, the following events can be simulated and studied:

- E1. Mismatch between plan and request for specific models due to changes in demand patterns. Therefore E1 is indicated by the combination of Q1 and Q2, i.e., Q1-Q2.
- E2. Mismatch between the planned dispatch time to customer and the order completion time. This may be due to manufacturing problems, problems because of suppliers, etc. Therefore E2 is indicated by the combination of Q3 and Q4, i.e., Q3-Q4.
- E3. Mismatch between expected delivery time and the shipment arrival time. Therefore E3 is indicated by Q4 and the transportation time.

Event E occurs if if $\delta_{F1} \geq |Q_1 - Q_2|$, or $\delta_{F2} \geq |Q_3 - Q_4|$, or $\delta_{F3} \geq |Q_3 - T_1|$.

5.4.1.4 Economic design of monitoring chart

The economic model is solved for determining the sampling frequency and the control limit for the three deviational variables. The parameters for the control chart are based on parameters taken up by Panagos (1985) in his study:

Table 5-4 Data used for the estimating the parameters for the control chart (from Panagos, 1985)

Chart	M	∂	λ	E	b	C	W	T	V_0	S	S_1	D_1
1	50	1	0.01	0.05	0.5	0.1	35	50	50	10	0.05	4
2	100	1	0.01	0.05	5.0	1.0	250	500	150	100	0.05	4
3	50	2	0.01	0.05	5.0	1.0	250	50	50	10	1.00	40

Further, the functional form of the net hourly income is considered to be a monotonically decreasing form and is of the following form:

$$V(t) = V_1 - \theta_1 t \qquad \text{wherein } \theta_1, t \geq 0 \tag{1}$$

Similarly, Z can be expressed as

$$Z(t) = Z_1 + \theta_2 t \qquad \text{wherein } \theta_2, t \geq 0 \tag{2}$$

Substituting the equations in the previous cost function and simplifying,

$$\begin{aligned} E(NI) &= (V_0 / \lambda) + [V_1 - \theta_2 + (2\theta_1 / \lambda)](T_3 - \tau) - \frac{\alpha Y}{exp(\lambda h) - 1} - bE(T') / h \\ &- \theta_1 \{T_3^2 + (2T_3 - h)[(1/\lambda) - \tau]\} \end{aligned} \tag{3}$$

where

$$T_3 = T_2 - jh \qquad \text{wherein } \theta_2, t \geq 0 \tag{4}$$

5.4.1.5 Solutions and results

The mapping function that has been developed is as follows:

Table 5-5 SVM based pattern identification for historical I Series performance data

Serial No.	Metrics	Std. Error	SVM Weights				
			SLT	DD	TT	PLT	PD
1	PIL	0.021	0.00	0.32	0.36	0.10	0.22
2	WIP	0.122	0.23	0.19	0.16	0.12	0.30
3	SOF	0.113	0.14	0.33	0.28	0.03	0.12
4	ABL	0.119	0.21	0.09	0.12	0.32	0.26
5	ABS	0.011	0.22	0.08	0.15	0.33	0.22
6	BCS	0.071	0.26	0.22	0.0	0.21	0.21
7	BF	0.322	0.26	0.22	0.0	0.21	0.21
8	ATC	0.044	0.22	0.12	0.0	0.28	0.38

Taking $\theta_1 = 0.6$ and $\theta_2 = 0.9$. The solution for the economic model is as follows:

Table 5-6 Economic design of the control chart

Chart	h	K	E(NI)/E(T)
1	3.12	3.14	42.11
2	6.92	2.94	143.78
3	7.01	3.46	38.06

5.4.1.6 Discussion and extensions

In this chapter, the economic design of the Shewhart chart used for monitoring the logistic variables was discussed. The assumption is that the mean shift is detectable by the Shewhart chart. One effective alternative to the Shewhart chart when small shifts are of interest is the cumulative sum (CUSUM) control charts proposed by Page (1954), It has been shown that CUSUM schemes perform better than Shewhart charts in detecting small process shifts.

Although the CUSUM chart has good protection against small process shifts, it is not effective in detecting large process shifts. Several enhancements have been proposed to improve the performance of the CUSUM schemes. Lucas (1982) proposed a parabolic V mask to improve the responsiveness of CUSUM to large shifts. In a subsequent paper, Lucas showed that a combined Shewhart-CUSUM scheme gives improved properties when both large and small shifts are to be detected. Crosier proposed a modified CUSUM scheme and showed its performance slightly better than the conventional CUSUM scheme. Another control scheme useful for detecting small shifts is the Exponentially Weighted Moving Average (EWMA) control scheme proposed by Roberts. Lucas and Saccucci evaluated the EWMA scheme and showed that EWMA and CUSUM control schemes have similar performance properties. Combining the two charts with weights given appropriately can have the properties of both the charts, and we can move from one chart to the other easily.

The trend displayed by deteriorating systems can help estimate the occurrence of an event before it actually occurs, and so the costs can be considerably reduced due to the reduction of the down times. However to exploit this potential saving, there is a need to estimate the MTTF along with the parameters discussed before. The economics of the process is highly dependent on the reliability of the estimates and the predictability of the MTTF on the basis of the same. The following three issues are critical for the success of the predictive approach:

1. The events need to be recurrent. If in case the event is non-recurrent, then predictive approaches loose a bulk of their utility unless the effect is similar to other events.
2. There should be detectable patterns in the data for us to be able to predict the occurrence of the impending events. This is a necessary condition. If in case we don’t have any detectable pattern, then the predictability is lost and there is little that we can do. The logistic cost is infinity and we have to use either the corrective or the preventive approach to EM.
3. There should be sufficient time gap between the detection and time to respond. If in case we don’t have this scenario, the returns will be low, as we will be operating very close to either the preventive approach or the corrective approach.

5.5 Conclusions

The traditional approach to production planning applies functional decomposition to planning tasks: first, process engineers locally optimize all process parameters, including the selected process plan and cutting speeds, and, only then, production engineers optimize the production quantities and schedules in order to obtain the maximum profit when all process variables are considered as fixed parameters. The EM-based integrative models try to identify variables that may have significant impact on the company profitability and decompose the planning problem accordingly.

The existing HPP approach creates a master and slave hierarchy where the solutions at one level form the starting point for decisions to be made at the next level. In EM approach, only “hard” parameters, such as tools and material characteristics are fixed, while most of the other parameters such as cutting speeds, feed rates, and depth of cut are considered for being soft parameters, modeled as global variables in the aggregate model. Handbook-based parameters are usually considered being “optimal”; however, they are usually based on local operational rules without a global view of long-term profitability.

Chapter 6

Insights and perspectives on why the AHFM approach is doing well

6.1 Introduction

In the previous chapters, the AHFM approach has been adapted and applied in various domains, including discrete manufacturing, process industries and airline fleet scheduling problems. Early investigations into these problem domains have yielded remarkable improvements in performance (up to 59% improvement in a discrete manufacturing environment, 84% improvement in process industries and up to 25% improvement in fleet scheduling). For the case study taken up for the IBM supply chain (presented for illustration in this chapter), the indicated improvements are of 27%. While these improvement percentages cannot be generalized for evaluating gains from implementing AHFM, the examples illustrate that at least in some cases the magnitude of potential profits cannot be ignored. In addition, the conversion of parameters to variables adds flexibility to the models. In fact preliminary results of AHFM to the airline fleet scheduling problem indicate that the technique yields a feasible solution to problems that were considered infeasible.

The benefits are substantial enough to incite the need for further evaluation and explanations of the reason for the improvement. The question that needs to be answered is what is the underlying reason behind the improvement? The focus of this chapter is precisely on addressing these issues. The rest of the chapter is organized as follows: Section 2 is on determining the reason for the high improvements that are observed from the alternative problem partitioning approach that AHFM essentially provides. Section 3 takes up the IBM case study to illustrate the working of the AHFM approach and the Sense-Response enabled EM system. The conclusions and future work are presented in Section 4.

6.2 Examination of the AHFM approach: How does it work?

Conceptually, the AHFM approach aggregates the flexibility/redundancies that are available at the various planning levels and then optimally assigning the resource/time from the aggregate pool to deal with various disturbance and problems. Since the model spans over multiple levels of the activity circles, it does not compromise of the size of the solution space and hence the feasibility issue can be addressed in a straightforward manner.

To address the issue related to the solution quality, it is imperative that we look at what AHFM approach does from a dynamical perspective. However, before doing that, the time persistent perspective on EM is briefly revisited.

From the time persistent perspective, “event” is the cause or the consequence of a mean shift in the state variables. Further, the condition of y^P equaling x^p under the influence of a mean shift translates to studying the response of the dynamical system to a step change. The supply chains, production systems, and other material flow processes have internal dynamics, and the transition of states (when we are deriving the new optimal solution by perturbing or re-optimizing) is equivalent to the response of these dynamical systems to the step changes.

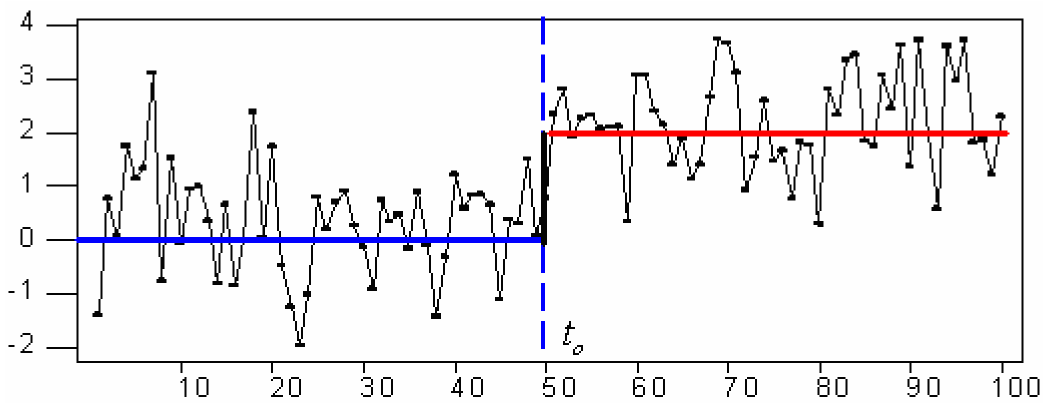


Figure 6-1Time persistent perspective of Δ change

The transience represents the cost that is incurred because of the time delays involved in the system to respond. The objective of EM is to decrease the area under the curve that represents the difference between the actual trajectory and the desired trajectory by as much as possible. This can be achieved by responding to arbitrary changes in the desired nominal condition with well behaved transients in state and control variables.

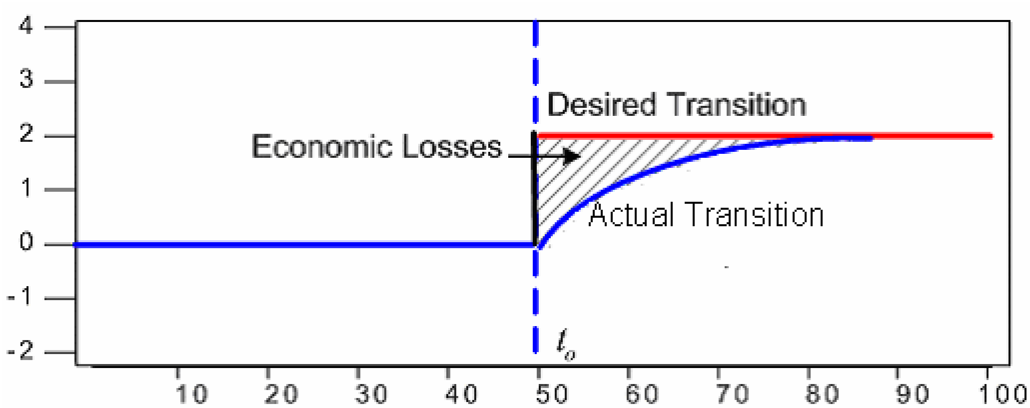


Figure 6-2 Response of systems to step change in state variables: Effect of system delay

The response of the dynamical system to a step change in the input however varies with the properties of the system. The supply chains typically are high order systems and so that economic loss that is shown in Figure 6-3 is high.

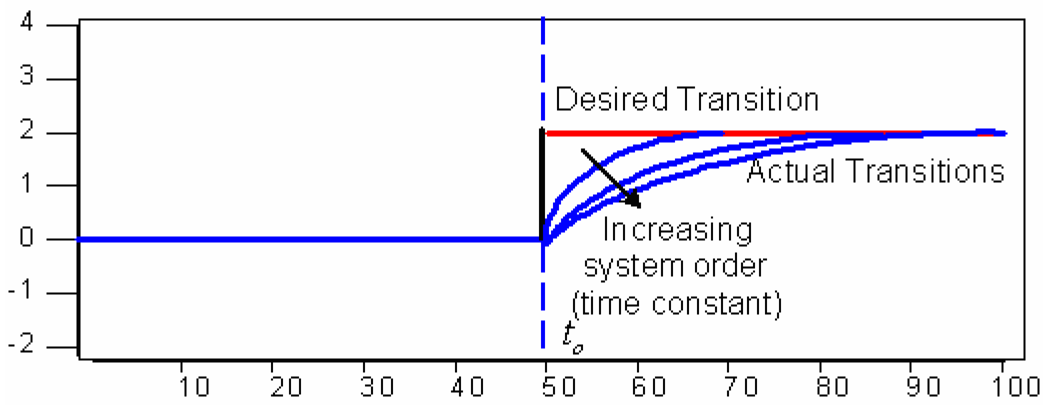


Figure 6-3 Response of systems with different dynamics to step change in state variables

However, the components of the activity levels shown in Figure 1-1 show multiple time scale characteristics. When arranged in order of decreasing magnitude of the average characteristic time between consecutive decision/action implementations, they can be grouped into multiple categories (Caramanis, 2002):

1. Investment and strategic decisions (alliances, service rate targets) [time scale, years]
2. Allocation of resources (machines, tools, manpower by skills, floor space) and production responsibilities (part types and associated rout segment operations) among cells; cell layout; production process; maintenance policy, and the like [time scale, months]
3. Plant-wide production and delivery promise schedule with cell-specific weekly production targets, cell-specific manpower and shift/overtime planning, contingency planning (safety stock, kanban size, and other contingency planning policies) [time scale, week]
4. Cell-specific day to day and hour to hour work center production dispatch, detailed operations scheduling, response to contingencies such as absenteeism, quality control, yield, delivery requirements/demand variations, failures and repairs, and short term horizontal coordination between interacting cells (e.g. sequencing of runs in the presence of significant setup delays) [time scale, hour to shift]

The system can therefore be decomposed into multiple sub-systems, each linked to each other. The overall transfer function remains the same.

The problem partitioning within the AHFM approach allows the breaking up the system in Figure 1-1 into multiple sub-systems, each linked to the next one because of the common high fidelity variable in terms of which the AHFM has been formulated. The sense capabilities further enhance the system by allowing intermediate measurement points.

The AHFM approach builds on the fact that the transfer functions have been decomposed serially and that intermediate measurement points are available, and develops a cascading control system. The solution to the AHFM problem can be seen as the gains for the controllers.

Cascade control can improve the response to a set point change by using an intermediate measurement point and multiple feedback controllers. This improvement can be shown by examination of the transfer functions between the system output and the disturbance for the cascade and conventional single feedback loop control configurations, respectively. For a cascade control, from Figure 6-4, the disturbance transfer function with d_1=0 is given

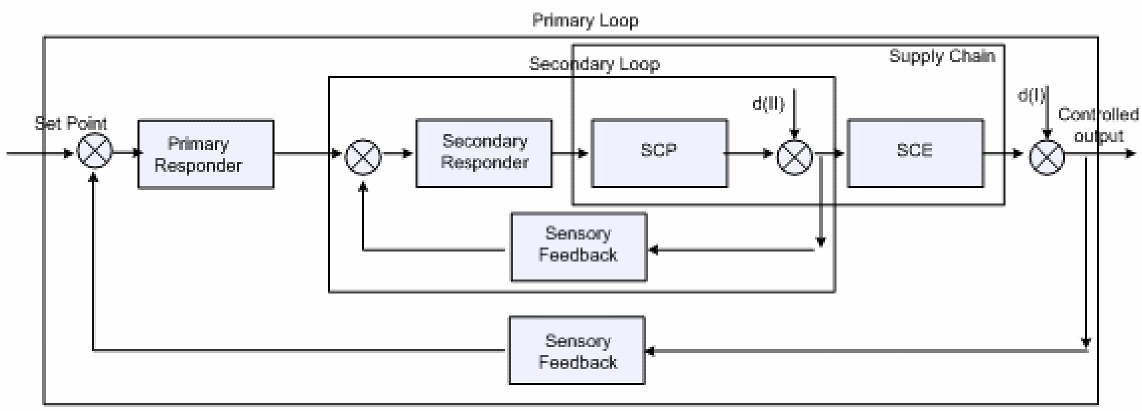


Figure 6-4 Cascade control architecture for SCEM

The improvement can be shown by examination of the transfer functions between the system output and the disturbance for the cascade and conventional single feedback loop control configurations, respectively. For a cascade control, the disturbance transfer function with d_1=0 is given by

$$\frac{Y_1}{D_2} = \frac{G_{p1}}{1 + G_{c2}G_{p2} + G_{c1}G_{c2}G_{p1}G_{p2}} \tag{6-1}$$

Without the inner feedback loop, and of course no controller G_{c2}, the transfer function between the same variables is

$$\frac{Y_1}{D_2} = \frac{G_{p1}}{1 + G_{c1}G_{p1}G_{p2}} \tag{6-2}$$

Because of the extra degrees of freedom, when appropriate values of the parameters of the two controllers are chosen, the cascade control will generally result in a better response.

However this is not all. The parameter variation approach works in conjunction to the AHFM approach and looks for solutions in the neighborhood of the previous solution. The changes in the plan imply that the state of the system is also affected by this change. The approach therefore modifies the dynamics of the execution system as well. It becomes time variant and by changing in conjunction with the change in the planning system, the losses are further reduced.

Conceptually, the idea can be explained as follows:

1. The AHFM approach decreases the step size that the execution system sees.
2. The parameter variation approach changes the dynamics of the execution system.

Together, the reduction in losses due to the state transition is more than when either of the two approaches are used independent of each other.

Further, the cascade control structure has the following advantages over a single feedback loop control system:

1. The secondary controller is used to correct disturbances arising within the inner loop before they can affect the controlled variable.
2. The effect of parameter variations in the process G_{p2} are corrected in the inner loop by the secondary controller.
3. The effect of any phase lag existing in G_{p2} may be reduced by the secondary loop, thus allowing the speed of response of the primary loop to be improved.

6.3 Implementation: IBM Case study

The aggregate high fidelity model for the system can be developed/represented as follows: Objective is to maximize the profits over multiple time horizons. These profits can be weighted or simple

$$Max\ Z_1 : P_1 + P_2 + P_3 + P_4 + \ \ldots.. + P_n \tag{6-3}$$

Subject to:

$$P_i = [Pr(t_i) - Cp(t_i) - A(t_i)]\, f(t_i) \tag{6-4}$$

Where P_i is the profit in the i^{th} planning horizon and Cp refers to the production costs.

$$Pr(t_i) = \frac{a_p}{v_i} + b_p v'_i + c_p \tag{6-5}$$

The production costs can be expressed as a function of the capacity as follows

$$Cp(t_i) = \frac{a}{v_i} + bv'_i + c \tag{6-6}$$

$$f(t_i) = kv_i \tag{6-7}$$

Further,

$$v_i^{low} < v_i^{actual} < v_i^{high} \tag{6-8}$$

Where v_i^{low} and v_i^{high} are functions of v'_i.

Note that everything has been expressed in terms of the capacity. Capacity can be manipulated by changing the process variable (operating conditions), by allowing overtime, and by inter facility transshipment. The dynamics of the change in the process variable is much faster than the dynamics of the inter facility transshipment. The economics of the choice will be provided by the AHFM approach.

The execution system can be considered as consisting of two parts, process 1 and 2. Process 1 is the primary and has the output that we want to control, that is the productivity of a manufacturing facility. Process 2 is the secondary has an output that is a that essentially involves changing the ORS and the operating parameters that aid ramp up of productivity that we are not really interested in controlling, but which effects the output we want to control. Disturbance arising within the secondary loop are corrected by the secondary controller before they can affect the values of the primary controller.

Note that the closed loop response of the primary loop is influenced by the dynamics of the secondary loop, whose open loop transfer function is equal to:

$$G_{secondary} = G_{c2}G_{p2} \qquad (6\text{-}9)$$

The stability of the secondary loop is determined by the roots of its characteristic equation

$$1 + G_{c2}G_{p2} = 0 \qquad (6\text{-}10)$$

Consider a process with the following transfer functions for its primary and secondary elements.

$$G_{p1} = \frac{1}{(1+0.5s)(1+s)} \qquad (6\text{-}11)$$

$$G_{p2} = \frac{1}{(1+0.1s)} \qquad (6\text{-}12)$$

The secondary process is faster than the primary, as can be seen from the time constants. When we use the simple feedback control, the open loop transfer function can be modeled as

$$G_{c1}G_{p1}G_{p2} = K_c\left(1+\frac{1}{s}\right)\frac{1}{(1+0.1s)}\frac{1}{(1+0.5s)}\frac{1}{(1+s)} \qquad (6\text{-}13)$$

The cross over frequency can be found from the equation that sets the total phase lag equal to -180^0;

$$\tan^{-1}\left(\frac{-1}{\omega_{co}}\right) + \tan^{-1}(-0.1\omega_{co}) + \tan^{-1}(-0.5\omega_{co}) + + \tan^{-1}(-\omega_{co}) = -180^o \qquad (6\text{-}14)$$

and is equal to

$$\omega_{co} = 4.45 rad/\min \qquad (6\text{-}15)$$

Also the overall amplitude ratio is given by

$$AR = K_c\sqrt{1+\frac{1}{\omega^2}}\frac{1}{\sqrt{1+(0.1\omega)^2}}\frac{1}{\sqrt{1+(0.5\omega)^2}}\frac{1}{\sqrt{1+\omega^2}} \qquad (6\text{-}16)$$

for $\omega = \omega_{co} = 4.45$, and $AR = 1$ we get $K_c = 11.88$. Therefore, when the disturbance d_2 changes, the simple feedback controller can use a gain upto 11.88 before the system becomes unstable. Also, given the fact that the overall system is third order, the closed loop response to changes in d_2 will b sluggish.

The open-loop transfer function for the secondary loop when we consider a simple proportional controller is given by:

$$G_{c2}G_{p2} = K_{c2}\frac{1}{(1+0.1s)} \qquad (6\text{-}17)$$

There is no crossover frequency for the secondary control loop, and therefore we can use large values for the gain which produce very fast closed loop response to compensate for the changes in the disturbance arising within the secondary process.

Further, if we look at the area under the curve for the IAE behavior for the two systems, A1 is 108 whereas A2 is 78 units. This indicates an improvement of about 27% in the IAE.

6.4 Technology road-mapping: the case of Thermal Products Solutions

In the semiconductor manufacturing industry, thermal processes such as reflow, curing, and baking are integral for several binding, drying, and packaging operations. The nature of these process heating operations significantly influences the cost and the quality of the finished product. Till recently, the emphasis was more on the quality aspects of the solutions and not too much on the efficiency and the efficacy of the solutions for these thermal processes. However, the need for efficient solutions has become the need of the day primarily because (1) inefficiency here leads to errors downstream, (2) the thermal operations are becoming the bottlenecks in the chip manufacturing process, and (3) the potential to decrease the operating costs and increasing the throughputs is obvious in these processes.

We see these three aspects as the primary drivers for technology innovation for the next generation thermal solutions for the semiconductor industry. We are focusing on

1. Improving quality (in terms of uniformity of heating, efficiency in following the heating and cooling curves while making the design modular so that the same solution fits all),
2. Improving throughput (while decreasing the footprint of the systems), and
3. Decreasing costs (by improving the efficiency of the thermal systems and increasing the reliability of the machines).

It is envisioned to achieve the three objectives by synergizing the breakthroughs in three pertinent areas:

- Process Heating: Where TPS combines the latest developments in the radiant heating domain with the conventional convective heating.
- Process Control: Where the cascading control concepts have been introduced to increase the fidelity of the temperature control, and decrease the variance between the actual and the required temperature profiles to less than 0.5C.
- System Design: Where the benefits of both batch and continuous processing are pooled in the latest line of in-line vertical heaters.

The Generation II and II* systems and design will be compatible to various safety standards (SEMI S2, and CE), and material handling standards (SMEMA), and can be used for multiple input chip sizes (200-300mm). The design is modular and the same Generation II system Legos can be configured for handling various operations-including the combination of reflow and curing in the same unit.

6.4.1 Technology Update: Process Heating

Conventional heating technology
Current technology is primarily convective: the technology constraints the capabilities that can be achieved. In conventional or surface heating, the process time is limited by the rate of heat flow into the body of the material from the surface as determined by its specific heat, thermal conductivity, density and viscosity. Surface heating is not only slow, but also non-uniform with the surfaces, edges and corners being much hotter than the inside of the material. Consequently, the quality of conventionally heated materials is variable and frequently inferior to the desired result.

Imperfect heating causes product rejections, wasted energy and extended process times that require large production areas devoted to ovens. Large ovens are slow to respond to needed temperature changes, take a long time to warm up and have high heat capacities and radiant losses. Their sluggish performance makes them slow to respond to changes in production requirements making their control difficult, subjective and expensive.

Forced convection is however useful as it carries away the flux, and provides the uniformity that is much desired in the products.

Variable-frequency microwave heating
The utilization of microwave energy for processing advanced materials has now been made practical with the advent of variable-frequency microwaves (VFM). This technology offers an attractive alternative to existing thermal treatments based on convection and IR heating. Fixed-frequency, multimode microwave ovens, such as those utilized in home kitchen applications, cannot be used for processing advanced materials for two main reasons:
1. The lack of field uniformity, and
2. The absence of process control and reliability.

In VFM, the electric-field distribution, on a time-averaged basis, is uniform throughout the entire cavity volume, leading to uniform exposure of the processed material(s) to the microwave energy. Furthermore, the incident micro-wave frequencies can be changed to optimize the microwave energy absorption by the material(s) of interest.

6.4.2 Proposed mixed mode heating technology

TPS will primarily focus on radiant heat transfer and hybrid techniques that combine convective heat transfer with the radiant counterpart.
Compound Frequency Modulation: The epoxy and the solder are compounds. They are not excited by a single frequency-but a combination for various components. The

compound frequency modulation is built on the fact that these frequencies need to be identified and the shift between these frequencies at predefined rates so as to achieve fast and uniform heating. The concept is built upon the existing and proven VFM technology, but is customized to the particular needs of the semiconductor industry.
Forced Convection: Build upon concepts that TPS has already achieved market supremacy in. The system will use the technology that is tested in its Tiros products.

Together, the two thermal processing technologies will provide some highly desirable features such as:

- Uniformity of temperature profile inside the oven as well as in the heated material
- High-fidelity and quick response to required changes in temperature change
- Focused heating that reduces the curing and pre heat time by a factor of almost 10
- Energy efficiency that goes along with the focused heating
- Flux removal and disposal opportunity due to the forced convection
- Inert atmosphere and low pressure operation capability.

6.4.3 Thermal Solutions Roadmap

TPS plans on systematically integrating the developments in technology and process control in the Generation II and II* thermal solutions. The up gradation is envisioned as follows in the feature lines for curing and reflow:

	Conventional	Generation II	Generation II*
Process Heating	Convective	Convective supported by Radiant	Radiant supported by Convective
Process Control	Feedback Control	Cascade Control	Advanced Adaptive control
System Design	Vertical in-line	Vertical in-line Oven Legos	Vertical in-line Oven Legos

The milestones that we plan are provided in the timeline presented in Figure 1

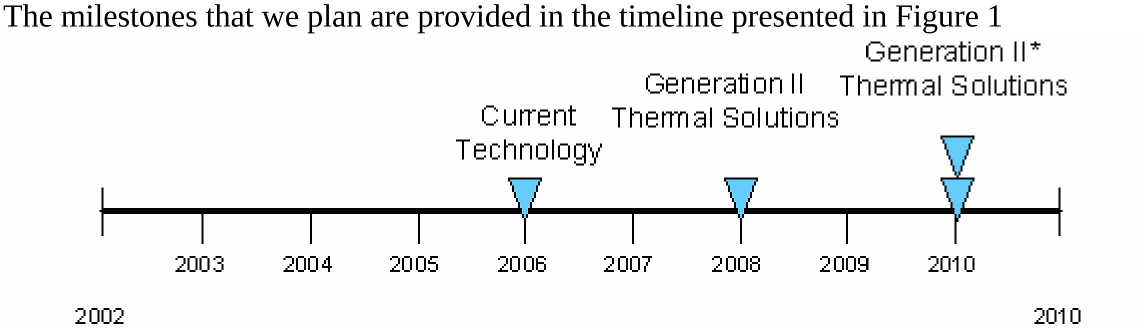


Figure 6-5 Milestones and timeline

The standard features that these will be equipped with are:

1. Modular design: The same oven Legos will be used for building up the solution, they can be assembled and reassembled as per requirements.

2. Integratability: Compliant with the SMEMA standards and provides interface for both local control and integration for advanced process control for fab wide control.
3. Compliance with the SEMI S2, and CE standards. With special emphasis on flux containment and the provision of cooling zones
4. Compliant with the JEDEC standards
5. Class 10,000 and class 1,000 clean room compatibility

Projected energy efficiency

We define efficiency in terms of input power consumed and the heat energy absorbed by the inputs to the oven.

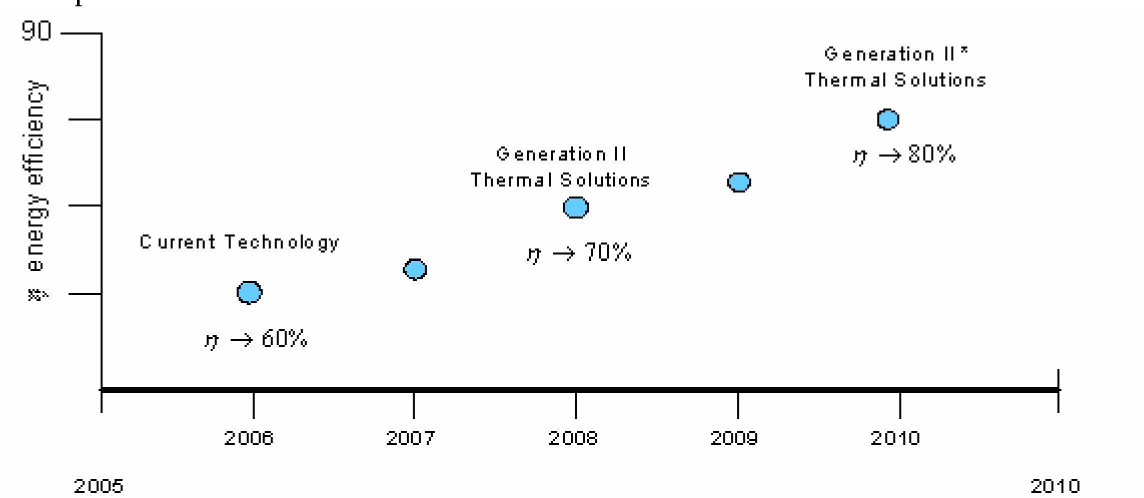


Figure 6-6 Projected Energy Efficiency

Projected throughput

The throughput is increased because of the reduction in the residence time. This is made possible due to the radiant heating technique that is being used as opposed to the convective heating.

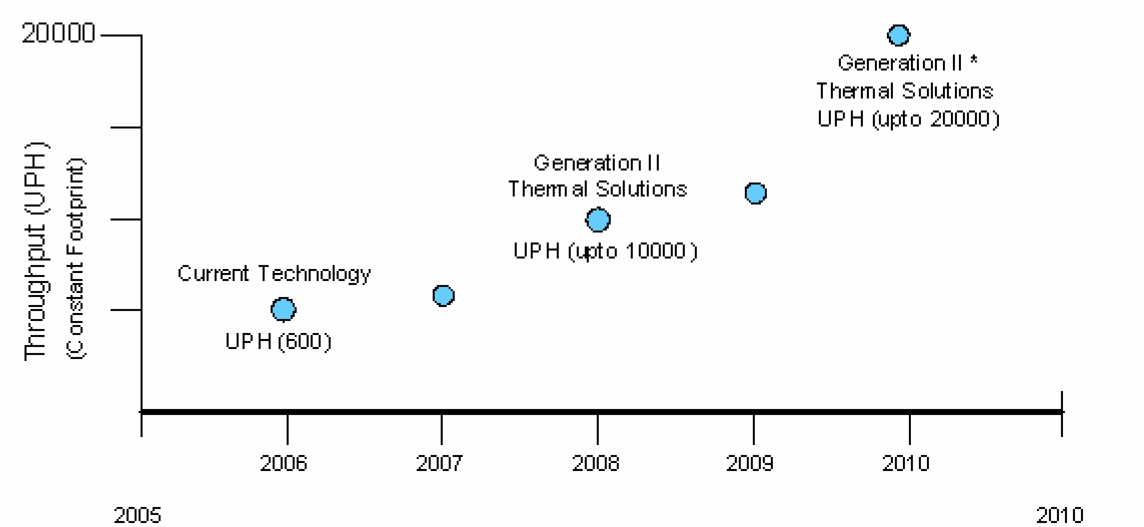


Figure 6-7 Projected Throughput

Projected temperature profile compliance

Higher fidelity in the control system implies that there is a better compliance. It may be noted that due to the focused heating-the compliance between the temperature inside is the higher. For instance for reflow, we can do it at a lower temperature of about 250 C as compared to 270 C otherwise. The advancement in the control system is targeted to get the temperature compliance better in the Generation II* systems.

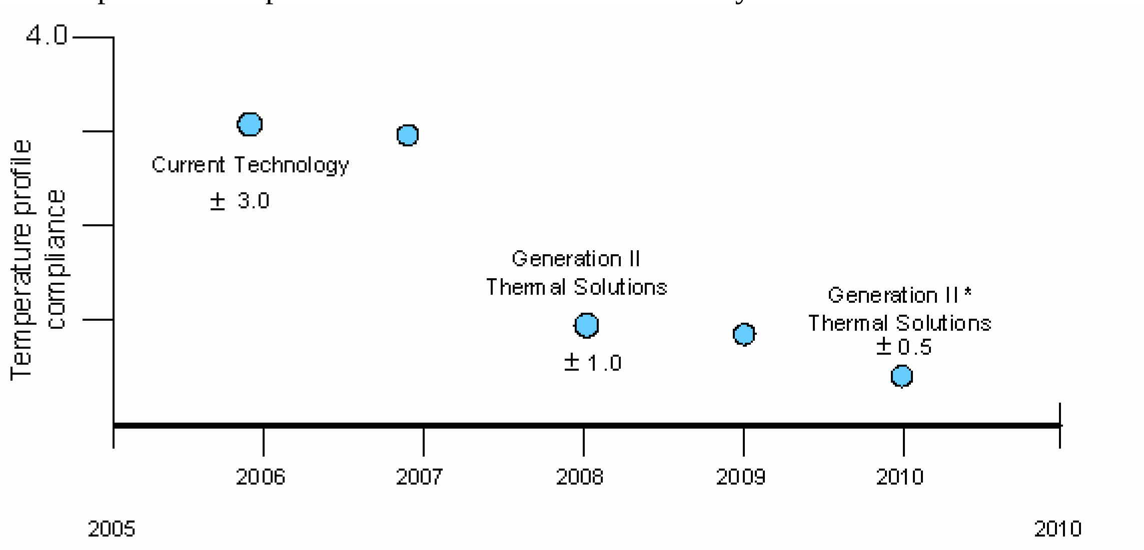


Figure 6-8 Projected deviations in temperature profile

Projected affordability

Affordability is calculated on the basis of the projected initial cost, the expected throughput, and energy efficiency, and the cost of defects due to the thermal processing.

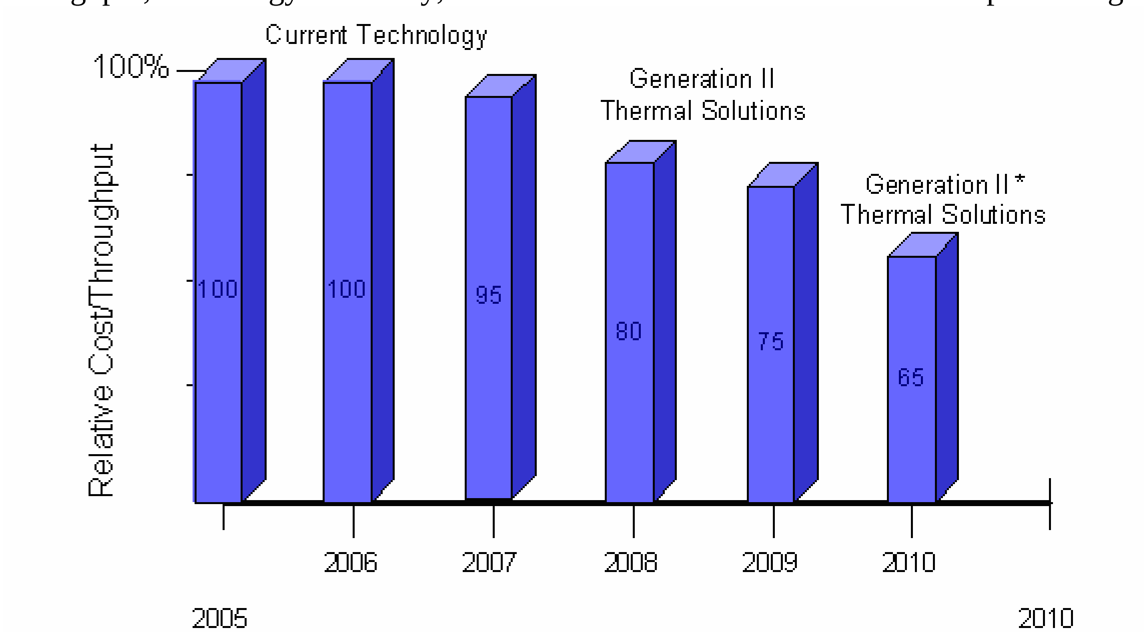


Figure 6-9 Projected affordability

6.4.4 Conclusions

The Generation II ovens are built using the force convection and compound frequency modulation techniques having advanced process control features for recipes for individual requirements.

Even though the expectations from the thermal solutions (in terms of throughput, and efficiency in terms of cost and quality) are increasing, we are providing a solution roadmap wherein the costs are decreasing to about 65% of the current costs, the quality of the thermal solutions is improving, and the projected needs of the semiconductor industry (in terms of throughput, chip size, and quality) are satisfactorily met.

6.5 Marketing Engineering: The Case of Kimberly Clark

The Marketing engineering project provides a set of analytical tools for identifying and exploiting opportunities on the market as well as the production front simultaneously that could enable the maximization of the operating profits for the business system under consideration (in this case, the Fem Care business of Kimberly Clark in Colombia). On the production side, the tools help identify the opportunities (and saving) that could be created by efficient batching, regularized changeovers, and exploiting the economies of scale. On the market side, the focus is on identifying the response of the market to marketing mix (advertisement, promotions, pricing, numeric distribution)-we consider the case of multiple products in a competitive market. The inferences drawn from the two components are combined and integrated into a mathematical model that provides suggestions for tactical production and marketing plans. These tactical plans also form the basis for generating the operational plans as well.

The optimization model that forms the core of this decision support system is an operational profits (OP) maximization model that we will discuss next. The input to this model (Equations 1-13) are the inferences that are drawn from the production and market side (parameter estimates for the structural equations (Equations 4-12) included in the model), the P&L files from the SAP B/W, and any other goal or requirements. The output of this model are the product mix and lot sizes for production, the desired marketing mix split (promotional and advertisement (A&P) spending, pricing and numeric distribution), and other information which forms the input to the operational planning models. The operational planning models provide detailed production plans and marketing strategies for implementing and achieving the projections form the mathematical model.

The Fem care products industries in Colombia see a mature market, characterized by high concentration of product providers, product features and options, large advertisement-to-sale ratios, and aggressive introduction of new products. We are trying to estimate the elasticities (sensitivities of market share to impacts of price, advertisement, trade and consumer promotion, and numerical distribution) and the market power that different sub-brands-not just KC, but competitors such as Sancela and J&J have in the Fem care segment in the Colombian market. We are also trying to break down these estimates of the elasticities by geographical regions and the channel of sale for KC. The key idea is that the visibility in the elasticities and the sensitivities of market share of a sub-brand to the marketing mix variables can enable

1. Applying targeted marketing mix strategies to the targeted market through the targeted channel
2. Developing goals for achieving specific market shares at specific points of time constrained by the budget availability for applying the marketing mix and the production capabilities.

The proposed exercise relies on the ability to consistently estimate the demand and the market share, and establish the relationship between the marketing mix variables and the market share, while explicitly considering the competitors strategies. The complexity arises because:

- We have limited visibility into the competitors strategy.
- We have limited amount of data
- Explicit experiments are not possible or are too expensive as we don't control the competitors behavior.
- We don't know explicitly the impact of the heterogeiety in the consumer behavior
- We don't know the response that the same marketing mix will have under different market conditions.

I use a three dimensional panel of quantities and process for 25 sub-brands in 6 zones spread over 3 channels of sale over a period of at least 1 year (data made available bi-monthly). The complexities in these calculations arise while we are making these estimates, we are explicitly considering that the competitor responds to the strategies.

6.5.1 Tactical Planning Model (Common Calendar)

The objective is to maximize the operating profits $(OP)_i$ over a fixed time interval T . It is also desired to keep this growth sustainable. This translates to:

$$\text{Max } \frac{1}{T}\sum_{i}^{T}(OP)_i \tag{5}$$

$$\text{s.t } (OP)_{i+1} \geq (OP)_i \tag{6}$$

Where $(OP)_i$ is the operating profit estimated for month i .

However, $(OP)_i$ is defined within the P&L files as:

$$(OP)_i = (NS)_i - (COGS)_i - (DC)_i - (SMG\,\&\,A)_i - (A\,\&\,P)_i \tag{7}$$

Where $(NS)_i$ is the Net Sales for month i , $(COGS)_i$ is the Cost of goods sold in the month i , DC is the distribution cost, SMG&A is the spending on sales, marketing, and G&A in the month i , and finally A&P is the spending on advertisements and promotions in that month. These terms can be individually modeled as:

$$(NS)_i = \sum_{j,k,l}(SP)_{ijkl} X_{ijkl} + \sum_{j,k,l}(M)_{ij} Y_{ij} \quad (8)$$

Where SP(i,j,k,l) is the selling price of the jth product in zone k through channel l in time period. X(i,j,k,l) is the quantity sold. Similarly, Y(i,j) is the quantity exported and M(ij) is the markup on that product type j.

$$(COGS)_i = \sum_j \sum_m \frac{\alpha^c_{jm}}{\sum_{kl}(X_{ijklm}) + Y_{ijm}} + \sum_j \sum_m \beta^c_{jm} \left(\sum_{kl}(X_{ijklm}) + Y_{ijm} \right)^{\theta(jm)} + \sum_j \sum_m \chi^c_{jm} \quad (9)$$

COGS is assigned a structural form as presented in Equation 5, and the coefficients are determined apriori from the production data from SAP. Note that X(i,j,k,l,m) is the quantity of the product j made on machine m in the ith time period and sold in the local markets whereas Y(ijm) is the quantity of the product j made on machine m but is exported.

$$(DC)_i = \alpha^{dl} \sum_{jklm}(X_{ijklm}) + \alpha^{de} \sum_{jm}(Y_{ijm}) + \beta^d e^{C*(\Delta ND)_i} \quad (10)$$

The distribution cost has three components: the first two being the cost for the local and the export markets that grows linearly with the distribution volume, and the third being the exponential increase due to change in the numeric distribution.

$$(SMG \& A)_i = \chi^{SM}_i + \alpha^{SM}_i (\sum_{jklm} X_{ijklm}) \quad (11)$$

SMG&A is seen to have a constant component and a volume dependent component.

$$(A \& P)_i = A_i + P_i \quad (12)$$

Advertisement and promotions are seen as incentives, and are not given any functional form here (the form and structure are derived from the market side model)

The market side model considers $(SP)_{ijkl}$, and $(ND)_{ijkl}$ as incentives along with A_i P_i. that are offered to the consumers (and comprise of the marketing mix that we need to determine), Currently, we have developed a non-parametric model to describe these relationships and therefore the structural form of the market side model is:

$$(I)_{ijkl} = f\left(A_i, P_i, (SP)_{ijkl}, (ND)_{ijkl}\right) \quad (13)$$

And, the incentives are included into the parametric model as:

$$\sum_m X_{ijklm} = \alpha^{IN}_{ijkl} + \beta^{IN}_{ijkl}\left(1 - e^{-C*(I)_{ijkl}}\right) \quad (14)$$

The production side model provides a rough capacity constraint for the given product mix

$$\sum_j \frac{\alpha^T_j}{\sum_{kl}(X_{ijklm}) + Y_{ijm}} + \sum_j \beta^T_j \left(\sum_{kl}(X_{ijklm}) + Y_{ijm} \right)^{\theta(j)} + \sum_j \chi^T_j \le K_M \quad (15)$$

$$\sum_m X_{ijklm} = X_{ijkl} \quad (16)$$

$$\sum_{m} Y_{ijm} = Y_{ij} \tag{17}$$

The model (Common Calendar) therefore involves the solution of Equations 2-8, and 10-13 simultaneously, with Equation 1 as the objective function.

6.5.2 Inferences from the market and the production: parameter estimation

For the solution of the model, several parameters and function forms that go into the analytical model have to be estimated. Some of these are:

1. On the production side, the parameter values for Equations 5 and 11
2. From the P&L, the parameter values in Equations 6 and 7 need to be determined
3. And from the market side, the function form of Equation 9 and the parameters in Equation 10 have to be determined.

Estimation of parameters from the production data:
We have the following data available:

1. The setup time and the changeover time.
2. The daily production time and cost for a given SKU on a given machine.

We have to infer the parameters that go into Equations 5 and 11 from this data.

It may be noted that the structural form of the two equations (5 and 11) require a non-linear least square estimation. However this becomes difficult as the model scales, so, break up the equation as follows:

$$T_{ijm} = \frac{\alpha_{jm}^{T}}{\sum_{kl}(X_{ijklm}) + Y_{ijm}} + \beta_{jm}^{T1}\left(\sum_{kl}(X_{ijklm}) + Y_{ijm}\right) + \beta_{jm}^{T2}\left(\sum_{kl}(X_{ijklm}) + Y_{ijm}\right)^{3} + \chi_{jm}^{T} \tag{18}$$

This equation is in the standard from

$$Y = \beta X + \varepsilon \tag{19}$$

Where X comprise of 4 columns presented in Equation 16

$$X = \left[1, \frac{\alpha_{jm}^{T}}{\sum_{kl}(X_{ijklm}) + Y_{ijm}}, \left(\sum_{kl}(X_{ijklm}) + Y_{ijm}\right), \left(\sum_{kl}(X_{ijklm}) + Y_{ijm}\right)^{3}\right] \tag{20}$$

Therefore:

$$\hat{\beta} = \left(X^{T} X\right)^{-1} X^{T} Y \tag{21}$$

Where

$$Y = \left[T_{ijm}\right] \tag{22}$$

The estimation of the cost parameters is exactly the same-the Y vector however changes and reflect the cost rather than the time numbers-i.e.,

$$C_{ijm} = \frac{\alpha_{jm}^{C}}{\sum_{kl}(X_{ijklm}) + Y_{ijm}} + \beta_{jm}^{C1}\left(\sum_{kl}(X_{ijklm}) + Y_{ijm}\right) + \beta_{jm}^{C2}\left(\sum_{kl}(X_{ijklm}) + Y_{ijm}\right)^{3} + \chi_{jm}^{C} \tag{23}$$

$$X = \left[1, \frac{\alpha_{jm}^{T}}{\sum_{kl} (X_{ijklm}) + Y_{ijm}}, \left(\sum_{kl} (X_{ijklm}) + Y_{ijm} \right), \left(\sum_{kl} (X_{ijklm}) + Y_{ijm} \right)^{3} \right] \quad (24)$$

$$\hat{\beta} = \left(X^{T} X \right)^{-1} X^{T} Y \quad (25)$$

But Y in this case is:

$$Y = \left[C_{ijm} \right] \quad (26)$$

For determining the optimal batch size, we have

$$\frac{\partial t_{ijm}}{\partial x} = 0 \text{, where } t_{ijm} = \frac{T_{ijm}}{x} \text{, and } x = \left(\sum_{kl} X_{ijklm} + Y_{ijm} \right) \quad (27)$$

6.5.3 Estimation of parameters from the P&L data:

The distribution cost and the SMG&A are currently determined from the P&L data itself. This part can potentially be improved upon in the future. As in the previous section wherein the parameter values for the COGS are estimated using LSE, we estimate the distribution cost coefficients use LSE. The local and the exports distribution costs are estimated separately. We enhance the local distribution cost estimate by adding a component for change in numeric distribution from the Nielsen Data.

$$(DC)_{i} = \alpha^{dl} \sum_{jklm} (X_{ijklm}) + \alpha^{de} \sum_{jm} (Y_{ijm}) + \beta^{d} e^{C*(\Delta ND)_{i}} \quad (28)$$

Further, we linearize the system by breaking up the exponential term into two components and taking a linear combination of the two. The local distribution cost can therefore be split up as shown in Equation 24.

$$(DC)^{local} = \alpha^{dl} (X_{1}) + \beta_{1}^{d} (X_{2}) + \beta_{2}^{d} (X_{3}) + \varepsilon \quad (29)$$

Where

$$X_{1} = \sum_{jklm} X_{ijklm} \quad (30)$$

$$X_{2} = e^{(\Delta ND)} \quad (31)$$

$$X_{3} = e^{3*(\Delta ND)} \quad (32)$$

Note that the coefficients used in equations 27 and 28 need to be tuned. This equation is in again of the standard from presented in Equation 15

Where X comprise of 3 columns

$$X = \left[X_{1}, X_{2}, X_{3} \right],\ \hat{\beta} = \left(X^{T} X \right)^{-1} X^{T} Y \text{, and } Y = \left[DC^{local} \right] \quad (33)$$

Similarly, we can estimate the coefficients for the export market:

$$(DC)^{e} = \alpha^{de} (X_{1}) + \varepsilon \quad (34)$$

Where

$$X_1 = \sum_{jm} X_{ijm} \tag{35}$$

The estimation therefore comprises of:

$$X = [1, X_1],\ \hat{\beta} = \left(X^T X\right)^{-1} X^T Y \text{, and } Y = \left[DC^{\exp orts}\right] \tag{36}$$

The total distribution cost comprises of the terms in Equations 29 and 32

$$(DC)_i = (DC)^l + (DC)^e = \alpha^{dl} (\sum_{jklm} X_{ijklm}) + \alpha^{de} (\sum_{jm} Y_{ijm}) + \beta_1^d e^{(\Delta ND)} + \beta_2^d e^{3(\Delta ND)} \tag{37}$$

The estimation of SMG&A is more straightforward. It is again a LSE with the X and the Y terms being:

$$X = \left[1, \sum_{jklm} X_{ijklm}\right],\ \hat{\beta} = \left(X^T X\right)^{-1} X^T Y \text{, and } Y = [SMG \,\&\, A] \tag{38}$$

6.5.4 Estimation of parameters and function form of the market side

As mentioned earlier, the market side model is non-parametric and takes the utility formulation perspective

$$u_{ijt} = \delta_{jt}\left(x_{jt}, p_{jt}, \xi_{jt}; \theta_1\right) + u_{ijt}\left(x_{jt}, p_{jt}, v_i, D_i : \theta_2\right) + \varepsilon_{ijt} \tag{39}$$

It links the utility that the consumer perceives with the incentives that are offered with the product. Therefore the marketing mix variables: Numeric distribution, promotional spending, and media spending (x(jt)) are linked to the market share through this functional form. The price and the also explicitly modeled through the p(jt) term. The consumer demographics (D) and preferences (v) are also modeled. The functional form of δ taken in the model is as presented in Equation 36-the utility increases with the Numeric distribution, promotional spending, and media spending, whereas it decrease with the prices.

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} \tag{40}$$

For estimation purposes, the can be re-written as:

$$u_{ijt} = \lfloor -p_{jt}, x_{jt} \rfloor \left(\Pi D_i + \sum v_i\right) \tag{41}$$

We therefore have

$$A_{jt}\left(x_{.t}, p_{.t}, \delta_{.t}; \theta_2\right) = \left\{\left(D_i, v_i, \varepsilon_{i0t}, \varepsilon_{i1t}, \ldots \varepsilon_{iJt}\right) \mid u_{ijt} > u_{ilt} \forall l = 0,1,\ldots, J\right\}$$

and market share is:

$$s_{jt}\left(x_{.t}, p_{.t}, \delta_{.t}; \theta_2\right) = \int_{A_{jt}} dP(D, v, \varepsilon)$$

$$s_{jt}\left(x_{.t}, p_{.t}, \delta_{.t}; \theta_2\right) = \int_{A_{jt}} dP(\varepsilon \mid D, v) dP(v \mid D) dP(D)$$

the elasticities are therefore given by:

$$\eta_{jkt} = \frac{\partial s_{jt} p_{kt}}{\partial p_{kt} s_{jt}} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_i s_{ijt} \left(1 - s_{ijt}\right) dP(D) dP(v) & \text{if } \mathrm{j} = \mathrm{k} \\ \frac{p_{jt}}{s_{jt}} \int \alpha_i s_{ijt} s_{ikt} \, dP(D) dP(v) & \text{otherwise} \end{cases}$$

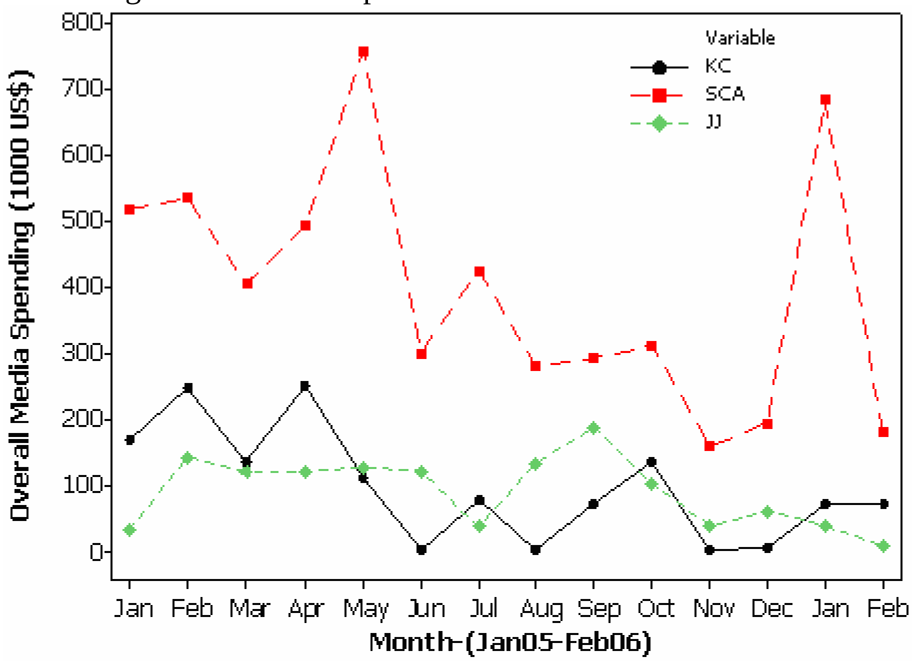
Estimation:

$$\underset{\theta}{Min} \left\| s\left(x, p, \delta(x, p, \xi; \theta_1); \theta_2\right) - S \right\|$$

6.5.5 Analysis: Media Spending

The media spending form a significant component of the marketing and sales budget for KC. Its main purposes is to increase visibility (and to a certain extent, awareness) for the brand, and the impact is therefore observed at the brand level.

The last 15 months or so (Jan 05-Feb 06) have witnessed a steady decrease in the media spending by all three of the major players in the fem care sector, namely SCA, KC, and J&J (Figure 1 displays the media spending-adjusted for devaluation). The reason for this has been attributed to the maturating of the market and the product offerings, and the maturating of the consumer preferences.

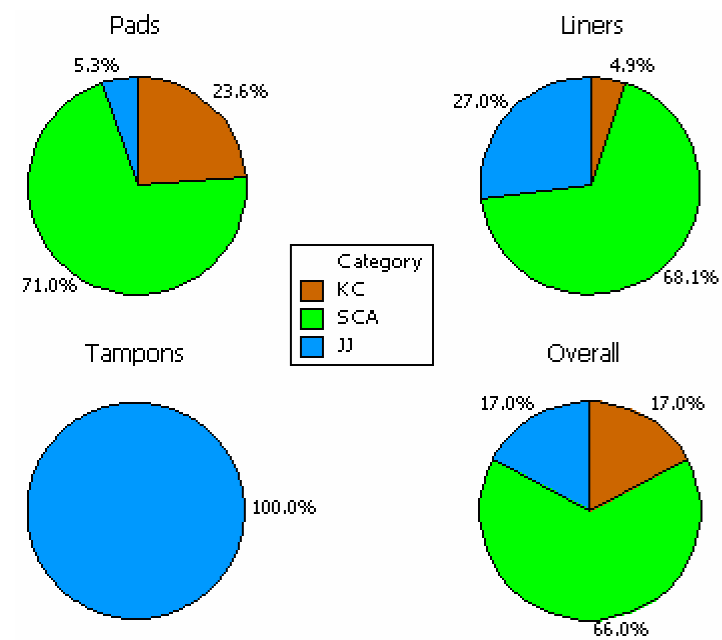


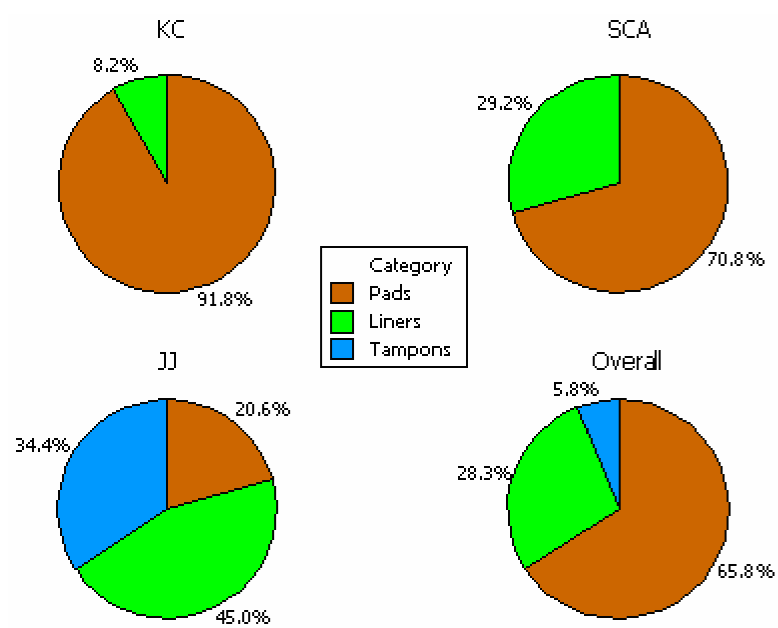
The focus of this section is to

1. Identify and quantify the impact that the media spending has on the volume share of the products in the Colombian market
2. Identify and quantify the competitor (individual firms) strategies so that we can isolate the impact of the media spending strategy that KC plans on adopting for the future

3. Based on the understanding gained from the previous two points, identify how a strategic use of media spending can be used to increase the market share in the seemingly mature market

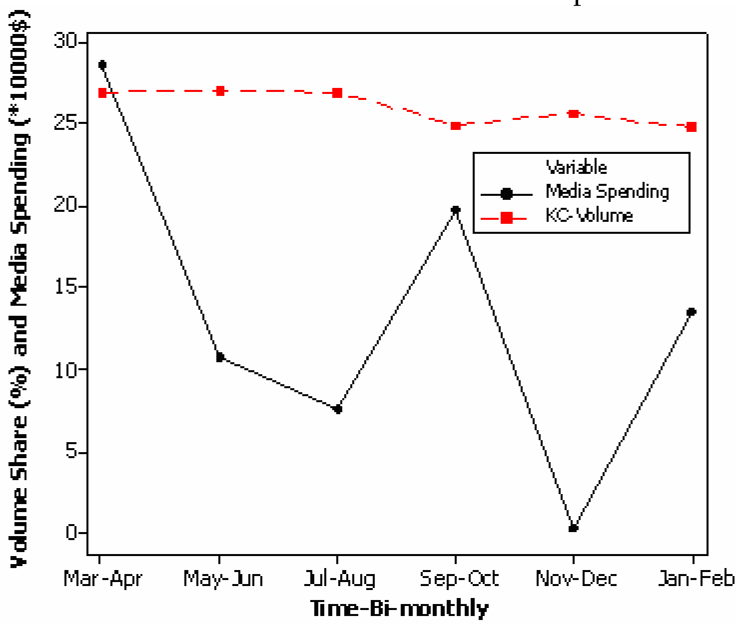
Overview
Approximately $7 million was spent on media by SCA, KC, and J&J combined in the year 2005. KC contributed about 17% to this spending ($1.2million) while SCAs share was around 66%. Further, 92% of the KC media spending budget was allocated to pads, while 0% was spent on Tampons. The distribution of the spending by KC, SCA, and J&J are presented in Figure 2.





Media Spending-Pads

Approximately 92% of the spending by KC on media (in 2005) was in the pads segment- It therefore is imperative to understand the impact of media spending within this segment. Statistical analysis shows that there is a high positive cross-correlation between the media spending by KC in a given time period (bi-monthly period) and the market share in the next time period. An increase in spending in a given time period implies that there is an increase in the market share in the next time period and vice versa.



Cross Correlation Function: KC-Volume Share, KC-spending

```
CCF - correlates KC-Volume Share(t) and KC-Media spending(t+k)

            -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
-2   0.194                             XXXXXX
-1   0.638                             XXXXXXXXXXXXXXXXX
 0   0.054                             XX
 1   0.229                             XXXXXXX
 2  -0.211                        XXXXXX
```

Further, it may be noted that the correlation is more or less negligible (statistically insignificant) for the other time periods-indicating the possible absence of a long time impact of the media spending. Not only that, if we look into the cross correlation between the media spending of KC with the market share of SCA, a relatively high negative correlation is observed-implying that a decrease in spending by KC implies an increase in the share of SCA in the next time period.

Cross Correlation Function: KC-spending, SCA-Share

```
CCF - correlates KC-spending(t) and SCA-Share(t+k)

            -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
-1  -0.452                  XXXXXXXXXXXX
 0  -0.270                      XXXXXXXX
```

```
 1   0.110                          XXXX
```

Therefore, increase in media spending in a given time period implies an increase in market share in the next time period for KC. However, a decrease in the media spending implies not only a decrease in the market share in the next time period, but also an increase in the market share taken up mainly by SCA. Another interesting observation is that the media spending by SCA and the market share by KC in the next time period are positively correlated-implying that even when SCA increases its media spending, market share of KC increases-implying that SCA follows KC media spending, but the impact of the following is low. So, there is a scope to increase media spending and gain market share.

Cross Correlation Function: SCA-spending, KC-Share

```
CCF - correlates SCA-spending(t) and KC-Share(t+k)

            -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
-1   0.433                          XXXXXXXXXXXX
 0   0.356                          XXXXXXXXXX
 1   0.572                          XXXXXXXXXXXXXXX
```

The questions that come up next are related to the sensitivities-how sensitive is the change in market share to this change in media spending and what is the connection between this change and the competitors change in strategy?

Currently, there isn't sufficient data to comment about the sensitivities- but the hypothesis is that it is more difficult to gain back the market share lost due to decrease in the media spending-therefore a higher resistance is encountered to restart the process.

Analysis of the media spending:
Assuming that KC comes up with an advertisement strategy (media spending): what can potentially be the response that the competitors will have? The focus of this section is on identifying these responses from the past data provided by Patricia.

Amongst other options, we will evaluate three obvious ones that the competitors might take (based on what they have done in the past)

1. SCA and JJ follow their own advertisement strategy (internally defined)-the strategy is not conditioned to the strategy KC takes
2. SCA and JJ respond to the KC strategy and match the investment made by KC; i.e., they follow KC strategy.
3. There is a mix of the two-they follow it to a certain extent, but stick to their own internal one to an extent as well.

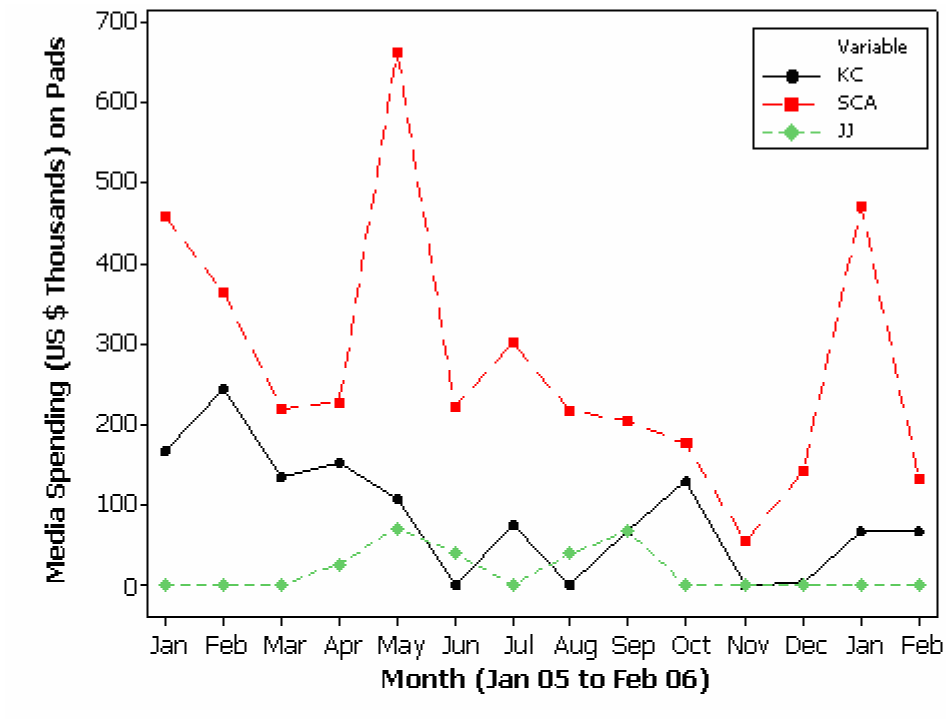
Having a sense of these strategies are important for determining how the market share will evolve over time in a competitive environment.

The three options imply:

1. The spending by KC, SCA, and JJ are uncorrelated, but there is auto-correlated
2. The spending by KC, SCA, and JJ are cross-correlated

3. There exists both cross and auto-correlation. They are related by simple descriptive models

We will look into the three options sequentially. The time series plot is as follows:



Cross Correlation Function: KC, SCA

```
CCF - correlates KC(t) and SCA(t+k)
                -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
                +----+----+----+----+----+----+----+----+----+----+
 -3   0.032                                XX
 -2   0.232                                XXXXXXX
 -1   0.108                                XXXX
  0   0.427                                XXXXXXXXXXXX
  1   0.093                                XXX
  2   0.075                                XXX
  3   0.674                                XXXXXXXXXXXXXXXXXX
```

Inferences:

1. KC follows the campaigns to a certain extent-however there is a lag of 2 time periods.
2. SCA matches the campaigns that KC comes up with in the same time period. However, it is arguable as to who follows whom-it can also be inferred as KC tries to match the campaigns by SCA.
3. There is high cross correlation at 3. This might be attributed to some seasonality function. Or is can be attributed to the fact that for every big effort by KC, SCA comes up with a big campaign after a lag of approximately 3 months. Further, this might be seen as the time it takes for the campaign to show significant results from a gain in market share perspectives.

Cross Correlation Function: KC, JJ

```
CCF - correlates KC(t) and JJ(t+k)
             -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
             +----+----+----+----+----+----+----+----+----+----+
 -4  -0.016                             X
 -3  -0.445                  XXXXXXXXXXXX
 -2  -0.252                       XXXXXXX
 -1  -0.093                           XXX
  0  -0.192                        XXXXXX
  1   0.004                             X
  2   0.207                             XXXXXX
  3   0.314                             XXXXXXXXX
  4   0.476                             XXXXXXXXXXXXX
```

Inferences:

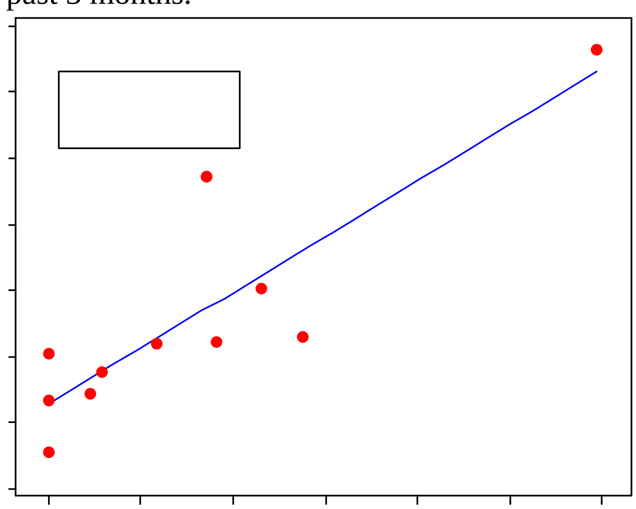
1. J&J follows the campaigns to a certain extent-however there is a lag of 2-4 time periods.

Cross Correlation Function: SCA, JJ

```
CCF - correlates SCA(t) and JJ(t+k)
             -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
             +----+----+----+----+----+----+----+----+----+----+
 -4   0.015                             X
 -3  -0.359                    XXXXXXXXXX
 -2  -0.284                      XXXXXXXX
 -1   0.021                             XX
  0   0.312                             XXXXXXXXX
  1   0.090                             XXX
  2  -0.053                            XX
  3   0.341                             XXXXXXXXXX
  4   0.635                             XXXXXXXXXXXXXXXXX
```

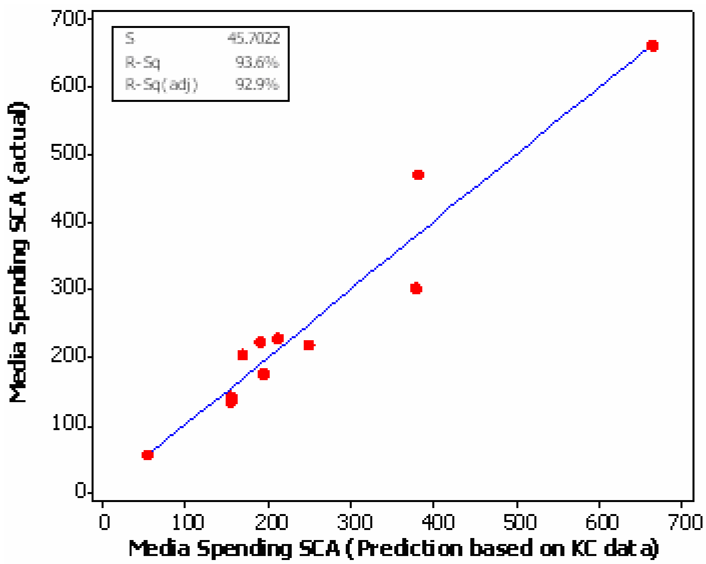
Prediction on the basis of the inferences

The current strategy by SCA can be predicted to a reasonable extent based on the media spending by KC in the past 3 months.



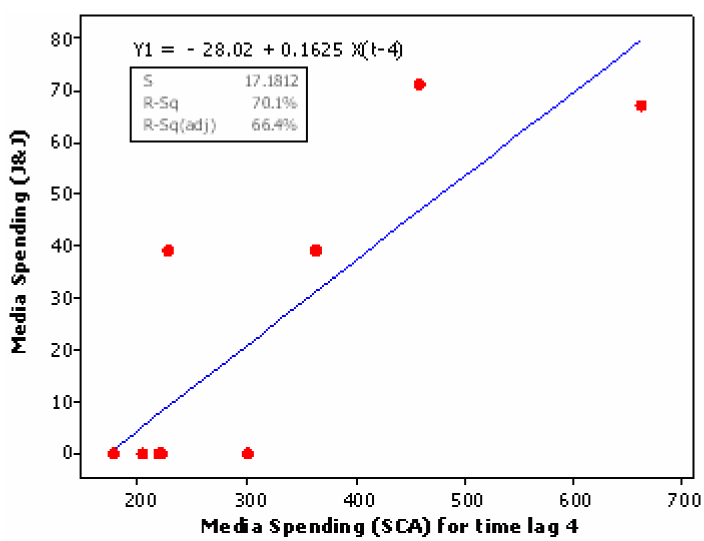
More intricate models can also be generated that have a better predicting capability: for instance, approximately 93% of the variability can be explained using the following model structure:

```
Y = 222 - 0.940 X(t-1) - 0.679 X(t-2) - 0.565 X(t-3) + 0.0137 X(t-3)^2
```



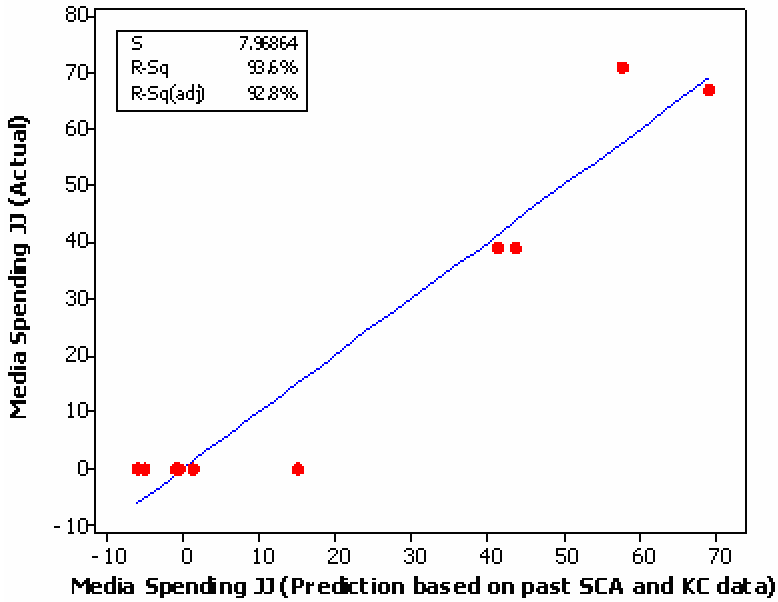

```
S = 55.9735   R-Sq = 93.6%   R-Sq(adj) = 89.4%
```

For the prediction of the J&J media spending, we use the SCA data-this is because apparently, J&J spending is based on the spending of SCA (the market leader) and not directly to KC



More intricate models can also be generated that have a better predicting capability: for instance, approximately 93% of the variability can be explained using the following model structure:

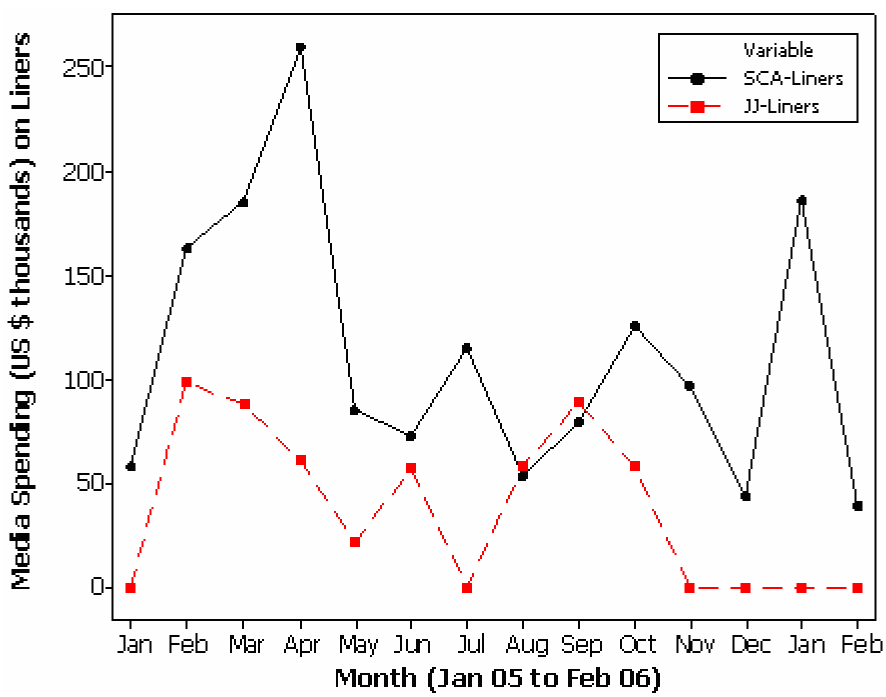
```
Y1 = - 46.1 + 0.120 X1(t-4) - 0.0262 X2(t-2) + 0.0634 X2(t-3) + 0.145 X2(t-4)
```




```
S = 10.0796   R-Sq = 93.6%   R-Sq(adj) = 88.4%
```

Analysis of media spending on Liners:

Note that in this case, the spending by KC is almost negligible. The spending split by SCA between pads and liners is not based on some intrinsic ratio, but is primarily driven by the spending by J&J on liners.



6.6 Conclusions

The traditional approach to production planning applies functional decomposition to planning tasks: first, process engineers locally optimize all process parameters, including the selected process plan and cutting speeds, and, only then, production engineers optimize the production quantities and schedules in order to obtain the maximum profit when all process variables are considered as fixed parameters. The AHFM-based integrative models try to identify variables that may have significant impact on the company profitability and decompose the planning problem accordingly.

The existing HPP approach creates a master and slave hierarchy where the solutions at one level form the starting point for decisions to be made at the next level. In AHFM approach, only “hard” parameters, such as tools and material characteristics are fixed, while most of the other parameters such as cutting speeds, feed rates, and depth of cut are considered for being soft parameters, modeled as global variables in the aggregate model. Handbook-based parameters are usually considered being “optimal”; however, they are usually based on local operational rules without a global view of long-term profitability.

This is a very basic thing. We can investigate the enhancements in the following aspects:

1. Extend the cascade controllers to take care of systems with different delays
2. Extend the cascade controllers to take care of parallel systems
3. Integrate the cascade controllers with the MPC architecture.

Chapter 7

On hindsight: A static perspective on events and the EM problem: definitions, classifications & implementation

7.1 Introduction

The planning problem can be viewed from both a static as well as a dynamic perspective. The static perspective is based on snapshots of the estimated states of the system at various instances of time, and if in case the static perspective is taken, the planning problem can, in general, be transformed into the standard LP form. Further, an event is defined as the cause or a consequence of a Δ change in one or more parameters/variables that are related to the initial state of the system, as defined in the planning problem. Starting with a planning problem formulated as a LP, post optimality analysis and synthesis techniques can potentially be of use to characterize the event types, the impact that they have, and the corresponding state transformations that these events cause. In this chapter, sensitivity and parametric analysis approaches have been used to provide a framework for classifying the events and the impact that they have on the plan. Sensitivity analysis is used to study the effect of discrete impact size Δ on different components of the planning problem. These universal characteristics have been used for classifying the event types into 5 categories.

Not only are the characteristics of the transformation due to the different event types different, but these characteristics change due to different impact sizes (Δ) as well. Parametric programming has been used to characterize this and the effect of the event is classified into 4 categories. Further, the approaches to deal with the impact of the events are also classified into 2.

The rest of the chapter is organized as follows: The planning and the EM problem are presented in Section 2.2. The sensitivity and parametric analysis solutions are presented in Sections 2.3 and 2.4 respectively. The classification of the events is presented in Section 2.3, while the classifications of the impact of events are presented in Section 2.4. A case study for the planning problem encountered on the manufacturing shop floor is presented in Section 2.5. Different types of events and their effects are simulated and studied in this section. Finally, the insights drawn that will form the basis of most of the work in the later chapters of the thesis are presented in Section 2.6. An extensive list of events, the category to which they belong, and the associated performance indicators has been developed and presented in the Appendix. Further, looking at the potentials of EM using post optimality analysis provides the foundations for several theory based implementations. One such application, namely the attributes charts for prioritizing event notifications is presented in the addendum to the chapter.

7.2 Events and the EM problem

The supply chain planning (SCP) problem can be stated as: given the demand forecast $F(t)$ for each period t in the planning horizon which extends over T periods, determine the production level $P(t)$, capacity $C(t)$, inventory level $I(t)$, marketing effort $M(t)$, work force level $W(t)$ for periods $t = 1,2,....,T$, which maximize the relevant profits over the planning horizon (Nahimas, 2000, Kogan and Khmelnitsky, 1995, Warburton, 2004). The production system, in general, is considered to be a production facility that produces a single product (multi-product firms are treated by aggregating individual products to some common unit such as cost, contribution, or direct machine hours). Time is divided into discrete periods and demand forecasts for each period are obtained by some appropriate forecasting technique. Capacity expansion and reduction, warehousing limitations, transportation delays, and SLRs are incorporated as constraints.

This problem can be represented in a general form presented in Equations (1-1) and (1-2). Where the production level $P(t)$, capacity $C(t)$, inventory level $I(t)$, marketing effort $M(t)$, and work force level $W(t)$ are part of the vector x. This solution is in fact the standardized LP formulation, and an entire body of literature has been developed for characterizing these problems and their solution strategies (Bazaara et. al., 2004) The solution to the problem x^p is the plan that is in place and needs to be executed. The elements of the vectors *A*, *b*, and *C* are indicative of the state of the system (define the system) and are called the state parameters. Together, the state parameters and the plan that is generated characterize the system and its state.

An event is now defined as the cause or a consequence of Δ change in one or more parameters/variables that are related to the initial state of the system. So if an event does occur, it will cause a Δ change in one or more parameters and variables that form a part of the planning problem (Equations (1-1) and (1-2)), i.e., will cause any of the transformations $A \rightarrow A', b \rightarrow b'$, or $C \rightarrow C'$. The change in the parameter value of A, b, or C is called the impact of the event.

This change in the parameter value of A, b, or C further implies that the operating point x^p, i.e., the solution to the planning problem, and/or the optimal value z^p changes and the original values may no longer be achievable or optimal. The new planning problem can be formulated as presented in (1-3) and (1-4). The solution y^* becomes the new operating plan y^P. EM problem is the study of the relationship between x^p and y^P given that a Δ change in the values of A, b, or C has occurred or can potentially occur.

Planning in business systems and supply chains has been one of the most active areas of application of OR based approaches in the past decade. Even though planning has received a lot of attention in this domain, the business world is rich with instance of exceptions to the planned. Since the cost associated with the breach of SLRs is high, there is a need to better understand the relationship between x^p and y^p.

7.3 Classification of events

The first thing that is needed to be known is given an impact of size Δ on the values of A, b, or C and causing the transformation $A \rightarrow A'$, $b \rightarrow b'$, or $C \rightarrow C'$, what is the new y^P (if it exists) and the relationship between x^p and y^P. Based on the description of the planning problem (Equations (1-1) and (1-2)) and the kind of impact that the event has on the planning problem and the parameter/variable that they affect, events can be classified into five categories, namely:

6. Those that change the cost coefficients
7. Those that change the capacity coefficients
8. Those that change the constraint coefficients
9. Those that add a new constraints
10. Those that add a new variable in the system

Discrete sensitivity analysis can help determine the effect that a delta change in one or more parameters will have (Bazaara, et. al, 2004). The analytical interpretations are based on the effect that the event will have on the optimal simplex tableau, which, at any iteration is of the form shown in Figure 7-1. Reference to this tableau is made for discussions pertaining to the analytical solution.

	z	x	RHS
z	1	$c_B B^{-1} a_j - c_j$	$c_B B^{-1} b$
x_B	0	$B^{-1} a_j$	$B^{-1} b$

Figure 7-1 Simplex tableau at any given iteration

The event categories and the effect that the events have are presented in the following table.

Type	Cause: Event Categories	Effect
1a	Change in cost coefficient (c_k) of a nonbasic variable	Only the coefficients of row 0 are affected. In particular, if c_k is changed to c'_k for nonbasic variable x_k, then only $z_k - c_k$ is affected. The primal optimality can be lost only when $z_k - c_k$ becomes negative. In that case, there is a need to perform additional simplex iterations to regain optimality. The optimal basis will also change. Graphically, when the cost changes, there is a corresponding change in the slope of the level curves (isoprofit lines).
1b	Change in cost coefficient (c_k) of a basic variable	If c_k is changed to c'_k for basic variable x_k, then c_B changes. Thus, in this case, all $c_B B^{-1} a_j - c_j$ and $c_B B^{-1} b$ are affected. If all $z_j - c_j \geq 0$, the current solution (extreme point) is still optimum, but the optimal objective value changes.

2	Change in RHS coefficient (b_j)	The addition of a new variable requires that updating of the current tableau by adding a new column with $z_j - c_j$ and $\alpha_j = B^{-1}a_j$. If $z_j - c_j \geq 0$ then the addition of the new variable x_j does not affect optimality and x_j becomes nonbasic in the optimal solution. Otherwise, the primal optimality is lost and there is a need to re-optimize using the primal Simplex method.
3a	Change in constraint coefficient (a_{ij}) for a nonbasic variable	If a_{ij} (x_j is nonbasic) is changed to a'_{ij}, then $B^{-1}a_j$ and $c_B B^{-1} a_j - c_j$ are affected. There may be a need to apply the primal simplex method to re-optimize.
3b	Change in constraint coefficient (a_{ij}) for a basic variable	If a_{ij} (x_j is basic) is changed to a'_{ij}, then B and therefore (B^{-1}) changes and the whole tableau is affected. There is a need to recomputed B^{-1} and update the whole tableau. This new tableau could be primal or dual infeasible, and so the impact is equivalent to solving the problem from the beginning.
4	Addition of a new variable	The addition of a new variable requires that updating of the current tableau by adding a new column with $z_j - c_j$ and $\alpha_j = B^{-1}a_j$. If $z_j - c_j \geq 0$ then the addition of the new variable x_j does not affect optimality and x_j becomes nonbasic in the optimal solution. Otherwise, the primal optimality is lost and there is a need to re-optimize using the primal Simplex method.
5	Addition of a new constraint	If the current x^* satisfies the new constrain, then the x^* remains optimal and the new constraint is not tight. If current x^* does not satisfy the new constrain, then there is a need to re-optimize. Row operations are also required to regain the canonical representation.

7.4 Parametric programming and classification of the impact

In the previous section, the effect of discrete impact size (Δ) and the effect of its variation were considered. However, the state transformations due to different event types have distinct characteristics, and not only are the characteristics of the transformation due to the different event types different, but these characteristics change due to different impact sizes (in terms of Δ mean shift that the event causes) as well. In this section, parametric programming is used for illustrating this effect of the continuous variation of the c_k and b_j values.

7.4.1 Systematic variation of cost vector c

This involves the replacement of c_k by c_k + tc_k. Since only systematic changes in the c vector are considered in the optimal tableau, only primal optimality (row 0 will be affected). In general, the primal optimality conditions are $z_j - c_j = c_B B^{-1} a_j \geq 0$ for all $j \in J$. The optimization problem is transformed into the form

Maximize $z(t) = (c + tc')x$ (7-1)

Subject to Equation (1-1) and

$-\infty \leq t \leq \infty$ (7-2)

Now primal optimality conditions are given by replacing c by $c + tc'$ to get

$(c_B + tc'_B)B^{-1}a_j - (c_j + tc'_j) \geq 0 \qquad j \in J$ (7-3)

$(c_B + tc'_B)B^{-1}a_j - (c_j + tc'_j) \geq 0 \qquad j \in J$ (7-4)

$(z_j - c_j) + t(z'_j - c'_j) \geq 0$ (7-5)

The objective function value becomes

$z + tz' = (c_B + tc'_B)B^{-1}b = c_B B^{-1}b + tc'_B B^{-1}b$ (7-6)

7.4.2 Systematic variation of capacity vector b

A change in b_j is equivalent to a change in the objective function coefficient of the dual. Analyzing the systematic variation of b is similar to analyzing the systematic variation of c vector.

Replacing b_j by b_j + t b_j, i.e.,

Maximize $z(t) = cx$ (7-7)

Subject to: $Ax = b + tb'$ and $x \geq 0$ (7-8)

$-\infty \leq t \leq \infty$ (7-9)

This change in b affects only primal feasibility (last column). It has no affect on the primal optimality (row 0). Feasibility conditions and objective function value for the original LP are

$x_B = B^{-1}b \geq 0$ (7-10)

$z = c_B B^{-1}b$ (7-11)

For the revised problem, the feasibility conditions and objective function value are:

$x_B = B^{-1}(b + tb') = B^{-1}b + tB^{-1}b' \geq 0$ (7-12)

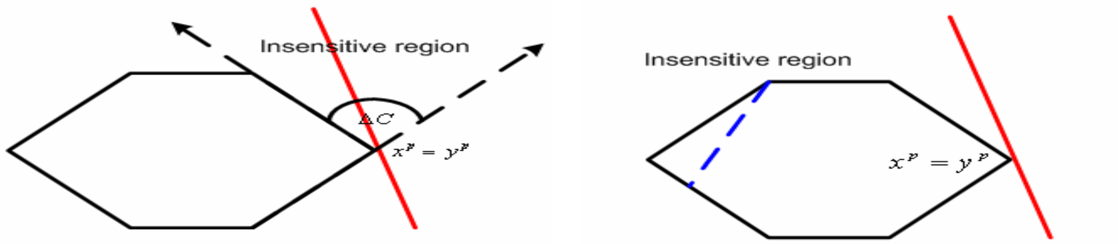
$z = c_B B^{-1}(b + tb') = c_B B^{-1}b + tc_B B^{-1}b'$ (7-13)

7.4.3 Classification of impact

Parametric programming provides the effect that the five categories of events have on the plan and the characteristics of the replanning problem in the solution space. Based on parametric programming, it can be seen that there are four distinct regions in which the impact of an event takes the existing plan.

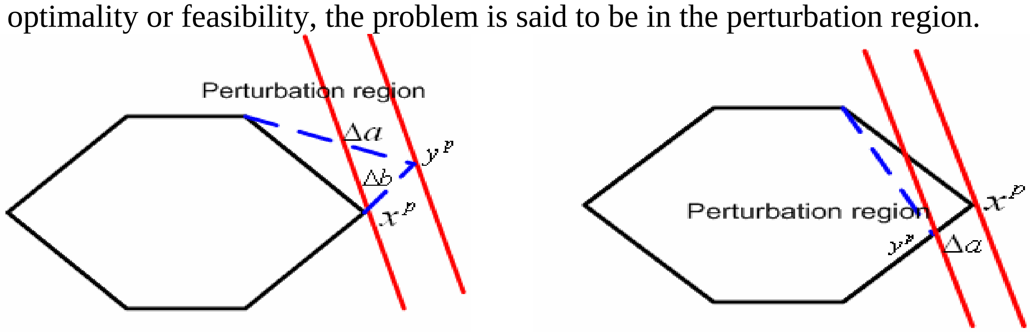
1. Insensitive region
2. Perturbation region
3. Reoptimization region
4. Infeasible region

The characteristics of these regions are as follows:
Insensitive region: When the basis vector remains the same after the impact of the event, and the problem does not require any further steps of either the simplex or the dual simplex algorithm to regain optimality, the event is said to lie in the insensitive region.



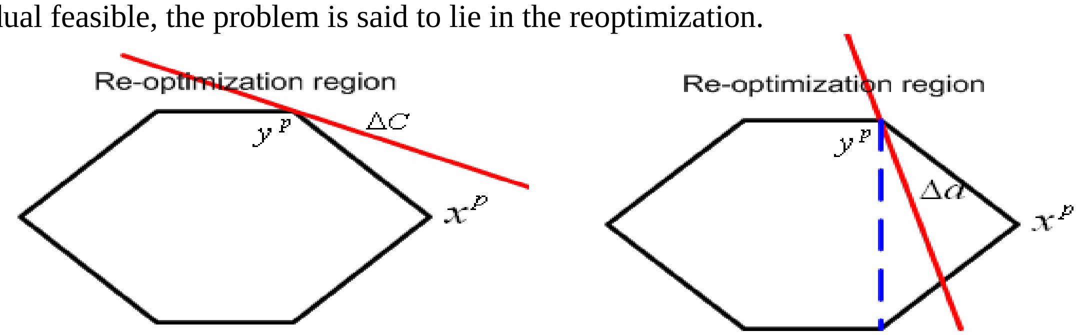
(a) Insensitivity to change in cost vector (b) Insensitivity to change in the capacity
Figure 7-2 Example of insensitive region in solution space

Perturbation region: When the basis vector remains the same after the impact of the event, but some steps of either the primal or the dual simplex algorithm are required to regain optimality or feasibility, the problem is said to be in the perturbation region.



(a) Sensitivity to change in capacity vector (b) Sensitivity to change capacity vector
Figure 7-3 Example of perturbation region in solution space

Reoptimization region: When the basis vector changes but the problem is either primal or dual feasible, the problem is said to lie in the reoptimization.



(a) Sensitivity to change in cost vector (b) Sensitivity to change in capacity
Figure 7-4 Example of re-optimization region in solution space

Infeasible region: When the solution to the problem no longer exists, then the event is said to take the system into the disruption region. There is a need for the reformulation of

the planning problem. Events can therefore be classified into four categories on the basis of the effect that they have on the existing plan.
The complexity of the problem and the solution strategy depends on the region in the solution space to which the impact of the event takes the system. Corresponding to the four categories of impact, there are three response categories, with the amount of work and complexity growing with each region.

1. Corresponding to the insensitive region, no response is required.
2. Corresponding to the perturbation and the re-optimization region, there is a need to find the solution to the replanning problem in the neighborhood of the previous solution. The optimal solution takes the best trajectory.
3. And corresponding to the infeasible region, reformulation and re-optimization is required.

Based on the four regions, the objectives of EM, and the economics, there are two distinct approaches to handling the events and their impacts, namely the passive and the proactive approaches.

7.5 Approaches to the EM problem

Given that an "event" causes a Δ change in one or more of the state variables, the expectations from EM primarily are two folds:

1. Prevention of the events from having an impact on the systems SLRs.
2. If prevention is not possible, then early response to limit the losses incurred due to the variance between the planned and the executed;

Corresponding to the two objectives, are two broad approaches to EM; the passive approach and the proactive approach.

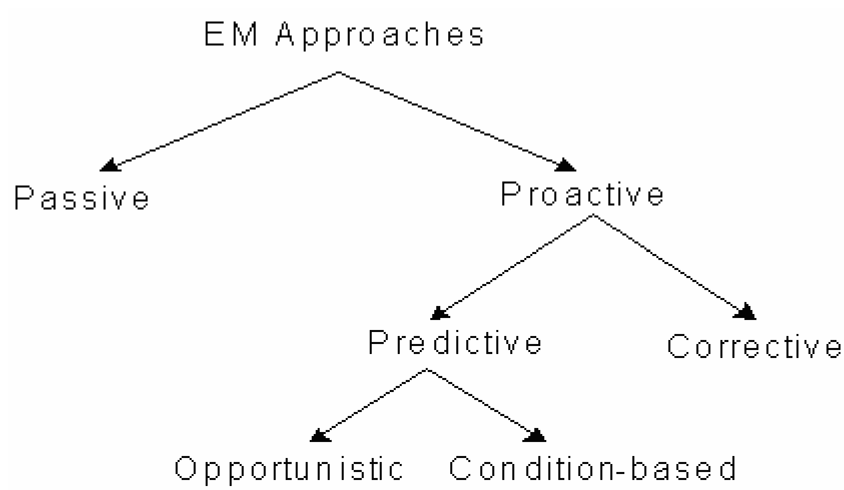


Figure 7-5 Approaches to EM

The **passive approach** aims at preventing the occurrence of the event, or at making plans that are insensitive to the event. Eliminate the possibility of occurrence of the event is the basic idea behind the approaches that are prescribed by the statistical process control. Concepts such as root cause analysis (RCA), continuous improvement, and reengineering fall in this category. Mathematically, this can be expressed as developing conditions that favor $P(E_i) \to 0$. Concepts of survivable and robust designs on the other hand fall under the second point. The idea is to introduce redundancy so that the impacts are absorbed and the problem remains in the insensitive region. Conservative estimates of the probability density function (PDF) of the monitored variable are taken by incorporating the possibility of events in the distribution itself. Mathematically, this idea can be expressed as $M' = \sum_i M \mid E_i \times P(E_i)$ and then use either M' or $E[M']$ for planning (as against using M), thereby giving the ability of disturbance rejection and uncertainty handling capabilities to the system.

The **proactive approach** addresses the issues that arise when the impact (or the potential impact) of the event takes the replanning problem into the perturbation, reoptimization or the infeasible region. The proactive approach relies on current information of the state of the system and estimates of the future state for proposing corrective actions. It is infact the application of optimal control theory and concepts to planning and execution in business systems. This essentially translates to the application of a controller that has the following two functionalities;

1. Estimation of the state of the system, and
2. Prescribing appropriate control action.

These functions correspond to the so called "sense" and "response" functionality, and describe the manipulation of the system state for the purpose of maximizing at least one of the measures of performance or minimizing a cost function.

Within the proactive approach itself, there are two sub categories: namely the corrective approach and the predictive approach to EM (Figure 7-5). The difference between the corrective and the predictive approach lies on the time of detection of the event and the time when the response was evaluated and put into effect. In the corrective approach, the occurrence of event is identified (or the response is evaluated) after the event has occurred, and in the predictive, the response is on the basis of anticipation of the occurrence of the event in the near future.

If in case the predictive approach is justified, and the time available to react before the event actually occurs is higher that the time required to respond and shift to the new plan, there is scope for condition based and opportunistic approaches to EM.

7.6 Selection of approach to EM

Several attributes must be taken into account at when selecting the EM approach for an event. This selection involves several aspects such as the investment required, failure costs, reliability of the approach, mean time between events (MTBE) and mean time to correct (MTTC) the problem. This section looks into the relationship between these attributes and the choice of the EM approach. The approaches considered herein are:

1. The passive approach
2. The predictive approach
3. The corrective approach

Some of the attributes that are important and need to be considered are as follows:

1. Criticality: Does the deviation take the process into region 1, 2, 3, or 4. How sensitive is the delta change to the transition from the process.
2. Available flexibility: Includes attributes such as available buffering while planning. The buffer may be in the form of spare capacity availability or spare part availability
3. Impact: Has two components: the cost and the frequency. The cost again has two components: the cost of repair and the cost due to the out of control process.
4. Predictability
5. Controllability: Has the MTTC as a component.
6. Propagation effect. The propagation effect takes into account the possible consequences of a machine failure on the adjacent equipment (domino effect).

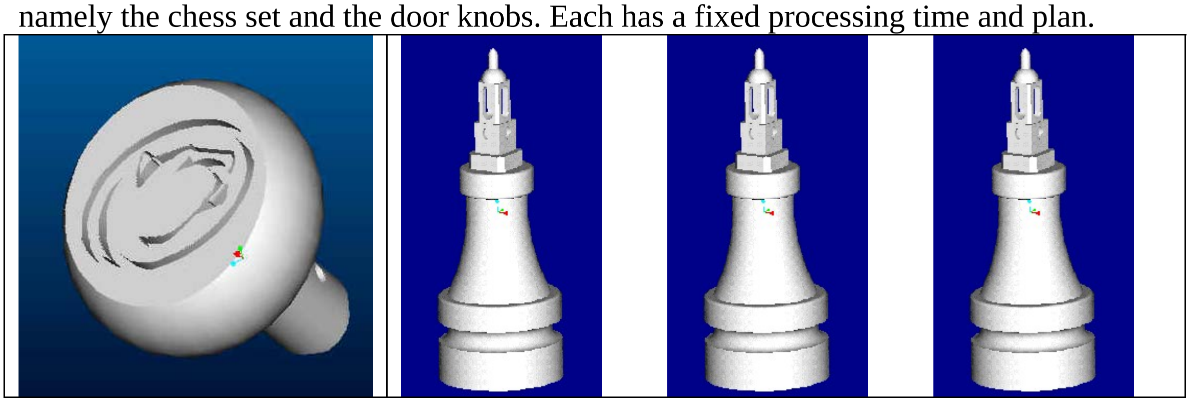
As an aid to the resolution of this problem, some multi-criteria decision making (MCDM) approaches are proposed in the literature. Almeida and Bohoris discuss the application of decision making theory with particular attention to multi-attribute utility theory. Triantaphyllou et al. suggest the use of Analytical Hierarchy Process (AHP).

The following three issues are critical for the success of the predictive approach:

4. The events need to be recurrent. If in case the event is non-recurrent, then predictive approaches loose a bulk of their utility unless the effect is similar to other events.
5. There should be detectable patterns in the data for us to be able to predict the occurrence of the impending events. This is a necessary condition. If in case we don't have any detectable pattern, then the predictability is lost and there is little that we can do. The logistic cost is infinity and we have to use either the corrective or the preventive approach to EM.
6. There should be sufficient time gap between the detection and time to respond. If in case we don't have this scenario, the returns will be low, as we will be operating very close to either the preventive approach or the corrective approach.

7.7 Putting things into perspective: case study

In this section, the post-optimality analysis is illustrated by integrating the execution system with the planning model. The aggregative production planner for manufacturing based on a real implementation in the Factory for Advance Manufacturing Education (FAME) lab at The Pennsylvania State University. The shop floor has two main products, namely the chess set and the door knobs. Each has a fixed processing time and plan.



Door knob	Chess Set

Figure 7-6. Example parts (the lion knob, and the PSU chess set)

7.7.1 The planning problem

Two cases have been considered for illustrating the sensitivity analysis and the parametric programming approach respectively:

The common assumptions regarding the system state are as follows:

1. Only work station 2 is available. Therefore there are three machines that are available, namely Haas VF-0E, VF-3, and turning center SL-20.
2. The daily production plans and the uncertainty therein is considered, wherein the machines are available for 300 time units (5 hrs of operational time) per day. Detailed ORS and process plans can be found in (Masin, et al. 2003). The objective is to maximize the profits.

Problem 1:

The manufacturer supplies the three chess pieces individually to the market, with each chess piece having different costs and production plans. The LP formulation for the planning in the sample problem is presented in equations 5-9.

Max $z = 10x_1 + 7x_2 + 6x_3$ (7-14)

s.t. $3x_1 + 2x_2 + x_3 \leq 36$ (7-15)

$2x_1 + 1x_2 + 1x_3 \leq 36$ (7-16)

$1x_1 + 1x_2 + 2_3 \leq 32$ (7-17)

Where x_1, x_2, and x_3 are the number of units of piece king, bishop and pawn manufactured respectively.

Problem 2:

Let x_1 be the number of units of product 1 (the door knob) produced and x_2 be the number of units of product 2 (chess set). The assumption is that whatever is produced is sold. Let the selling price of product 1 be 10.00, and that of product 2 be 40.00 per unit. The door knob (product 1) requires processing on the SL-20, and the VFOE, whereas the chess set (product 2) requires processing on all the three machines. The processing time per part is presented in Table 1. Based on the numbers, the production planning problem can be formulated as:

Table 7-1. Processing times of parts on the machines

Processing time	M1	M2	M3
Product 1	60	50	0
Product 2	30	50	75

Detailed ORS and process plans can be found in (Masin et al, 2003). The objective is to maximize the profits

Max $z = 1x_1 + 4x_2$ (7-18)

Subject to the constraint: $2x_1 + x_2 \le 10$ (7-19)

$x_1 + x_2 \le 6$ (7-20)

$x_2 \le 4$ (7-21)

$x_1, x_2 \ge 0$ (7-22)

7.7.2 Events: causes and affects

The events that can cause some problems in achieving the SLR are:

1. Machine failures, i.e., that translates to reduction in capacity that changes the right hand side (RHS) coefficient (b_j)
2. Quality regulations, i.e., that translates to the addition of a new constraint.
3. Addition of a new product/order, that translates to the addition of a new variable.
4. Operator related issues, that change the constraint coefficient (a_{ij})
5. Changes in the selling price of the product, which translates to the change in the objective coefficient (c_k) of a basic variable.

7.7.3 Sensitivity analysis approach: Problem 1

Cause: Event Categories	Effect
Change in cost coefficient (c_k) of a nonbasic variable If c_1 is changed from 10 to 14, i.e., c_1'=14, the only element that is affected is $z_1 - c_1$	c_1'=14, the only element that is affected is $z_1 - c_1$ As $z_1 - c'_1 \le 0$, the primal optimality is lost, and there is a need to perform additional Simplex iterations to regain optimality. The optimal basis will also change. However, for values of $c'_1 \in$[0,13], there is no change in the optimal basis.
Change in cost coefficient (c_k) of a basic variable. c_3 is changed from 6 to 5, the implications are that c_B changes to $c'_B = [7 \quad 0 \quad 5]$	the implications are that c_B changes to $c'_B = [7 \quad 0 \quad 5]$ In this case, all $c_B B^{-1} a_j - c_j$ and $c_B B^{-1} b$ are affected. Since all $(z_j' - c_j') \ge 0$, the current solution is still optimal. However the optimal objective function value changes to z' =138
Change in RHS coefficient (b_j) In this case, assuming that the availability of M3 is increased to 24, i.e., b_3 is changed from 24 to b_3'=24	The primal feasibility can be lost, but the dual feasibility is maintained. There is no change in the coefficients in the row 0. There usually is the need for the application of the dual simplex method to re-optimize. $z' = c_B x_B' = 156$
Change in constraint coefficient (a_{ij}) for a nonbasic variable. In this case, suppose the processing time of P3 on M1 is decreased to 1, i.e., a_{31} is changed from 2 to 1.	If a_{ij} (x_j is nonbasic) is changed to a'_{ij}**,** then $B^{-1}a_j$ and $c_B B^{-1} a_j - c_j$ are affected. There may be a need to apply the primal simplex method to re-optimize. The basis changes and therefore there is a need to change state. $(z_1' - c_1) = -2$

Change in constraint coefficient (a_{ij}) for a basic variable. suppose that the processing time of P2 on M3 is increased from 1 to 3, i.e, a_{23} is changed from 1 to 3	If a_{ij} (x_j is basic) is changed to a'_{ij}, then B and therefore (B^{-1}) changes and the whole tableau is affected. There is a need to recomputed B^{-1} and update the whole tableau. This new tableau could be primal or dual infeasible, and so the impact is equivalent to solving the problem from the beginning.
Addition of a new variable. Consider an addition of a new product type having a profit of 7 and processing times of 2, 3 on M1, M2 respectively.	The current tableau needs updating by the addition of a new column with $z_7 - c_7$ and $\alpha_7 = B^{-1}a_7$. If $z_7 - c_7 \geq 0$ then the addition of the new variable x_7 does not affect optimality and x_7 becomes nonbasic in the optimal solution. In this case, $z_7 - c_7 = c_B B^{-1} a_7 - c_7 = -1$ The primal optimality is lost and there is a need to re-optimize using the primal Simplex method.
Addition of a new constraint. Suppose a manufacturability constraint is added saying that the total productivity be greater than 20	$x_1 + x_2 + x_3 \geq 20$, then since $(0) + (14) + (8) = 22 > 20$, x^* satisfies the constraint, and there is no need for re-optimize. If however, the constraint that is added is $x_1 + x_2 + x_3 \leq 20$, x^* violates the new constraint. In this case there is a need for the performance of row operations to regain the canonical representation. The new tableau is primal infeasible, but dual feasibility is maintained. Dual simplex algorithm needs to be applied for re-optimization.

7.7.4 Parametric programming approach: Problem 2

The following cases are considered:

1. Cost parameters in the market. They change because of the fact that the orders come in from different sources, each having a different willingness to pay.
2. Capacity of the machines. They vary based on whether the machine is up or down, due to the tool changes, etc.

7.7.4.1 Systematic variation of the cost vector c

Consider the revised objective function:

Maximize $z(t) = (1+t)x_1 + (4-t)x_2$ with $-\infty \leq t \leq \infty$, where t is the parameter that defines the inflation rate. The quantity of interest is knowing how the optimal solution is affected as t is varied from $-\infty \leq t \leq \infty$

Range of t	Optimal Solution	Optimal Objective	Region
$-\infty \leq t \leq -1$	$X = (0,4,6,2,0)^T$	$Z(t) = 16 - 4t$	Re-optimization
$t = 0$	$X = (2,4,2,0,0)^T$	$Z(t) = 18$	Original Solution
$-1 \leq t \leq 3/2$	$X = (2,4,2,0,0)^T$	$Z(t) = 18 - 2t$	Insensitive Region

$3/2 \le t \le 7/3$	$X = (4,2,0,0,2)^T$	$Z(t) = 12 + 2t$	Re-optimization
$7/3 \le t \le \infty$	$X = (5,0,0,1,4)^T$	$Z(t) = 5 + 5t$	Re-optimization

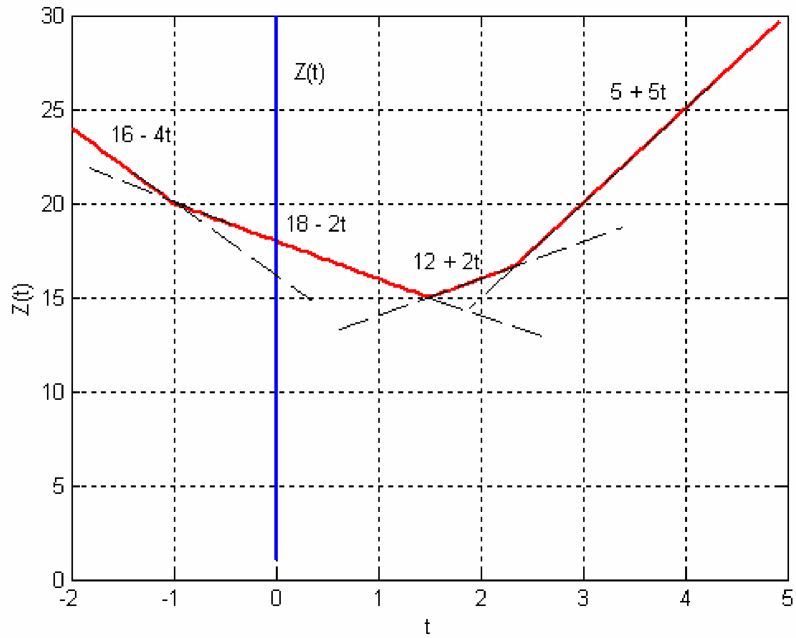


Figure 7-7. Transformation chart for systematic change in cost vector using parametric programming

7.7.4.2 Systematic variation of the capacity vector b

A change in the value of b is equivalent to a change in the objective function coefficient of the dual. Analyzing the systematic variation of b is similar to the systematic variation of c. Consider the following change in the b vector $b = [10,6,4]^T \rightarrow b' = [10,6+t,4-t]^T$
Note that the capacity cannot be negative. The part when the capacity becomes negative is not defined. In this case, it goes into the infeasible region.
Maximize $z(t) = x_1 + 4x_2$, subject to $Ax = b'$ with $-\infty \le t \le \infty$, where t is the parameter that defines the capacity changes. The interest here is in knowing how the optimal solution is affected as t is varied from $-\infty \le t \le \infty$
The optimal tableau at $t = 0$ is the same as presented in the earlier case

Range of t	Optimal Solution	Optimal Objective	Region
$-\infty \le t \le -6$	Infeasible	n.a.	Infeasible
$-6 \le t \le -1$	$X = (0,6+t,4-t,0,-2-2t)^T$	$Z(t) = 24 + 4t$	Re-optimization
$-1 \le t \le 2/3$	$X = (2+2t,4-t,2-3t,0,0)^T$	$Z(t) = 18 - 2t$	Perturbation
$2/3 \le t \le 4$	$X = (3+(1/2)t,4-t,0,-1+(3/2)t,0)^T$	$Z(t) = 19 - (7/2)t$	Re-optimization
$4 \le t \le \infty$	Infeasible	n.a.	Infeasible

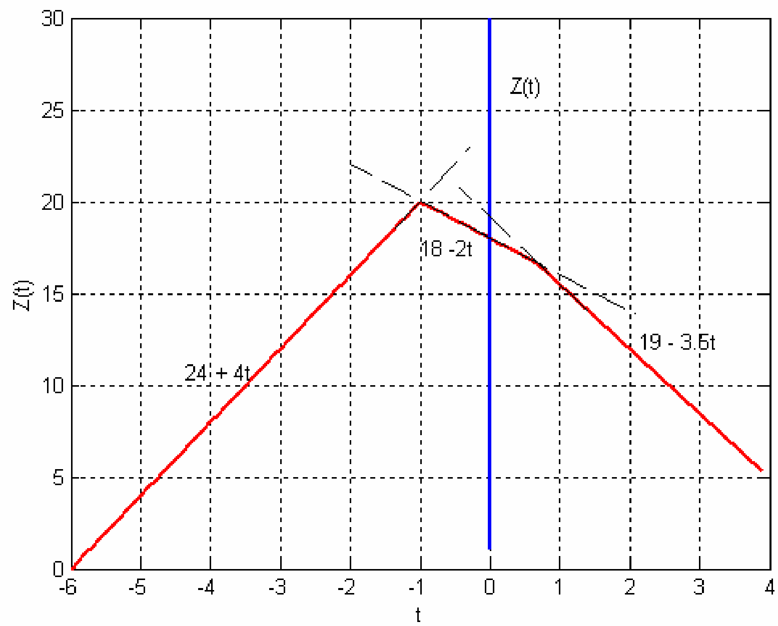


Figure 7-8. Transformation chart for systematic change in capacity vector using parametric programming

7.8 Conclusions and discussion

In general, the planning problem can be viewed from both a static as well as a dynamic perspective. The static perspective is based on snapshots of the state of the system (usually the estimated state, as the plans are made for the future) at various instances of time. Starting with a planning problem formulated as a LP, post optimality analysis and synthesis techniques were used to characterize the event types, the impact that they have, and the corresponding state transformations. Sensitivity analysis provides us with a framework for classifying the events and the impact that they have on the plan. The state transformations due to different event types have distinct characteristics, and need to be handled in different ways; that is one key idea that this chapter tries to put forth.

Based on the analysis and the requirements therein, the events have been classified into five while their impact the events are categorized into four. These four categories define the focus areas of the passive and the proactive approaches to EM taken up in the next chapter. It is important to understand these relationships so that quick actions can be determined (i.e., determine the solution to the replanning problem quickly, efficiently, and in general optimally), and design the preventive or predictive actions in accordance to the sensitivity of the impact and the system state.

Chapter 7 Addendum:

On the lookout for logistical anomalies: monitoring and prioritizing alerts for EM

7.9 Introduction

The need for a tighter coupling between p&e is gaining currency in industrial settings. It has led to the advent of EM systems that attempt to link the automated planning and execution systems. Conceptually, "smart" software agents are designed that monitor the execution of the plans and provide "alerts" when deviations between plan and execution are detected. However, a problem is encountered when attempts are made to introduce this smartness into the software agents: they have to be told when to provide these alerts, and how to associate priority levels with them.

The above mentioned problem is handled by using statistical process monitoring based approaches: albeit in a slightly ad hoc manner. Control limits are provided by the user, and fine tuned progressively to satisfy his or her needs. However, monitoring is not linked with the properties of the plan or the sensitivity of the plan to the deviation. That apart, seldom is priority associated with the alerts. Prioritization is an important issue from the managers' perspective, especially when he/she is inundated by the alerts messages, losses due to delayed responses can potentially be lowered by timely action.

The focus of this note is on addressing the issue of determining the control limits, linking it to the sensitivity of the plan to the deviation, and prioritizing the alerts on the basis of the effect that the deviation have on the overall system objective. The control limits are derived from post-optimality analysis of the plan that the monitored attribute has an impact on. Parametric programming is used on the plan, and the results are used for determining not only the control limits but for the integration of the prioritization of the alerts with the deviation that the alert signifies. The proposed parametric attribute charts have four distinct regions, namely, the insensitive, the perturbation, the re-optimization, and the infeasible with each region having specific priorities and alarm levels. Further, the approach removes the adhocness associated with the determination of the control limits as well as the priority given to the different classes of impacts that the manager observes in EM. The time consuming process of entering and fine-tuning the alarm bands and the priorities associated with the different alarm types when the event management system is installed can be automated.

The rest of the note is organized as follows: the concept behind the parametric control charts are presented in Section 7.10. The pre-requisites for understanding the concepts (description of events and the effects that they have on the monitored variables and the post optimality analysis) have already been presented in Chapter 7, and will not be discussed here. The case study discussed in Chapter 7 is enhanced to capturing the essence of the issues discussed so far is presented in section 7.11. Discussion and conclusions are presented in Section 7.12.

7.10 Parametric control limits and prioritization

There are three distinct parts in the proposed charting approach: The offline analysis of the planning problem, the online monitoring and data transformation, and the task prioritization aspect of data presentation.

7.10.1 Offline analysis of the planning problem

This comprises of performance of the post-optimality analysis discussed in the Chapter 2 for the static plan. Plans are not made or changed on the fly: system dynamics and the time delay involved prevent these activities. The same post-optimality analysis stays till a change in the plan is deemed necessary.

7.10.2 Online monitoring and data transformation

This comprises of the monitoring of the control variable. Control charts are widely used for controlling industrial processes, and tons of literature is available about the choice of control charts, the sampling frequency, the sample size, and other relevant data. Over the years, Shewhart control charts have been the most widely used control scheme in statistical process control (SPC), and we will be using the Shewhart control charts for deviation for illustration purposes. Based on the particular problem specifics, we can shift from the Shewhart chart, to the EWMA, or CUSUM, or whatever may be of interest. The transformation however is for the individual observations. Recently a lot of work has been done on identifying the performance indicators, the monitoring of operations, and the concept of six sigma supply chain.

7.10.3 Task prioritization

In this case, everything is converted to a single cost function. Therefore the task prioritization comprises of simple sorting of the tasks on the basis of the cost that the production unit can incur if the problem is not solved. The task prioritization however becomes more involved for the multi-criteria case, and inputs are required from the individual user for the task prioritization.

The detailed implementation of the approach is illustrated in the case study presented in the following section.

7.11 Case Study

7.11.1 The problem setting

The problem and the settings presented in Chapter 2 (Problem 2) will be taken up here.

7.11.2 Offline analysis of the planning problem

This analysis is presented in Chapter 2 problem 2. This has the two components

1. Systematic variation of the cost vector c
2. Systematic variation of the capacity vector b

Range of t	Optimal Solution	Optimal Objective	Region
$-\infty \le t \le -1$	$X = (0,4,6,2,0)^T$	$Z(t) = 16 - 4t$	Re-optimization
$t = 0$	$X = (2,4,2,0,0)^T$	$Z(t) = 18$	Original Solution
$-1 \le t \le 3/2$	$X = (2,4,2,0,0)^T$	$Z(t) = 18 - 2t$	Insensitive Region
$3/2 \le t \le 7/3$	$X = (4,2,0,0,2)^T$	$Z(t) = 12 + 2t$	Re-optimization
$7/3 \le t \le \infty$	$X = (5,0,0,1,4)^T$	$Z(t) = 5 + 5t$	Re-optimization

Range of t	Optimal Solution	Optimal Objective	Region
$-\infty \le t \le -6$	Infeasible	n.a.	Infeasible
$-6 \le t \le -1$	$X = (0,6+t,4-t,0,-2-2t)^T$	$Z(t) = 24 + 4t$	Re-optimization
$-1 \le t \le 2/3$	$X = (2+2t,4-t,2-3t,0,0)^T$	$Z(t) = 18 - 2t$	Perturbation
$2/3 \le t \le 4$	$X = (3+(1/2)t,4-t,0,-1+(3/2)t,0)^T$	$Z(t) = 19 - (7/2)t$	Re-optimization
$4 \le t \le \infty$	Infeasible	n.a.	Infeasible

7.11.3 Online monitoring and data transformation

The difference between the planned and the actual, i.e., the value of t is observed as a function of time. Shewhart control charts have been the most widely used control scheme in statistical process control (SPC), and the same will be used here for illustration purposes.

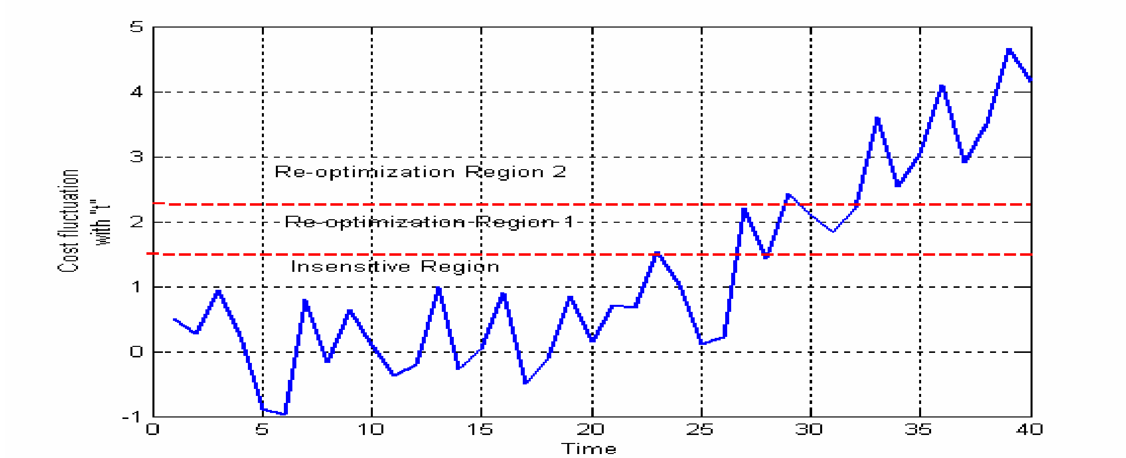


Figure 7-9 Transformed monitoring chart for systematic change in cost vector

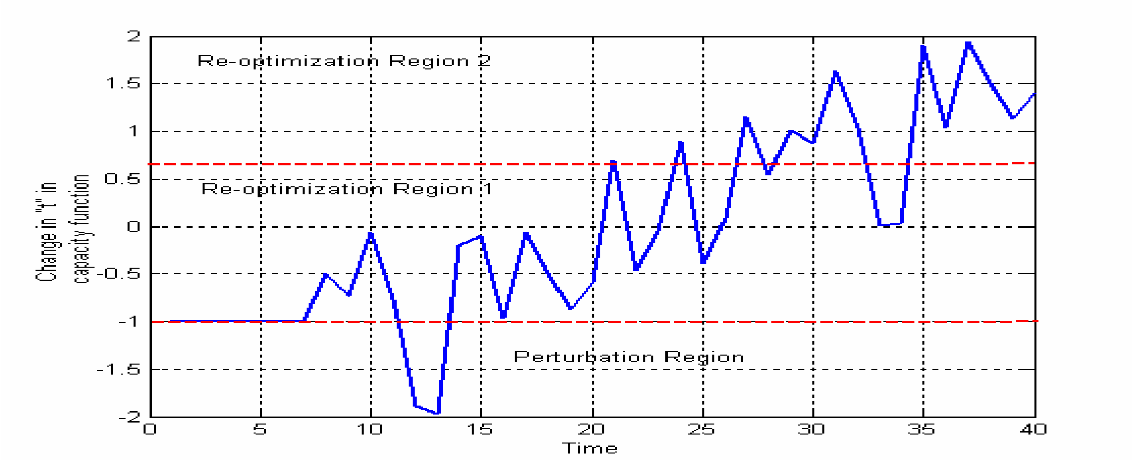


Figure 7-10: Transformed monitoring chart for systematic change in capacity vector

7.11.4 Task prioritization

Based on the priority values determined in the previous section, the task prioritization can be achieved by sorting the cost functions on the basis of the weights determined in the previous step. This is as shown in Figure 7-11.

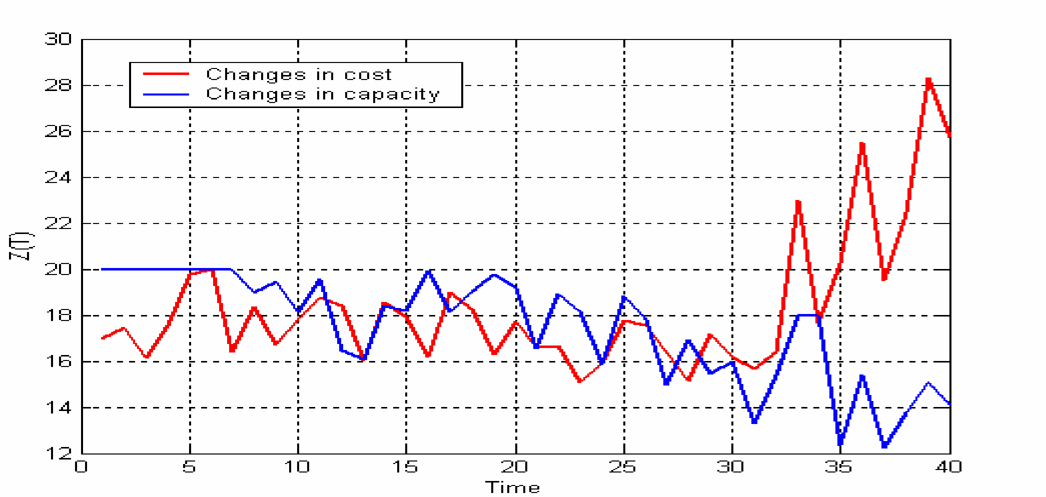


Figure 7-11 Task priority and significance of the deviation

7.12 Conclusions and discussion

The main problem is that the planning systems use mathematical models that are based on snapshots of the state of the system (usually the estimated state, as the plans are made for the future) at various instances of time, whereas the execution systems capture time persistent data. There is therefore a need to infer the affect that the current state of the system (based on input the time persistent inputs from the execution systems) has on the optimality and feasibility of the plan that was generated using the mathematical models. But, the current concept of use of control charts for monitoring the system state is not very effective in doing that.

When the logistical variables are being monitored, there is a need to know the severity of the deviation. The severity is divided into four regions by the sensitivity and the parametric programming approach. Further, the control charts that monitor the deviation, with the control limits determined by the parametric programming approach has been presented. Control limits are derived from post-optimality analysis of the plan that the monitored attribute has an impact on. This approach allows direct integration of the prioritization of the alerts with the deviation that the alert signifies.

The proposed approach removes the adhocness associated with the determination of the control limits as well as the priority given to the different classes of impacts that the manager observes in EM systems. Further, the time consuming process of entering and fine-tuning the alarm bands and the priorities associated with the different alarm types when the EM system is installed can be automated.

References

1. http://www.highjumpsoftware.com/solutions/supply-chain-event-management.asp
2. http://www.sap.com/solutions/business-suite/scm/pdf
3. http://itresearch.forbes.com/rlist/term/Supply-Chain-Event-Management.html
4. http://www-1.ibm.com/solutions/businesssolutions/doc/content/solution/642839107.html
5. http://www.sapient.com/pdfs/what/Sapient_SCEM_POV_.pdf
6. http://www.viewlocity.com/
7. Adler, I., Resende, M. G. C., Velga, G., and Karmarkar, N. (1989), “An implementation of Karmarkar’s algorithm for linear programming”, Mathematical Programming, Vol 44, No. 3, pp. 297–335.
8. Akella, R. (1989). “Real time part dispatch in flexible assembly test and manufacturing systems”, Operations research models in flexible manufacturing systems, Wien: Springer, pp. 1–73.
9. Akturk, M.S., and Avci, S. (1996) Tool allocation and machining conditions optimization for CNC machines, European Journal of Operations Research, Vol 94, pp. 335-348.
10. Akturk, M.S., and Ozkan, S. (2001) Integrated scheduling and tool management in flexible manufacturing systems, Int. J. Prod. Res., Vol 39, No. 12, pp. 2697-2722.
11. Almeida, A.T., and Bohoris, G.A., 1995, “Decision theory in maintenance decision making”. Journal of Quality in Maintenance Engineering, pp. 39–45.
12. Anderson, T.W., 1984, “An Introduction to Multivariate Statistical Analysis”, New York: Wiley.
13. Aviv, Y., and A. Federgruen. 1997. Stochastic Inventory Models With Limited Production Capacity And Periodically Varying Parame-Ters. Probab. Engrg. Inform. Sci. 11(2) 107–135.
14. Baganha, Manuel P. And Morris A. Cohen, 1998, “The Stabilizing Effect Of Inventory In A Supply Chains,” Operations Research, Vol. 46, No. 3, 1998, pp. 72- 83
15. Barnes-Schuster, D., Y. Bassok, and Anupindi, R., 2000, ``Optimized Deliver Leadtime / Inventory Placement in a Two-Stage Production/Distribution System", submitted to European Journal of Operations Research
16. Bazaraa, M.S., Jarvis, J.J., and Sherali, H.D., Linear Programming and Network Flows, Wiley-Interscience; 3 edition (December 3, 2004), ISBN: 0471485993
17. Benjaafar, S., (1996)“On production batches, transfer batches and lead times”, IIE Transactions, Vol. 28, No. 5, pp. 357-362.
18. Bertsimas, D., and Sim, M., 2003, “Robust discrete optimization and network flows”, Mathematical Programming Series B, v98, pp 49-71.
19. Bertsimas, D., and Sim, M., 2004, “The price of robustness”, Operations Research, Vol. 52, No. 1, pp 35-53.
20. Bertsimas, D., and Thiele, A., 2004, “A robust optimization approach to supply chain management”, http://web.mit.edu/dbertsim/www/papers.html.

21. Bevilacquaa, M., and Bragliab M., 2000, "The analytic hierarchy process applied to maintenance strategy selection", Reliability Engineering & System Safety, Vol 70, pp 71-83.
22. Bitran, G. R. and Morabito, R., 1996, "Open queueing networks: optimization and performance evaluation models for discrete manufacturing systems," Production and Operations Management, 5(2), 163-193.
23. Bonvik, A. M., C. E. Couch and S. B. Gershwin, 1997, "A comparison of production-line control mechanisms," International Journal of Production Research, 35(3), 789-804.
24. Boser, B. E., Guyon, I., and Vapnik, V. N., "A training algorithm for optimal margin classifiers," Proceedings of the Fifth Annual Workshop of Computational Learning Theory, 1992, Pittsburgh, PA, USA, ACM, 144-152.
25. Braun, R. D.; and Kroo, I. M., (1995), "Development and Application of the Collaborative Optimization Architecture in a Multidisciplinary Design Environment", Multidisciplinary Design Optimization: State of the Art, N. Alexandrov and M. Y. Hussaini, eds., SIAM J. Optimization.
26. Brooke, A., Kendrick, D., Meeraus, A., and Raman, R., (1998), "GAMS: A Users Guide", GAMS Development Corporation, http://www.gams.com/.
27. Buzacott, J. A., and J. Shanthikumar, 1993, "Stochastic Models of Manufacturing Systems," Prentice Hall, Englewood Cliffs, New Jersey 07632.
28. Cachon, Gerard P. And Marshall Fisher, 2000, "Supply Chain Inventory Management And The Value Of Shared Information" Management Science, Vol.46,No.8,August2000pp.1032–1048
29. Cachon, Gerard P. And Martin A. Lariviere, 1999, "Capacity Choice And Allocation: Strategic Behavior And Supply Chain Performance," Management Science, Vol 45, No 8, 1999, Pp. 1091-1108
30. Cachon, Gerard P. And Paul H. Zipkin, 1999, "Competitive And Cooperative Inventory Policies In A Two-Stage Supply Chain," Management Science, Vol. 45, No. 7, 1999, Pp.936-953
31. Cachon, Gerard P., 1999, "Managing Supply Chain Demand Variability With Scheduled Ordering Policies," Management Science, Vol. 45, No. 6, 1999, Pp. 843-856
32. Caramanis, M., 2000, "Multiple Time Scale Decomposition in the Production Planning of Stochastic Manufacturing Supply Chain Systems", Working paper, http://www.bu.edu/pcms/docs/mcc/mtsdppsc.pdf
33. Caramanis, Michael C., And Osman M. Anli, 1998a, Dynamic Production Planning And Control For Enterprise Integration," Proc. Of WODES98, 1998, Cagliari, Italy
34. Caramanis, Michael C. And Osman M. Anli, 1998b, "Manufacturing Supply Chain Coordination Through Synergistic Decentralized Decision Making," Proc. Of Rensselaer's International Conference On Agile, Intelligent, And Computer-Integrated Manufacturing, 1998
35. Caramanis, Michael C. And Osman M. Anli, 1999a, "Modeling Load-Dependent Lead Time Constraints For Efficient Supply Chain Coordination: Convexity Issues" Proceedings of the IEEE Conference on Decision and Control, Phoenix, AZ, December 7-10, 1999, p900-906

36. Caramanis, Michael C., and Osman M. Anli, 1999b, "Dynamic Lead Time Modeling for JIT Production Planning", Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, MI, May 10-15, 1999b, v2, p1450-1455
37. Caramanis, Michael C., Ionnis Ch. Paschalidis, And Osman M. Anli, 1999, "A Framework For The Decentralized Control Of Manufacturing Enterprise," Proceedings of the 1999 DARPA-JFACC Symposium on Advances in Enterprise Control, San Diego, CA, November 15-16, 1999, p99-109.
38. Caraminis, M., 1998, "Production planning under dynamic lead times and distributed production control," Proceedings of the NSF Grantees Conference in Design and Manufacturing, 1998, pp. 291-292.
39. Caramanis, M. and Sharifnia, A., 1991, "Near Optimal Manufacturing Flow Controller Design," International Journal of Flexible Manufacturing Systems, Vol. 3, pp. 321-336, 1991.
40. Carver, R.H., (2003) "Doing Data Analysis with MINITAB 14", Duxbury Press; 2 edition.
41. Chang, T.C., Wysk, R. A. and Wang, H.P., (1992) Computer Aided Manufacturing, Prentice-Hall, Inglewood, NJ.
42. Chen, Fangruo 1999a, "Decentralized Supply Chains Subject To Information Delays," Management Science/Vol. 45, No. 8, August 1999, pp. 1076–1090
43. Chen, Fangruo, 1999b, "94%-Effective Policies For A Two-Stage Serial Inventory System With Stochastic Demand," Management Science, Vol. 45, No. 12, 1999, Pp. 1679-1696
44. Chen, Frank, Zvi Dresner, Jenifer K. Ryan, David Simchi-Levi, 2000, "Quantifying The Bullwhip Effect In A Simple Supply Chain: The Impact Of Forecasting, Lead Times, And Information", Management Science Vol. 46, No.3, March 2000, 436-443
45. Cheng, D. W. and Yao, D. D., 1993, "Tandem queues with general blocking: a unified model and comparison results," Discrete Event Dynamic Systems: Theory and Applications, 2, 207-234.
46. Cho, S. and Prabhu, V.V., 2002a, "A Vector Space Model for Variance Reduction in Single Machine Scheduling," IIE Transactions, Vol. 34, No. 11, pp 933–952, November 2002.
47. Cho, S. and V.V. Prabhu, 2002b, "Sliding Mode Dynamics in Continuous Feedback Control for Distributed Discrete-Event Scheduling," Automatica, Vol. 38, No. 9, pp 1499-1515, September 2002.
48. Cho, S. and V.V. Prabhu, 2001, "Distributed Control of Machine Capacity and Production Scheduling to be presented at NAMRC, 2001, Gainesville, FL.
49. Cho, S., 2000, Unification of Distributed Scheduling and Machine Capacity Control, Ph. D thesis, Pennsylvania State University, 2000.
50. Clausen, J., Larsen, J., Larsen, A., and Hansen, J., 2001, "Disruption Management-Operations Research between Planning and Execution", OR MS Today, Vol. 28, No. 5, pp. 40-43.
51. Coddington, E.A. and Levinson, N., 1955, "Theory of Ordinary Differential Equations," McGraw-Hill, 1955.

52. Connors, D., Feigin, G. and Yao, D., 1994, "Scheduling semiconductor lines using a fluid network model," IEEE Transactions on Robotics and Automation, v 10 n 2, pp. 88-98, April 1994.
53. Cortes, C., and Vapnik, V., 1995, "Support Vector Networks," Machine Learning, 20:3,273-297.
54. Cristianni, N., and Shawe-Taylor, J., 2000, An Introduction to Support Vector Machines and other Kernel-based Learning Methods, Cambridge University Press, Cambridge, UK.
55. Clark, Kim B. and Wheelwright, S.C., (1993) Managing New Product and Process Development, Free Press, New York, NY.
56. Dallery, Y. and Gershwin, S. B., 1992, "Manufacturing flow line systems: a review of models and analytical results," Queueing Systems, 12, 3-94.
57. Davis, W. J., & Thompson, S. D. (1993). Production planning and control hierarchy using a generic controller. IIE Transactions, 25(4), 26–45.
58. Dejonckheere, J., S. M. Disney, et al. (2003). "The dynamics of aggregate planning." Production Planning and Control 14(6): 497-516.
59. Dewan P. and Joshi, S., 1998, "Distributed Control of Manufacturing Systems with Due date Constraints", Proceedings of the 6th Industrial Engineering Research Conference, Banff, Canada, 1998.
60. Duffie, N.A., Prabhu, V.V., and Kaltjob, P., 2003, "A Closed-Loop Approach To Real-Time Distributed Scheduling of Heterarchical Manufacturing Systems," Journal of Manufacturing Systems, Vol. 21, No. 6, pp. 409-418, 2002.
61. E.M.Goldratt.TheGoal.NorthRiverPress,GreatBarrington,Massachusetts, 1984.
62. Ettl, Markus, Gerald E. Feigin, Grace Y. Lin, and David D. Yao, 2000, "A Supply Network Model With Base-Stock Control And Service Requirements," Operations Research, Vol. 48, No. 2, 2000, Pp. 216-232
63. Federgruen,A. and P. Zipkin. 1986a. An Inventory Model With Limited Production Capacity And Uncertain Demands, I: The Average Cost Criterion. Math. O. R. 11, 193–207.
64. Federgruen,A. and P. Zipkin. 1986b. An Inventory Model With Limited Production Capacity And Uncertain Demands, Ii: The Discounted Cost Criterion. Math. O. R. 11, 208–215.
65. Feigin, G.E., 1998, "Inventory Planning In Large Assembly Supply Chains", in Tayur, S., Ganeshan, R. and Magazine, M. (edt.), 1998, "Quantitative models for supply chain management", Kluwer, Norwell, MA, U.S.A.
66. Filippov, A.F., 1960, "Differential Equations with Discontinuous Righthand Sides," American Mathematical Society Translations, Vol. 42, pp. 199-231, 1960.
67. Filippov, A.F., 1988, "Differential Equations with Discontinuous Righthand Sides," Kluwer, 1988.
68. Gallego, G., and Scheller-Wolf, A., 1998, "Capacitated Inventory Problems with Fixed Order Costs: Some Optimal Policy Structure," to appear, European Journal of Operations Research
69. Gavirneni, Sriganesh, Roman Kapuscinski, and Sridhar Tayur, 1999, "Value Of Information In Capacitated Supply Chains," ManagementScience/Vol.45,No.1,January1999, pp. 0025-1909

70. Gershwin, S. B. and J. E. Schor, 2000, ``Efficient Algorithms for Buffer Space Allocation," Annals of Operations Research, Volume 93, pp 117-144, 2000.
71. Gershwin, S. B., 1994, "Manufacturing Systems Engineering," PTR Prentice Hall, Englewood Cliffs, NJ.
72. Gershwin, S.G., 1989,"Hierarchical flow control: a framework for scheduling and planning discrete events in manufacturing systems," IEEE Proceedings, Vol. 77, No. 1, pp. 195-208, 1989.
73. Gershwin, S. B. (1988). A hierarchical framework for discrete event scheduling in manufacturing systems. Proceedings of IIASA Conference on Discrete Event System: Models and Applications (Sopron, Hungary, 1987), Berlin: Springer, pp. 197–216.
74. Glasserman Paul And Yashan Wang, 1998, "Leadtime-Inventory Trade-Offs In Assembly-To-Order Systems", Operations Research, Vol.46, No. 6, November-December, 1998
75. Glasserman, P. and Tayur, S., 1996, "the Stability of a Capacitated, Multi-Echelon Productin-Inventory System". Naval Research Logistics 43, 41-58.
76. Glasserman, P. and Yao, D. D., 1994, "Structured buffer allocation problems," Working paper, Columbia University, New York, NY.
77. Glasserman, P., 1998, "Service Levels And Tail Probabilities In Multistage Capacitated Production Inventory Systems",in Tayur, S., Ganeshan, R. and Magazine, M. (edt.), 1998, "Quantitative models for supply chain management", Kluwer, Norwell, MA, U.S.A.
78. Goldrat, E.Y and Jeff Cox, 1984 "The goal: excellence in manufacturing", New York : North River Press, c1984.
79. Gou, L., Luh, P.B., and Kyoya, Y., 1999, "Holonic manufacturing scheduling: architecture, cooperation mechanism, and implementation," to appear in Computers in Industry, 1999.
80. Graves, Stephen C. And Sean P. Willems, 2000, "Optimizing Strategy Safety Stock Placement In Supply Chain," Manufacturing & Service Operations Mangement, Vol. 2, No. 1, 2000, Pp. 68-83
81. Gstettner, S. and Kuhn, H., 1996, "Analysis of production control systems kanban and CONWIP," International Journal of Production Research, 34(11), 3253-3273.
82. Hoitomt, D. J., Luh, P. B., and Pattipati, K. R., 1993, "A Practical Approach to Job Shop Scheduling Problems," IEEE Transactions on Robotics and Automation, Vol. 9, No. 1, February 1993, pp. 1-13.
83. Gunn, S.R., 1998, "Support Vector Machines for Classification and Regression," ISIS Technical Report, Princeton, NJ.
84. Hall, N.G., Lesaoana, M., and Potts, C.N., 2001, "Scheduling with Fixed Delivery Dates", Operations Research, 49, pp134-144.
85. Hax, A.C. and Meal, H.C. (1975) Hierarchical Integration of Production Planning and Scheduling, Studies in Management Sciences.
86. Han, J., and Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, CA.
87. Hendricks K.B., and Singhal V., "The effect of supply chain glitches on shareholder wealth", Journal of Operations Management, 2003.

88. Hillier, F.S., and Lieberman, G.J., Introduction to Operations Research, McGraw-Hill Science/Engineering/Math, 7 edition (March 22, 2002), ISBN: 0072535105
89. Hong, J. and V. V. Prabhu, 2003a, “Modeling and performance of distributed algorithm for scheduling dissimilar machines with setup,” to appear in International Journal of Production Research.
90. Hong, J. and Prabhu, V.V., 2003b, “Distributed Reinforcement Learning Control for Batch Sequencing and Sizing in Just-In-Time Manufacturing Systems,” to appear in Applied Intelligence.
91. Hong, J. and V.V. Prabhu, 2001, J. Hong and V.V. Prabhu, “Distributed Learning and Control for Manufacturing Systems Scheduling,” Engineering of Intelligent Systems (Eds. Monostori, L., Vancza, J., and Ali, M., (Eds.) LNAI 2070, Springer-Verlag, 2001, pp. 582-591.
92. Hong, J. Ph.D., 2001, “A Heterarchical Shop Floor Control Using Distributed Arrival Time Control and Reinforcement Learning,” Ph. D thesis, Pennsylvania State University, 2001.
93. Hong, J., V.V. Prabhu, and R.A. Wysk, 2001, “Real-Time Batch Sequencing using Arrival Time Control Algorithm,” accepted for publication, International Journal of Production Research.
94. Hotelling, H., 1947, “Multivariate Quality Control Illustrated by Testing of Sample Bombsights”, In O. Eisenhart (Ed.), Selected Techniques of Statistical Analysis, pp 113-184. New York, McGraw-Hill.
95. Hopp, W. J. and M. L. Spearman, 1996, “Factory Physics: Foundations of Manufacturing Managemen,”. IRWIN, Burr Ridge, IL.
96. Hopp, W. J. And Roof, M. L., 1998, Setting WIP levels with statistical throughput control (STC) in CONWIP production lines. International Journal of Production Research, 36(4), 867-882.
97. Jackson, J.E., 1991, “A Users Guide to Principal Components”, New York: Wiley-Interscience.
98. Jacobs, D. and Meerkov, S. M., 1995, “A system-theoretic property of serial production lines: improvability,” International Journal of Systems Science, 26(4), 755-785.
99. Johnson, D., (2003), “A framework for reducing manufacturing throughput Time”, Journal of Manufacturing Systems, Vol. 22, No. 4,pp. 283- 298.
100. Kelton, D.W., Sadowski, R., and Sadowski, D., (2001). “Simulation with ARENA”, McGraw-Hill Education; 2nd Ed edition.
101. Kapuscinski, R, and Tayur S.R., 1998a, “Optimal Policies and Simulatin Based Optimization for Capacitated Production Inventory Systems”, in Tayur, S., Ganeshan, R. and Magazine, M. (edt.), 1998, “Quantitative models for supply chain management”, Kluwer, Norwell, MA, U.S.A.
102. Kapuscinski, R., S. Tayur. 1998b, “A Capacitated Production-Inventory Model With Periodic Demand”. Oper. Res. 46(6) 899–911.
103. Kaskavelis, Christos A. and Michael C. Caramanis, 1998, "Efficient Lagrangian Relaxation Algorithms for Industry Size Job-Shop Scheduling Problems", IIE Transactions, v30, n11, p1085-1097, 1998
104. Kelton, D.W., Sadowski, R., and Sadowski, D., (2001). “Simulation with ARENA”, McGraw-Hill Education; 2nd Ed edition.

105. Kim, S. H., 1985, “Mathematical foundations of manufacturing science: theory and implications,” Ph.D. Thesis, MIT, May 1985.
106. Kimemia, J., & Gershwin, S. B. (1983). An algorithm for the computer control of a flexible manufacturing system. IIE Transactions, Vol. 15, No. 4, pp. 353–362.
107. Kogan, K. and E. Khmelnitsky (1995). "An optimal control method for aggregate production planning in large-scale manufacturing systems with capacity expansion and deterioration." Computers & Industrial Engineering 28(4): 851-859.
108. Koomsap, P., 2001, High Performance Computing for Intelligent Control of Laser Manufacturing Processes and Systems, Ph. D thesis, Pennsylvania State University, 2001.
109. Koomsap, P., V.V. Prabhu, N.I. Shaikh, J.T. Schriempf, and J. H. McDermott, 2003, “Condition Monitoring and Lifetime Estimation of a CO2 Laser,” to appear in Journal of Laser Applications.
110. Kopac, J. and Dolinsek, S. (1992) Cutting Parameters Optimization in Direct Connection Technological Data Base-N.C. Program, Factory 2000, Competitive Performance Through Advanced Technology, Third International Conference on (Conf. Publ. No.359)pp. 217-222.
111. Krasovski, N.N., 1963, “Stability of Motion - Application of Lyapunov’s Second Method to Differential Systems and Equations with Delays”, (Translated by J.L. Brenner) Stanford University Press.
112. Kroo, I, (1995) “Decomposition and Collaborative Optimization for Large Scale Aerospace Design.” ICASE/LaRC/SIAM Workshop on Multidisciplinary Optimization, Hampton, VA, March 1995.
113. Kroo, I.; Altus, S.; Braun, R.; Gage, P.; and Sobieski, I. (1994), “Multidisciplinary Optimization Methods for Aircraft Preliminary Design”, AIAA-94-4325, 5th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization. Panama City Beach, FL, Sept. 7–9, 1994. Vol. 1, pp. 697–707.
114. Larrabee, G.B., “The intelligent microelectronic factory of the future”, IEEE/SEMI Internaitonal Semiconductor Manufacturing Science Symposium, 1991, pp. 30-34.
115. Larsen, N. E. (1993) Methods for Integration of Process Planning and Production Planning. Int. J. Computer Integrated Manufacturing, 6, n1/ 2, 152-162.
116. Lawson, David G. And Evan L. Porteus, 2000, “Multistage Inventory Management With Expediting”, Operations Research, Vol. 48, No. 6, 2000, Pp. 878-893
117. Lee, H., P. Padmanabhan, S. Whang. 1997a. The bullwhip effect insupply chains. Sloan Management Rev. 38 93–102.
118. Lee, H., P. Padmanabhan, S. Whang. 1997b. Information distortion in a supply chain: The bullwhip effect. Management Sci. 43 546–58.
119. Lee, Hau L., Kut C. So, Christopher S. Tang, 2000, “The Value Of Information Sharing In A Two-Level Supply Chain,” Management Science, Vol. 46, No. 5, May 2000 pp. 626–643

120. Lee, Hau, And Seungjin Whang, 1999, “Decentralized Multi Ecelon Supply Chains:Incentives And Information” Management Science/Vol. 45, No. 5, May 1999, pp. 0025-1909
121. Lieberman, Marvin B. And Lieven Demeester, 1999, “Inventory Reduction And Productivity Growth: Linkages In The Japanese Automotive Industry,” Management Science, Vol.45, No. 4, April 1999
122. Lin, Grace, et al., 2000, “Extended-Enterprise Supply-Chain Management At IBM Personal Systems Group And Other Divisions,”Interfaces 30: January-February 2000, 7-25
123. Magretta, J. "The Power of Virtual Integration: An Interview with Dell Computer's Michael Dell," Harvard Business Review, March-April 1998, pp. 74-84.
124. Malakooti, B. (1989). A gradient-based approach for solving hierarchical multi-criteria production planning problems. Computers and Industrial Engineering, 16(3), 407–417.
125. Markowetz, F., 2002, “Support Vector Machines in Bioinformatics,” Diploma Thesis, Mathematics Department of Ruprecht-Karls, University Heidelberg.
126. Masin, M, and Prabhu, V.V., 2001, “AWIP: A Simulation-based Feedback Control Algorithm for Scalable Design of Self-regulating Production Control Systems”, IIE Transactions (submitted September 2001).
127. Masin, M., Herer, Y.T. and Dar-El, E. M., 2000, “SWIP: A Unified Model of Self-regulating Production Control Systems,” Working paper, Tel Aviv University, Tel Aviv, Israel.
128. Masin, M., Herer, Y.T. and Dar-El, E. M., 2000b, “Design of Self-regulating Production Control Systems by Tradeoffs Programming,” accepted, IIE Transactions.
129. Masin, M., Shaikh, N.I., and Wysk, R.A. (2003) Aggregative parameter variation in optimization modeling, IEEE Transactions on Robotics and Automation, Vol. 19, No. 4, pp. 529-542.
130. Mehra, A., Minis, I. and Proth, J.M. (1995) A Hierarchical Planning Model for Multistage Production Systems, Proceedings of INRIA / IEEE Symposium on Emerging Technologies and Factory Automation, Vol. 3, pp. 219 -228.
131. Meier, K. (1989) Commande Hierarchisee d’un Systeme de Production, Technical Report. These faculte des sciences de F universite de Metz.
132. Mitra, D. and I. Mitrani, 1990, “Analysis of a Kanban Discipline for Cell coordination in Production Lines, Part I,” Management science, 36(12), 1548-1568.
133. Nahimas, S., Production & Operations Analysis, McGraw-Hill/Irwin; 4 edition (November 10, 2000)
134. Nguyen, P. L., & Dupont, L. (1993). Production management of a steel manufacturing system: a hierarchical planning model. Computers and Industrial Engineering, Vol. 25(1–4), pp. 81–84.
135. NRC, 1998, “Visionary manufacturing challenges for 2020,” National Academy Press, 1998.
136. NRC, 2000, “Surviving Supply Chain Integration: Strategies for Small Manufacturing,” National Academy Press, 2000.

137. Ogata, K., 2001, Modern Control Engineering (4th Edition), Prentice Hall; 2001, ISBN: 0130609072
Page, E.S., 1954, Continuous Inspection Schemes, Biometrika 41, 100-115.
138. Pakin, S. and M. Lauria and A. Chien 1995, "High Performance Messaging on Workstations: Illinois Fast Messages (FM) for Myrinet," Proceedings of Supercomputing, December, 1995.
139. Pancucci, D. (1999) Shop floor to top floor integration, Manufacturing Computer Solutions, Vol5, No9.
140. Papadopoulos, H. T., Heavey, C. and Browne, J., 1993, "Queueing Theory in Manufacturing Systems Analysis and Design. Chapman & Hall," London, UK.
141. Parunak, H. V. D., "Applications of Distributed Artificial Intelligence in Industry", Foundations of Distributed Artificial Intelligence, O'Hare, G. M. P., and Jennings, N. R., (eds.), Wiley Inter-Science, 1994 .
142. Pearson, K., 1901, "On Lines and Planes of Closest Fit to Systems of Points in Space", Philosophihcal Magazine Series B, 2, 559-572.
143. Perkins, J. R. and Kumar, P.R., 1989, "Stable, Distributed, Real-Time Scheduling of Flexible Manufacturing / Assembly / Disassembly Systems," IEEE Transactions on Automatic Control, Vol. 34, No. 2, pp. 139-148, February 1989.
144. Phukan, A., M. Kalava, and V.V. Prabhu, 2003, "Complexity metrics for manufacturing control architectures based on software and information flow," conditionally accepted, Computers and Industrial Engineering
145. Pontil, M., Rifkin, R., and Evgeniou, T., 1999, "From Regression to Classification in Support Vector Machines," European Symposium on Artificial Neural Networks, Bruges, Belgium, 38:1, 9-13.
146. Prabhu, V.V. and N.A. Duffie, 1996, "Modeling and Analysis of Heterarchical Manufacturing Systems Using Discontinuous Differential Equations," Annals of the CIRP, Vol. 45, No. 1, 1996, pp. 445-448.
147. Prabhu, V.V. and N.A. Duffie, 1999, "Nonlinear Dynamics in Distributed Arrival Time Control of Heterarchical Manufacturing Systems," IEEE Transactions on Control Systems Technology, Vol. 7, No. 6, November 1999, pp. 724-730.
148. Prabhu, V.V., 1995, "Real-Time Distributed Arrival Time Control of Heterarchical Manufacturing Systems," Ph.D. Thesis, University of Wisconsin-Madison, 1995.
149. Prabhu, V.V., 1997, "Performance of Nonlinear Distributed Arrival Time Controllers in Part-Driven Heterarchical Manufacturing Systems," (invited paper) Proceedings of Sixth Industrial Engineering Research Conference, Miami, FL, pp. 310-315, 1997.
150. Prabhu, V.V., 1998, "Lyapunov Stability of Distributed Control in Multiple Machine Heterarchical Manufacturing Cells," Proceedings of XXVI North American Manufacturing Research Conference, Atlanta, GA, pp. 311-316, 1998.
151. Prabhu, V.V., 1999, "Performance of Real-Time Distributed Arrival Time Control in Heterarchical Manufacturing Systems," IIE Transactions, Vol. 32,pp. 323-331, 2000.
152. Prabhu, V.V., 2003, "Stability and Fault Adaptation in Distributed Control of Heterarchical Manufacturing Job Shops," IEEE Transactions on Robotics and Automation, Vol. 19, No. 1, pp. 142-147, February 2003.

153. Prabhu, V.V., S. Cho, and P. Koomsap, 1999, “High Performance Computing and Communication for Real-Time Feedback in Distributed Manufacturing Control Systems,” CIRP Journal of Manufacturing Systems, Vol. 29, No. 1, pp. 79-83, 1999.
154. Qiu, M. M., & Burch, E. E. (1997). Hierarchical production planning and scheduling in a multi-product, multi-machine environment. International Journal of Production Research, Vol. 35, No. 11, pp. 3023–3042.
155. Quintana, R., 1998, “A Recursive, Closed-Loop Production Methodology for Agile Manufacturing,” Proceedings of NSF Grantees Conference, 1998.
156. Ramesh, R., S. Y. Prasad and M. V. Thirumurthy, 1997, “Flow control in kanban based multicell manufacturing: I. A structural analysis of general blocking design space,” International Journal of Production Research, 35(8), 2327-2344.
157. Rao, Uday, Alan Scheller-Wolf, And Sridhar Tayur, 2000, “Development Of A Rapid-Response Supply Chain At Caterpillar,” Operations Research, Vol. 48, No. 2, March-April 2000, 189-204
158. Rosipal, R., Girolami, M., Trejo, L. J., and Cichocki, 2001, A., “Kernel PCA for Feature Extraction and De-Noising in Non-linear Regression,” Neural Computing and Applications, New York: Springer-Verlag Inc, 10:3, 231-243.
159. Saad, G. H. (1988). An optimization approach for the disaggregation of aggregate production plans. Proceedings of the Second International Conference on Expert Systems and the Leading Edge in Production Planning and Control (Charleston, SC, 1988), New York: North-Holland, pp. 305–318.
160. Sarmiento, A. And Nagi, R., 1999, “A Review Of Integrated Analysis Of Production-Distribution Systems”, IIE Transactions; Norcross; Nov 1999, Pp. 1061-1074
161. Schilling, G. and Pantelides, C. C., 1996, “A simple continuous-time process scheduling formulation and a novel solution algorithm,” Computers in Chemical Engineering, 272:1001-1004, 1996.
162. Schölkopf, B., Smola A.J., 2002, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA.
163. Schor, J. E., 1995, Efficient Algorithms For Buffer Allocation. M.S. Thesis, Mit Eecs; Mit Laboratory For Manufacturing And Productivity Report Lmp-95-006.
164. Schwartz, L.B., and Weng, Z.K., 2000, “The design of a JIT supply chain: the effect of leadtime uncertainty on safety stock,” Journal of Business Logistics, 21, 231-253
165. Seong, D., S.Y.Chang, and Y. Hong, 1994 , Heuristic Algorithms for Buffer Allocation in a Production Line with Unreliable Machines. International Journalof Production Research (accepted for publication).
166. Shaikh, N.I. and Prabhu, V.V., (2003), “Intelligent condition-based maintenance of reactive ion etching process”, to appear in, Smart Engineering System Design, ASME press, 2003.
167. Shapiro, J., 2000, “Modeling the Supply Chain”, Duxbury Press; 1 edition (December 20, 2000).
168. Shewhart, W.A., 1931, “Economic Control of Quality of Manufactured Product”, Princeton, NJ,: Van Nostrand.

169. Siha, S., 1994, “The pull production system: modeling and characteristics,” International Journal of Production Research, 32(4), 933-949.
170. Simpson, N. C. (1999). Multiple level production planning in rolling horizon assembly environments. European Journal of Operational Research, Vol 114, No. 1, pp. 15–28.
171. Simpson, N. C., & Erenguc, S. S. (1998). Improved heuristic methods for multiple stage production planning. Computers and Operations Research, Vol 25(7/8), pp. 611–623.
172. Smola, A. J., and Schölkopf, B., 1998, “A Tutorial on Support Vector Regression,”euroCOLT2 Technical Report Series, NC2-TR-1998-030, Berlin, Germany.
173. Smola, A. J., Schölkopf, B., and Müller, K. R., 1998, “The Connection Between regularization Operators and Support Vector Kernels,” Neural Networks, 11:637-649.
174. Sobieszczanski-Sobieski, J. and Haftka, R. T., (1997), "Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments," Structural Optimization, Vol. 14, No. 1, pp. 1-23.
175. Son, Y. J. and Wysk, R. A. (2001) Automatic Simulation Model Generation for Simulation-based, Real-time Shop Floor Control, Computers in Industry, 45 (3), 291 - 308.
176. Spearman, M. L., 1992, “Customer service in pull production systems,” Operation Research, 40(5), 948-958.
177. Steele, J. W., Son, Y. J. and Wysk, R. A. (2001) Resource modeling for the integration of the manufacturing resource, Journal of Manufacturing Systems, Vol. 19, No. 6.
178. Stephanopoulos, G., 1983, “Chemical Process Control: An Introduction to Theory and Practice”, Prentice Hall PTR; 1 edition (October 1, 1983), ISBN: 0131286293.
179. Stradtler, H., and Kilger, C. (edt.), 2000, “Supply Chain Management and Advanced Planning: Concepts, Models, Software and Case Studies”.
180. Suri, R. and Fu, B-R, 1994, “On Using Continuous Flow Lines to Model Discrete Production Lines,” Discrete Event Dynamic Systems: Theory and Applications, v 4, n 2, pp. 129-169, 1994.
181. Taha, H.A., “Operations Research: An Introduction (7th Edition)”, 2002, Prentice Hall, ISBN: 0130323748
182. Takeuchi, K., Collier, N., 2002, “Use of Support Vector Machines in extended named entity,” CoNLL-2002, Taipei, Taiwan.
183. Tayur, S. R., 1992, “Properties of Serial Kanban Systems,” Queueing Systems, 12, 297-318.
184. Tayur, S. R., 1993, “Structural Properties and Heuristic for Kanban-controlled Serial Lines,” Management Science, 39(11), 1347-1368.
185. Tayur, S., Ganeshan, R. and Magazine, M. (edt.), 1998, “Quantitative models for supply chain management”, Kluwer, Norwell, MA, U.S.A.
186. Triantaphyllou, E., Kovalerchuk, B., Mann, L., and Knapp, G.M., 1997, “Determining the most important criteria in maintenance decision making. Journal of Quality in Maintenance Engineering vol 3, pp.16–24.

187. Thirumalai, R., “Multi Criteria – Multi Decision Maker Inventory Problems for Serial Supply Chains”, Ph. D thesis, Pennsylvania State University, 2001.
188. Vapnik, V., 1995, The Nature of Statistical Learning Theory, Springer, N.Y.
189. Veeramani, D., Joshi, P., and Sharma, P., “Critical Research Issues in Agent-Based Manufacturing Supply Webs, ” Workshop of the 3rd International Conference on Autonomous Agents (Agent’99) – Agents for Electronic Commerce and Managing the Internet-Enabled Supply Chain, Seattle, Washington, May 1999.
190. Venkatasubramanian, V., Rengaswamy, R., Yin, K., and Kavuri, S.N., “A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-based Methods”, Computers and Chemical Engineering, 27 (2003) 293-311.
191. Villa, A. (1989). Hierarchical architectures for production planning and control. Operations Research Models in Flexible Manufacturing Systems, Wien: Springer, pp. 261–288.
192. Warburton, R. D. H. (2004). "An exact analytical solution to the production inventory control problem." International Journal of Production Economics In Press, Corrected Proof
193. Wiendhal, H. P. and Scholtissek, P., 1994, “Management and control of complexity in manufacturing,” Annals of the CIRP, Vol. 43, No. 2, 1994, pp. 533-540, 1994.
194. Wold, H., 1982, “Soft Modeling, The basic design and some extensions” Systems under indirect observations, Amstradam: North Holland.
195. Wold, S., 1978, “Cross-validatory estimation of the number of components in factor and principle component models”, Technometrics, 20 (4), 397-405.
196. Wold, S., Albano, C., Dunn III, Edland, W.J., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., and Sjostrom, M., 1984, “Multivariate Data Analysis in Chemistry”, Chemometrics; Mathematics and Statistics in Chemistry.
197. Wold, S., Esbensen, K., and Geladi, P., 1987, “Principal Component Analysis” Chemometrics and Intelligent Laboratory Systems 2, pp 37-52.
198. Wooldridge, M.J., and Jennings, N.R., "Agent Theories, Architectures and Languages: A Survey," in Intelligent Agents: ECAI-94 Workshop on Agent Thoeries, Architectures, and Languages, eds. Wooldridge, M.J., and Jennings, N.R., pp. 1-39, Springer-Verlag, Berlin, 1994.
199. Wu, S.D. and Harker, P.T., 1998, “Toward a Distributed Paradigm for Manufacturing Logistics,” Proceedings of NSF Grantees Conference, 1998.
200. Wysk, R.A. (1977) Process Planning systems, MAPEC Module, West Lafayette, IN: Purdueue University; School of Industrial Engineering.
201. Yan, H. S., Wang, N. S., Zhang, J. G., & Cui, X. Y. (1998). Modelling, scheduling and simulation of flexible manufacturing systems using extended stochastic high-level evaluation Petri nets. Robotics and Computer-Integrated Manufacturing, 14(2), 121–140.
202. Yeh, H. M., Tarng, M. Y., & Chen, W. C. (1988). An iterative algorithm to solve hierarchical production planning problems. The Collection of Refereed Papers Presented at the 9th International Conference on Production Research (Cincinnati, Ohio, 1987), Amsterdam: Elsevier, pp. 209–215.

Vita

NAZRUL I. SHAIKH
310 Leonhard Building
University Park PA-16802
nazrul@psu.edu

Education:

Dual PhD in Industrial Engineering and Operations Research — Aug 2002-Dec 2006
PhD minor in Statistics
The Pennsylvania State University, University Park.

Dissertation: Intelligent Sense-Response System for Supply Chain Event Management
Advisor: Dr. Vittal V. Prabhu

Aug2000–Aug 2002
Dual M.S. in Industrial Engineering and Operations Research
The Pennsylvania State University, University Park.
Thesis: Mathematical Modeling and Control of Cryogenic Tunnel Freezers

Aug 1996-May 2000
B.Tech in Chemical Engineering
Indian Institute of Technology (IIT), Bombay, India.
Thesis: Optimization of the Pooling and the Trans-shipment Problems in a Refinery Supply Chain

Representative Publications:

1. Shaikh, N.I. and Prabhu, V.V., (2005), "Simulation-control based resource allocation on conveyor networks with merging configurations", submitted to International Journal of Production Research.
2. Koomsap, P., Shaikh, N.I., and Prabhu, V.V., (2005), "Integrated process control and condition based maintenance scheduler for distributed manufacturing control systems," International Journal of Production Research, Volume 43, Number 8, pp 1625 - 1641.
3. Shaikh, N.I., Masin, M., and Wysk, R.A., (2004), "Shop floor planning using a parameter variation modeling approach", IIE Transactions (under review).
4. Shaikh, N.I. and Prabhu, V.V., (2004), "Integrated condition-based maintenance and process control of reactive ion etching process", International Journal of General Systems, (paper code 2004-009).
5. Masin, M., Shaikh, N.I., and Wysk, R.A., (2003), "A parameter variation modeling approach for enterprise optimization", IEEE Transactions on Automation and Robotics, V-19, No 4, pp 529-542.
6. Shaikh, N.I. and Prabhu, V.V., (2003), "Intelligent condition-based maintenance of reactive ion etching process", Intelligent Engineering Systems through Artificial Neural Networks: Smart Engineering.