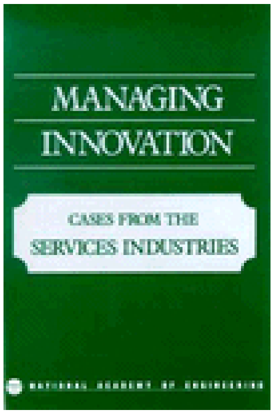

Managing Innovation: Cases from the Services Industries

Bruce R. Guile and James Brian Quinn, Editors;
National Academy of Engineering

ISBN: 0-309-59480-4, 224 pages, 6 x 9, (1988)

Series on Technology and Social Priorities

NATIONAL ACADEMY OF ENGINEERING

Managing Innovation

Cases from the Services Industries

Bruce R. Guile and James Brian Quinn Editors

NATIONAL ACADEMY PRESS
Washington, D.C. 1988

National Academy Press 2101 Constitution Avenue, NW Washington, DC 20418

NOTICE: The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievement of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

Funds for the National Academy of Engineering's Symposium Series on Technology and Social Priorities were provided by the Andrew W. Mellon Foundation, Carnegie Corporation of New York, and the Academy's Technology Agenda Program. This publication has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee. The views expressed in this volume are those of the authors and are not presented as the views of the Mellon Foundation, Carnegie Corporation, or the National Academy of Engineering.

Library of Congress Cataloging-in-Publication Data

Managing innovation: cases from the services industries/ Bruce R. Guile and James Brian Quinn, editors.

p. cm.—(Series on technology and social priorities)
Bibliography: p.
Includes index.

ISBN 0-309-03926-6: $32.50.
ISBN 0-309-03891-X (pbk.): $22.50

1. Service industries—Technological innovations— Management. I. Guile, Bruce R. II. Quinn, James Brian, 1928– . III. Series.
HD9980.5.M343 1988
658.5/14—dc19 88-19631

CIP

Printed in the United States of America

Preface

Services now account for more than two-thirds of the U.S. gross national product and close to three-quarters of U.S. employment. This is not a new phenomenon, but important questions remain unanswered. How sophisticated are services industries in their use of technology? Are the productivity improvement issues in services fundamentally different from those in manufacturing businesses? Is the competitive performance of U.S. services industries threatened by underinvestment in technology or by poor management of innovation? What is the effect of global economic integration—brought about in part by advances in services such as transportation and communications—on the structure and performance of U.S. services industries? These questions, and other similar questions, are addressed in this book and its companion volume *Technology in Services: Policies for Growth, Trade, and Employment*.

The two books—one focused on the management of technology in services and the other on public policy issues—set forth an important message: services such as banking, software development, communications, air transportation, and health care are technologically dynamic and crucial to the performance of both U.S. manufacturing and the entire economy.

Whether the task is linking world financial markets through computer and communications networks, lowering the cost of delivering volatile materials or urgent packages, or designing dams and bridges for U.S. or foreign governments, the challenges to technology and management in services industries are significant. Productivity in such industries suffers from many of the same problems identified for manufacturing industries. As the papers in this volume demonstrate, capital availability, human resource development, quality, and scale economies are as important in services as they are in manufacturing. R&D or innovation in services industries can create a competitive edge just as they do in manufacturing. In short, if reasons exist for distinguishing between the importance and character of manufacturing and services, they do not arise from the role of technology in the industry.

The principal focus of this volume is on examples of the application of technology in services businesses, either at the level of an individual business or at the level of an industry. Much of the material in this book was presented at an NAE symposium entitled "Technology in Services: The Next Economy" held in Washington, D.C., on January 28 and 29, 1988.

I would like to thank James Brian Quinn, who chaired the NAE activity on technology in services, and Bruce R. Guile, the principal staff officer for the project, for their efforts. Together they assembled a first-rate advisory committee, worked with the advisory committee to organize an excellent workshop and subsequent symposium, and have moved quickly to get this material published.

Also, on behalf of the Academy, I would like to thank the advisory committee (listed on page 195) for the project and the authors who participated in the workshop and symposium. Special thanks are due to Jesse H. Ausubel, director of the NAE Program Office, Stephen L. Murphy, an NAE summer fellow who helped in the early stages of the project, H. Dale Langford, NAE editor, and to Marjorie D. Pomeroy, administrative assistant in the NAE Program Office who, although she has moved on to other employment, was involved in the project for almost 18 months.

ROBERT M. WHITE

PRESIDENT

NATIONAL ACADEMY OF ENGINEERING

Contents

Managing Innovation

Managing Innovation in Services

James Brian Quinn and Bruce R. Guile

> Short of cash for a movie and dinner out a New Yorker stops at an automatic teller machine to make a cash withdrawal on Saturday evening before going on to the theater.
>
> At 11:00 p.m. a California resident calls the catalog department of a store in Maine and orders, for express delivery to his house two days later, an item at a total price not much different from the price he would pay at a local department store.
>
> An automobile mechanic stands at a computer terminal in the service bay, checks a data base catalog concerning the car currently on the lift, confirms that the local wholesaler has the necessary replacement parts, and orders the parts all without leaving the service bay.
>
> An executive from Spain travels for an annual checkup to a medical facility in Cleveland, Ohio—the city's largest private employer, drawing 6 percent of its customers and 10 percent of its revenues from foreign patients.
>
> The trading day closes on an average day at the New York Stock Exchange—250 million shares changed hands. The reconciliation process—documenting trades, accounting for securities documents, verifying values the exchange process has set—continues into the evening and by morning of the next day the process begins all over again.

Services today account for 70 to 75 percent of all economic activity in the United States, and large services industries such as communications, transportation, health care, financial, and professional services have become capital intensive and technology based. As the chapters in this volume dem onstrate, innovation is crucial to success in the services industries and is in many ways similar to innovation in goods industries.

Like the floor of a manufacturing plant, the teller floors of a bank may be considered, at the most basic level, production centers amenable to process and productivity improvements through careful application of new technologies. And—just like a new household appliance, machine tool, or aircraft—a new commercial data base, kidney dialysis procedure, or communication service can be an extremely advanced product technologically. All these products—goods and services—pass through similar processes of invention, development, design, packaging, testing for customer acceptance, justification in the user's environment, distribution, diffusion, upgrading, and post-sale servicing.

The chapters in this volume illustrate that many of the principles that have proved so essential to innovation management in product or manufacturing environments are equally important in services. Among these are (1) developing a clear vision of the technological need in an atmosphere that tolerates the false starts, failures, and experimentation necessary to develop new technologies; (2) maintaining a close interactive relationship between developers and users of innovation; (3) designing in an integrated fashion for both productivity efficiencies and quality of output; (4) utilizing multiple approaches (or parallel development) to improve the quality of technological solutions, team motivation, and the probability of ultimate success; (5) keeping innovative organizations small scale, flat, and open to new ideas; (6) developing a high degree of internal technical expertise to participate in and evaluate potential solutions; (7) creating strong external linkages to both domestic and foreign sources of technological expertise such as suppliers and universities; (8) empowering champions or championing teams to drive the innovation process; (9) maintaining the continuity, balance, and enthusiasm of the innovating team; (10) providing the patient, long-term (3 to 5 year) support that technological innovations require; and (11) not over-managing technological innovations with detailed, rigid, progress plans, but being willing to move incrementally and opportunistically as problems and opportunities emerge.

Increasingly, managers in services industries must understand the innovation process and learn to manage technological development with a balance of insight, analysis, and instinct. The chapters provide some useful examples of how successful services enterprises such as AT&T, Federal Express, Citicorp, Bell & Howell, and the New York Stock Exchange have proceeded.

NEW PRODUCTS AND PROCESSES

The chapter by Carl Nehls describes the development and implementation of Federal Express's COSMOS IIB system that allows tracking of packages from pick-up through delivery. The system, developed over 7 years, has elements of both product innovation and process innovation. From a product innovation perspective, COSMOS IIB allows Federal Express to offer customers real-time, precise information about the location of a package anytime between pick-up and delivery. As a production process innovation, electronic custodial tracking substantially increases Federal Express's ability to control and transport packages efficiently. In addition, COSMOS IIB's systematic collection and electronic encoding of data also allow virtually real-time efficiency analyses of Federal Express's entire transport operation. In short, product and process innovations are often linked in services. An appreciation of both aspects is crucial for effective management of innovation and the financial evaluation of an innovation's outcome.

The chapter by Paul Glaser raises a similar set of issues. In the case of Citicorp's development of automated teller machines (ATMs), the product is the teller—client transaction-the exchange of money or financial documents relating to personal or business finance. ATMs provided a process innovation that changes the nature of the company's product delivery system—making the product available at more hours and locations with more predictable waiting times for users than a human teller system allowed. Particularly noteworthy in organizational terms was Citicorp's decision during the 1970s to establish a wholly owned subsidiary to develop the technology. As part of the pervasive mythology about services, one does not normally think of banks as having a substantial internal technological capability. But in Citicorp's case the development of this capability was obviously a key element in its innovative success.

In their chapter Frederick Fellowes and Donald Frey describe a classic challenge in business technology—keeping (or expanding) a market in light of changing demands and product technologies. Threatened by other firms in the microfiche publishing business and by new entrants offering catalog data bases, Bell & Howell reevaluated its entire approach to the automotive parts catalog publishing business. In the process the company discovered how a properly conceived system could provide significant efficiency improvements in automobile repair operations. Much of the hardware needed for the new system was "off the shelf." In this case the heart of the story is not the development of a new widget, but the company's persistent and iterative efforts to understand its customers' activities and to offer a system that could be tailored to, and justified in, the customer's business environment. As in many cases of technological development, initial expectations about what the new technology could offer were incorrect. It took both patience and serendipity to discover exactly what kind of system would add substantial value to automobile repair operations and thus justify the customers' purchase of Bell & Howell's parts catalog in electronic form rather than on microfiche or paper.

Express package delivery, retail banking, and automotive repair involve relatively concrete production processes. At one level, the same is true for a stock exchange. Connecting buyers and sellers, verifying and documenting trades, and handling or storing securities are processes that can be characterized as production activities centered on transactions. Christopher Keith and Allan Grody describe the incremental process of innovation at the New York Stock Exchange—spanning more than 100 years from the Morse code ticker to modem program trading. Process innovations often occurred in response to pressures for more Exchange capacity, which when achieved generally led to new "product" or services innovations, which in turn created dramatic increases in demand. Keith and Grody note, however, that an exchange—if its market processes are operating properly—creates another important product that is not directly related to the production aspects of executing transactions; an exchange is a mechanism for determining value.

The NYSE, in particular, occupies a central position in determining stock values, and through the Dow-Jones averages provides an important measure of the health of the U.S. economy and its relationship to other global financial systems. The importance of worldwide knowledge about, and confidence in, the quality and reliability of the services of the NYSE forces technological innovation at the Exchange to be a paradigm of innovation in a fish bowl. Mistakes (almost inevitable in any development process) can be very costly and visible; hence, new processes are carefully scrutinized by a variety of interested parties, both on the Exchange and at more distant vantage points. Based on their confidence in the NYSE's system, parties will participate or take their business elsewhere. Witness the attention paid to some of the unexpected impacts of program trading that showed up in the market crash of October 1987 and the reluctance of individuals to return to the market later. The degree of risk—and the public's interest in its systems—forces a degree of conservatism and incrementalism in innovation at the NYSE that might seem overly cautious in other environments but is probably essential given the character of the Exchange.

The chapter by Richard Larson takes a different view of innovations affecting service production processes. Larson focuses on a technology—the technology of operations research—and examines its application to three functional areas in a variety of service businesses. The chapter provides examples of operations research applications in logistics, work force planning, and services related to goods production. Larson examines these problems both as technical and investment decisions, providing information about returns on investment as well as the complexities of implementing solutions in a services environment. Operations research (OR) is an example of a "soft" technology of great importance, the widespread application of which is intimately coupled with the diffusion of the computing capability necessary for many applications. Larson's review of successful OR applications—some yielding very impressive returns on investment—is very provocative and suggests the range and potential payoffs of unexploited opportunities for this technology in virtually every service environment.

STRATEGY, STRUCTURE, AND TECHNOLOGY

The five chapters discussed above deal mostly with specific technological innovations. The remaining chapters deal more with the development and organization of entire services industries—and their relationship to other policy parameters—rather than with the application of a particular technology in one organization. The chapter by John Davis details the emergence of the cellular telephone industry in the United States and illustrates the long time horizons (15 years) and huge investments (hundreds of millions of dollars) necessary for innovation in the large services industries. In the cellular telephone industry the government's regulation of telephone service providers and its allocation of the available frequency spectrum dramatically shaped the development of this industry in the United States and the world. But the issues raised by Davis's analysis are not unique to cellular telephone service. Other services industries are among the most regulated industries worldwide. The financial, transportation, and communications industries in particular are subject to a type and amount of regulation not usually applied to goods industries. Although there are often good reasons for regulation in these industries, the fact that another layer of complexity is added to decision processes affecting innovation often slows and changes technological introductions, with profound competitive effects. In these industries, innovation is not likely to be helped by abolishing regulation—everyone would lose if the frequency spectrum fell into chaos for lack of orderly allocation or interconnect standards. But, as the Davis chapter illustrates, there may be substantial opportunities for improving technical innovation by enhancing the responsiveness of regulatory processes.

Juan Murillo's review of the evolution of bridge construction and design in Europe and the United States bears some similarities to the development of cellular telephone service. In the case of bridges the government affects innovation primarily through its standard setting and procurement processes. Virtually all bridges are bought by governments and, as public works, these must meet strict cost, safety, and durability criteria as defined by public policy processes. Procurement practices and liability concerns have combined with outmoded standard setting practices to erode seriously providers' ability to innovate. In particular, the requirement that design and construction be provided by separate firms has had a dampening effect on both design and construction innovation. Technological advance is important to this industry, and the most important opportunities for im proving the technology used in U.S. bridge design and construction may lie in changing government procurement practices in ways that encourage more experimentation, backed by professional testing facilities that are credible to all parties. Murillo also notes the high costs that are incurred for the industry and the nation as a result of failure to deal with highly litigious participants.

The chapter by Thomas Doorley, Alison Gregg, and Christopher Gagnon examines the implications of information technology for a number of professional services firms. The chapter provides data not available elsewhere on professional services firms and shows how professional firms' strategies, market positionings, and organization practices can affect (and be affected by) new technology developments. Particularly interesting is the observation that there are few economies of scale for firms in professional services, thus suggesting why applications of computer and information technologies do not show expected productivity returns in professional services. Much of the benefit from investigating technology in professional services industries shows up as quality improvements or increased services flexibilities, not measured in national productivity statistics. The chapter moves beyond that argument, however, to suggest how the activities of professional services firms can be segmented by functions so that one can analyze where "value added" or "productivity improvement" potentials can be most attractive for the application of new information technologies. The chapter concludes with an analysis of the barriers to effective implementation of information technologies in professional services firms, particularly noting how the partnership structures, which most professional services firms use, exacerbate problems in implementing technological change.

The introductory chapter by James Brian Quinn summarizes some measures of the importance of technology in the services industries broadly and then focuses on the particular interactions between services and manufacturing. It notes that some 75 percent of most manufacturers' internal costs and external purchases are for services activities. By using the new services technologies available, manufacturers can significantly decrease their costs and enhance their own value added. Quinn posits that the successful manufacturers of the future will be those who most effectively integrate flexible manufacturing systems and services technologies. The chapter suggests how technological advances in services have already forced major changes in the entire world economic order and, more particularly, in the structure of U.S. industry. It poses that integrated use of services and manufacturing technologies may already be leading to the "remanufacturing of the United States." The chapter demonstrates a variety of ways in which services technologies have created specific opportunities for increased competitiveness both in the services industries and at the interface with manufacturing. It presents a structure for looking at these relationships in new ways, a series of concrete examples of significant applications, and macroeconomic data about the benefits this interrelationship has offered.

SUMMARY AND CONCLUSIONS

Virtually every technological advance of the modern age is applied in services industries. Advances in communications and computing, medical care and delivery, transportation systems, energy production, and pollution control are all applied in or central to the development of services industries. The chapters in this volume illustrate several important lessons about managing innovation in services.

At the level of the production process, the chapters by Nehls, Glaser, Keith and Grody, and Fellowes and Frey illustrate that many innovations in services involve an intimate combination of both product and process technologies. These cases (especially those by Keith and Grody, Glaser, and Nehls) dispel any remaining myths about services and services-oriented businesses not needing their own in-house technical personnel, and in some cases, R&D capability. The core competencies of successful services firms will increasingly have to include a substantive ability to invent and apply technologies. No small part of this competency, of course, will be the ability to work effectively with highly sophisticated technological suppliers and with both technologically sophisticated and unsophisticated customers. Many services firms now perform the role of systems integrators—designing, implementing, and operating sophisticated information, transportation, distribution, or financial systems that integrate the activities of geographically dispersed and highly diverse customers and suppliers. Effective competition, and the company's ability to add value in this environment, will increasingly depend on the engineering and technological management the company can bring to bear throughout its system.

At the industry level, the chapters by Murillo, Keith and Grody, Davis, Doorley et al., and Quinn help a reader understand how changing business structures often pose major new strategic threats and opportunities for executives and regulators. From these chapters one can conclude that successful exploitation of the new services technologies will require development of an array of new organizational forms and relationships—including new coalition modes, government-industry consortia, cross-industry competition, disintermediation possibilities, and moves to transactional and functional (versus industry) regulation modes. All this will create exciting new challenges for managers in services, manufacturing, and the professions over the next decades.

This volume demonstrates in detail that services are a technologically dynamic portion of the U.S. economy and implicitly argues that a new set of assumptions should exist in policy debates: that significant opportunities exist for improving U.S. economic performance through intelligent and systematic application of advanced technologies in the services industries and at the services-manufacturing interface. A companion volume to this book—*Technology in Services: Policies for Growth, Trade, and Employment*—focuses on the public policy questions that affect services industry growth and performance.

Services Technology and Manufacturing: Cornerstones of the U.S. Economy

James Brian Quinn

Now, more than ever, the United States needs a vital services sector. Some U.S. manufacturing industries (such as basic steel, automobiles, or electronic appliances) have suffered serious competitive reverses in recent years. Merchandise trade deficits have risen to all time highs. And even services trade balances have begun to slip seriously—from a positive balance of $37 billion in 1982 to only $15 billion for the entire year 1987. Of the total, the selected business services category declined by more than $8.6 billion between 1982 and 1987, when the net balance was a mere $0.8 billion.[1] The most serious losses occurred in travel and transportation-related industries. Although U.S. airlines have maintained a steady 38 to 41 percent of the world's revenue passenger miles during the past decade (U.S. passengers made up nearly 46 percent of the world total), the once powerful international carriers PanAm and TWA have fared poorly at the hands of direct foreign competitors such as JAL, Swissair, and Singapore Airlines, which made heavy long-term investments in their fleets and paid close attention to the quality of care given to passengers.

MAJOR THREATS AND OPPORTUNITIES FOR THE 1990S

The threat to other services industries is real and immediate. Confrontations will not be limited to markets abroad; U.S. markets for services are no safer from foreign competition than were the domestic markets for manufactured goods. Indeed, foreign direct investment in the U.S. services sector has exploded since the mid-1970s, although the rate of increase was slowing until the mid-1980s, when the weak dollar began to attract further foreign investments. It is sobering to realize that many of the great names in services—such as 20th Century Fox, Stouffers Hotels and Restaurants, Intercontinental Hotels, Saks Fifth Avenue, Marshall Fields, Spiegel, A&P, Grand Union, and Giant Food—have foreign owners now. Even the great investment banks of the United States have strong new partial owners from abroad.

The United States needs to take a careful look at its services sector and place increasing emphasis on its competitiveness and value-added potentials. U.S.-based services will continue to offer dramatic opportunities for growth, productivity improvement, and innovation over the next decade. Few realize how completely technological innovations in services have already restructured the U.S. economy and its international competitive relationships—in the process antiquating many of our precepts about the roles of both manufacturing and services in a modern economy.

The companion volume to this book—*Technology in Services: Policies for Growth, Trade, and Employment*—contains a number of research and policy papers concerning the implications of this restructuring. In this volume we have collected some outstanding cases about services innovations. These can provide informative guidelines on how such innovations can be managed, or mismanaged if we are not careful.

NEW TERMS OF COMPETITION IN SERVICES

Although the effects of deregulation complicate the picture, it is new technologies that have most extensively altered and expanded the services industries in recent years. Unfortunately, such technologies have also made these industries vulnerable to the same modes of attack that earlier so rapidly undermined segments of the nation's manufacturing economy. If executives and policymakers understand these emerging forces, they will both exploit some potent opportunities to boost their companies' (and the nation's) performance and develop more effective responses to their foreign competition. The attempts and frustrations of Federal Express in opening the Japanese market provide a good example. If executives do not comprehend these potentials or if they act with the complacency and shortsightedness that characterized policies that affected manufacturing industries a decade ago, they could leave major portions of the U.S. economy wide open to further foreign incursions. What are the main patterns at work?

Concentration and Diffusion

In virtually every services industry, technological change has created vastly increased capacities and economies of scale. The New York Stock Exchange and Federal Express COSMOS II systems described in this volume provide cases in point. The first-order effect in most industries has been a revised competitive structure characterized both by increased concentration in the industry and by increased fragmentation and market segmentation. To exploit the new scale economies available, many intermediate and large companies in each services industry merged to form giant enterprises. But within each services industry smaller companies also identified local niches—or specialized services needs—and concentrated successfully on these, with investment returns following the classic V- or U-shaped distributions noted by Porter (1985) and others (Booz Allen & Hamilton, undated).

In the rental car industry, for example, the major companies in conjunction with the airlines initially locked up virtually all long-distance travelers with their instantaneous guaranteed reservation systems. They rapidly extended their scale through international subsidiaries and affiliates connected electronically. But soon, niche operators proliferated to exploit the rigidities of the giants' more standardized services. Automated computer and telephone-answering devices allowed both "Super-Elegant" and "Rent-A-Wreck" extremes to serve local markets with office-in-the-home operations. Only a few midsized operations such as Alamo or Agency could grow substantially by segmenting target groups with particular needs, such as cut-rate or conference rentals.

As the Keith and Grody paper notes, automation in the securities handling process has changed that industry's entire structure since the mid-1960s. Under the old paper-based system, brokers had to document all trades by hand, and shares had to be physically delivered from the seller's agent to the buyer's. As daily volumes approached 12 million shares, only the big firms could hire and manage enough people to keep up with their securities trades each day. Smaller firms began to fail because they could neither control nor process their securities in a timely fashion. Finally, Wall Street firms formed Central Certificate Services (later the Depository Trust Company), which brought essentially all securities certificates under one roof, where a single set of accounting entries could change their ownership. After 5 or 6 years, the system became totally electronic and smaller brokers could tie into the Depository. Today, such automated clearinghouses handle virtually all private and government transactions and are key linkages enabling the worldwide integration of financial markets.

New technology also produced important scale effects in medical care. Some high-cost technologies, such as computerized axial tomography (CAT), allowed much more accurate diagnoses, and other advanced technological systems opened the way for heart, brain, general surgical, and life support procedures never before feasible. In part because of the reimbursement system then in use, the initial effect of these capital-intensive technologies was to centralize treatment in the larger hospitals. Small practitioners and hospitals had neither the patients nor the resources to buy or constantly update the new diagnostic, surgical, and recovery equipment. This centralization led to highly specialized medical training in the medical centers, caused specialist practices to form around the biggest hospitals, and spurred the creation of large regional referral facilities to handle particularly difficult cases. Many smaller hospitals suffered, closed down, or joined in cooperative networks with the larger centers.

Overhead expenses grew and the average cost of an inpatient stay soared from $729 in 1972 to $2,898 in 1984, when diagnosis-related group restrictions devised by the government medical reimbursement programs began to shift cost patterns. With patient care becoming ever more impersonal in the large centers, new and less costly distribution systems began to emerge along both vertical (home care, primary care, specialty care) and horizontal (pediatrics, obstetrics, dermatology, internal medicine) axes. Such complex alternative systems as electronically linked home or outpatient care facilities, emergency cardiac systems, and health maintenance organization networks have grown almost 10 times in number since 1971, and the number of people they serve grew from 3.1 million to more than 13.6 million. Ambulatory surgical centers have exploded from 400 in 1982 to more than 1,200 in 1985. Soon management of these complex systems became so critical a factor that large private companies found it profitable to apply their skills to both hospitals and other parts of the system—thus starting another wave of consolidation.

Networks and Variety in Services

Beyond their important effects on scale, new services technologies often create powerful second-order effects—economies of scope—that allow entirely new service products to move through established networks or systems with little added cost. Once debugged, communications and information-handling technologies permit the distribution of a much wider set of services to a more diverse and dispersed customer base. In the process they often further decrease costs on old product lines as equipment, development, and software investments are allocated over the broader base of applications. In addition, new technologies frequently offer wide-ranging strategic benefits in terms of more rapid new-product introductions or faster response to competitors' moves.

An often cited but classic example is American Airlines' Sabre System, a computerized reservation system which when combined with sophisticated revenue management software allowed American to offer selected fares comparable with those of People Express, an upstart no-frills competitor, without jeopardizing its higher fare sales base. More recently, Sabre has expanded its services to allow customer companies to link directly into its system, make their own reservations, and eventually handle payment of their travel bills. Sabre, which is reportedly often more profitable than the air travel operation, is extending its services into telemarketing of products.

In this same vein, insurance companies first began automating their back-office activities for greater efficiency in the mid-1960s when their industry was stable and heavily regulated. As expected, better handling of premium billings and collections brought dramatic productivity gains. Then when wildly fluctuating interest rates hit the industry in the 1970s the companies had to alter their products rapidly to attract new premiums and to offset the effects of customers' borrowing against their policies at lower interest rates. In such an environment, only companies with flexibly designed back-office computer and control systems could design or deploy their products quickly enoug—within days or weeks—to get a competitive edge. Previously, insurance companies had brought out new rate books only once every 3 to 5 years. Surviving companies now note that, without effective electronic and software systems, they could neither have conceived of the variety of new products needed during this critical period nor could they have explained and introduced the products to their widespread agent and customer bases. Again, many smaller insurance companies, which could not afford the huge costs of sophisticated electronic networks, faltered, sold out, merged, or concentrated on localized or specialized services outside the interests of larger companies.

International Implications

Given both their economies of scale and scope, as well as the relatively low expense of transporting their services internationally, large technologically sophisticated U.S. services companies such as Citicorp or the deregulated AT&T system (see the Glaser and Davis chapters in this volume) should enjoy powerful international advantages. But the same is also true for selected foreign competitors. The cellular telephone case included here also shows how rapidly overseas competitors can seize and exploit an advanced U.S. innovation, when regulatory processes react too slowly.

In still another service arena, powerful European and Asian distribution companies are also proving to have the scale, talent, and management systems to acquire and upgrade poorly performing U.S. retail networks. For example, Britain's well-automated food retailers boast a hefty 23.4 percent return on capital employed. J. Sainsbury's (which is electronically automated from checkout scanning through stock control and employee scheduling) with four other chain stores accounts for more than half of Britain's grocery sales (*The Economist*, 1987b). Now, Sainsbury's (through its acquisition of Shaw's Supermarkets) and other big British retailers are pushing abroad both in the United States and the European Economic Community (EEC) (*The Economist*, 1987a). Increasingly, technology is making global competition among services giants both possible and essential to success. Few areas seem immune.

Segmentation and Responsiveness

As the chapter by Doorley et al. suggests, technology in professional services—although not enhancing "measured productivity" much—is enabling the management of greater levels of complexity, with higher output quality. Law firms can in a matter of hours complete more exhaustive back-ground searches, prepare more intricate contracts, and document resulting settlements more thoroughly than they previously could in weeks. Computer models and data base networks have become so powerful in some research fields that they can identify critical relationships and pose new hypotheses, not merely help analyze and test data. The Murillo chapter suggests the power of computer techniques in engineering design of structures. In other services industries, computer-based techniques are helping to find feasible structures for complex but as yet unknown proteins, identify potential disease causes from epidemiological or genetic models, or (as Richard Larson's chapter demonstrates) solve inordinately complex problems in delivering public services.

Within companies, the effects of complexity reinforce the strategic effects of economies of both scale and scope. Companies that can deliver better and more varied services without increasing their marginal costs can achieve competitive advantages through a higher degree of segmentation in their marketing activities, increasing value added for their customers, while lowering their own average operating costs. The Fellowes and Frey chapter provides an excellent example. But the technologies that make such systems possible are having their own interesting side effects. They are undermining established concepts of competition, strategic management, organization, and economic policymaking at a rate never before envisioned. They have shortened transaction times, heightened volatility, and created entirely new forms of internal organizations, external alliances, and industry competitive structures in both manufacturing and services.

Services technologies have created a new level of responsiveness to markets by disintermediating (eliminating intermediaries) between customers and the producers of goods or services. Direct access to financial services markets has, of course, short-circuited many traditional banker, agent, and broker relationships; and easy direct ticketing by airlines, hotels, tours, and theaters has cut out other intermediaries in these services fields. The entire production process from raw materials to ultimate customer purchase can now be integrated by electronics systems, without entailing the ownership and control patterns involved in earlier attempts at "vertical integration." Any level in the value chain from raw material producer to retailer can drive this process.

Intermediaries themselves (such as wholesalers) can simultaneously vertically and horizontally integrate their markets by providing a better level of price, quality, and service than their customers could possibly achieve alone. For example, McKesson and Super Valu provide their retailing customers with the benefits of vertical integration (without ownership) by helping them locate and design their stores, advising them on increased value-added product mixes, managing stocks and displays in key departments, processing medical insurance claims for customers, recycling customer wastes, handling customers' accounting and credit functions, and researching new uses for suppliers' products. Sears owns and controls many suppliers for its branded lines and provides after-sales services through its own extensive financial, credit, repair, and insurance networks.

Computerized reservation and product control systems allow airlines, automobile rental companies, and retailers to analyze their costs and customers buying behavior in such detail that they can optimize margins on each type of demand and meet each competitor's response. Research indicates that the resulting crazy quilt of prices leads customers to concentrate more on services provided. This in turn offers attentive services companies greater opportunities to segment their markets with highly personalized responses such as Domino's Pizza recognizing customers' names, addresses, and preferences instantly; airlines offering specialized meals, wheelchairs, luggage verification, and even counseling for nervous passengers; and banks offering uniquely tailored products and services to individual clients, depending on their roles in their particular companies and their personal banking and credit histories. Such possibilities have led to whole new approaches to designing flexible, responsive organizations that can delegate authority with "spans of control" as wide as 200 people and empower and motivate employees at all levels to innovate and exploit their own creative potentials and the opportunities that the new technologies allow (Carlzon, 1987).

Services technologies are thus breaking down the traditional boundaries of nations, industries, and even government versus private functions. With no element in its value chain immune to structural changes due to services technologies, each producer must constantly reassess who its true suppliers, competitors, and customers are and how each could enhance or subvert its competitive posture. Just as international banks already often find that their compatriots may be simultaneously competitors, customers, joint venture partners, and suppliers, other manufacturing and services concerns are entering a new competitive era where substantially increased cross-industry competition is a fact of life.

The various cases in this volume offer specific and detailed examples of major innovations that have helped restructure services companies and industries in the ways described above. The next section of this paper, however, will focus on another set of innovations not widely recognized but also radically changing the country's basic economic landscape and competitiveness-the manufacturing-services interface.

THE MANUFACTURING-SERVICES INTERFACE

In recent years many have expressed concerns over the decline of U.S. manufacturing. They fear a U.S. economy dominated by services will mean low wages, diminished real economic growth, and a society engaged in menial personal services tasks. These perceptions are incorrect.

Total employment in manufacturing has decreased only marginally from long-term trends (see Figure 1). Real value added attributable to manufacturing grew steadily until the mid-1980s (see Figure 2); and U.S. manufacturers' percentage of total goods trade (including their overseas production) has stayed between 20 and 22 percent since the 1960s. Many important U.S. manufacturing industries (such as aircraft, pharmaceuticals, computers, chemicals) are strong and viable, although some (such as basic steel, automobiles, or electronic appliances) have suffered real declines in employment, output, and competitiveness over the last 15 years.

Meanwhile the services sector has grown steadily in its contributions to U.S. gross national product (GNP) and value added and in 1986 accounted

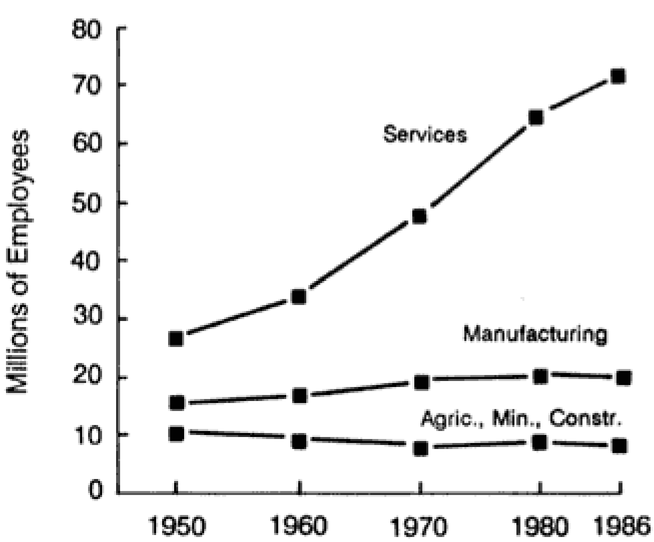

Figure 1 Employment trends by sector. SOURCE: Bureau of Economic Analysis, The National Income and Product Accounts of the United States; and Bureau of Labor Statistics, Establishment Database.

for 71 percent of the country's GNP and 75 percent of its employment (see Table 1). Far from being a negative development, today's services industries actually create major markets for consumer goods, lower virtually all manufacturers' costs, provide strong stable markets for capital goods producers, and enhance overall U.S. competitiveness in world markets. New technologies in services both have improved real economic growth significantly and offer manufacturers major new strategic opportunities and potentials for future growth.

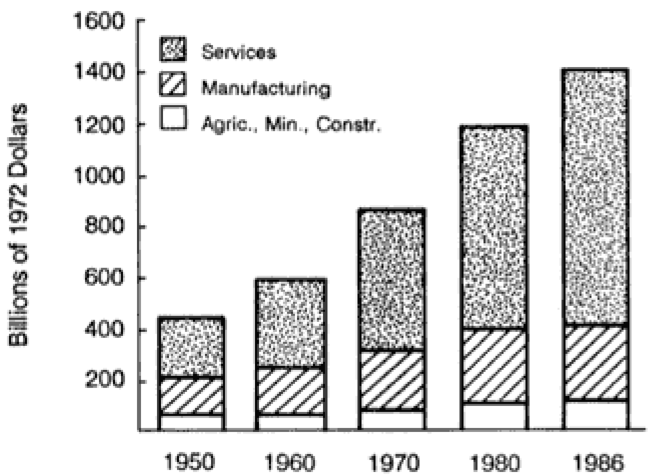

Figure 2 Real value added by sector. SOURCE: Bureau of Economic Analysis, The National Income and Product Accounts of the United States.

Hypothesizing a presumed conflict of services versus manufacturing is counterproductive. The services and goods-producing sectors of the economy are so intertwined that it is inappropriate to consider one as somehow subordinate or in opposition to the other.

Services Directly Substitute for Manufactures

Far from being inferior economic outputs, services are directly interchangeable with manufactures in a wide variety of situations. Few customers care whether a refrigerator manufacturer implements a particular feature through a hardware circuit or by internal software. New CAD/CAM software can substitute for added production or design equipment, and improved transportation or handling services can lower a manufacturer's costs as effectively as cutting its direct labor or materials inputs. These "services" investments improve productivity or add value just like any other new investment in physical-handling machinery or product features.

TABLE 1 U.S. GNP and Employment by Industry, 1986

	GNP		Employment	
	$ billions	% of total	Millions	% of total
Total economy	$4,235		108.0	
Agriculture, forestry, fisheries	93		1.7	
Mining and construction	293		5.7	
Manufacturing	825		19.1	
Total goods sector	$1,211	29%	26.5	25%
Finance, insurance, real estate	$695		6.5	
Retail trade	408		18.4	
Wholesale trade	295		5.8	
Transportation and public utilities	276		4.0	
Communications	115		1.3	
Other services	700		24.9	
Total private services	$2,489	59%	60.8	56%
Government and government enterprises	507		20.6	
Total services sector	$2,996	71%	81.4	75%
Rest of world and statistical discrepancy	30			

SOURCE: U.S. Bureau of Economic Analysis; U.S. Bureau of Labor Statistics.

The U.S. services sector is a natural and desirable outgrowth of an increasingly productive agricultural and industrial economy. People were able to buy ever more with what they earned each hour of their working time. But demands for products were inherently somewhat capped; people could only consume so many washing machines, sofas, or pounds of food. Consequently the relative utility of services grew apace. Simultaneously, new technologies also vastly improved performance and opened further markets in virtually all services industries.

From the manufacturer's viewpoint jet aircraft made long-haul passenger and freight handling much more efficient and convenient. New containerization, loading, refrigeration, and handling techniques for volatile liquids, by making it possible to transport virtually all goods safely and effectively, vastly extended international trade in manufactures. And electronics, information, and communications technologies stimulated new initiatives in virtually all services areas—most notably in retailing and wholesale trade, engineering design, financial services, communications, and entertainment. Note that virtually all these advances involve interactions between manufactured products and services.

However, total interactions between services and manufacturing are much more complex. Figure 3 suggests only some of the principal relationships. Economic benefits flow both ways between all these activities; the fact of trade means that each party must benefit from the presence of the other. Although most of the structure seems self-evident once disclosed, the role of the "service intermediary" may not be. Not all products can be used by the ultimate user without the intermediation of a professional who can harness its more arcane possibilities or control its potential hazards. Specialized technical representatives or consultants are often needed to introduce or support new product entries in high-technology fields such as nuclear energy, computers, or medical care systems. Without these services, market access would be prohibitively high.

SERVICES GENERATE MANUFACTURING MARKETS

Many have noted that some U.S. services industries are very dependent on manufacturing; many services companies exist primarily to provide transportation, finance, advertising, repair, distribution, or communications in support of goods manufactured here. A large number of these services would still be provided in the United States regardless of where the product was manufactured, for example, selling a Toyota in this country requires many of the same U.S.-based services activities as selling a Ford. However, important design support functions and supplier linkage services might move overseas if manufacture was not performed here.

Less recognized is the fact that a healthy manufacturing sector is probably equally dependent on services. First, if 75 percent of all people are employed in services, clearly they and their enterprises must be the major markets for most consumer and commercial products. Second, although the lagging data for U.S. input-output tables—and poor definition of sectors—do not allow complete intersectoral sales calculations, some specific studies suggest that 85 percent of the communications and related information technologies equipment sold in the United States in 1985 went to the information industries (dominantly services) (Roach, 1988), and 70 percent of all installed computer systems in Great Britain in 1984 were in services (*The Economist*, 1985).

In addition, services technologies can allow manufacturers to be much more responsive to fluctuating or individual demand patterns. Proper integration of these technologies throughout manufacturing and distribution can significantly increase the number and range of goods it pays to produce in the United States, rather than overseas. As personal affluence or the sheer size of markets grows, there is an increasing demand for differentiated,

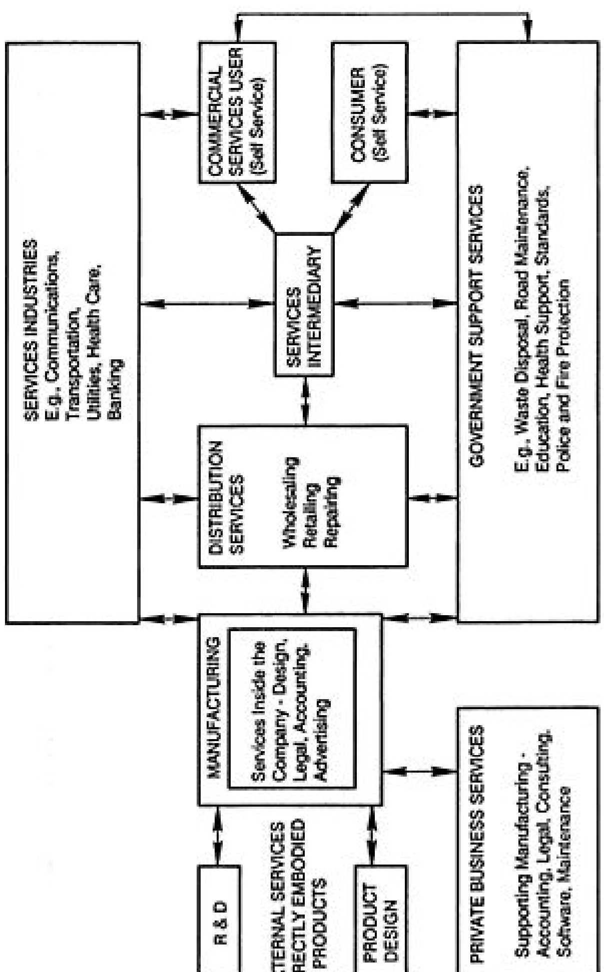

Figure 3 Some mutual interactions among manufacturing and service activities. Double-headed arrows indicate that each party benefits from the presence of the other in the trade.

individualized, or customized products. Thus, original equipment manufacturers (OEM) and wholesale-retail customers generally want shorter or just-in-time (JIT) inventory responses and higher variability in products available. Also final consumers want delivery on products that reflect their particular tastes now, rather than in 3–12 weeks. Both flexible manufacturing and automated services technologies can help manufacturers exploit these trends and increase their markets.

Services Technologies Improve Market Responsiveness

Manufacturing success today often requires more rapid feedback from the marketplace, better customized products, and more accurate delivery in shorter cycle times, all of which are dependent on services integration. So important is market integration that Texas Instruments tries to place its application-specific integrated circuit technical representatives in its customers' design shops, and consumer goods producers such as Mr. Ge van Schaik of Heineken's sometimes note, "We are just a marketing company with a production facility." Other specific examples will suggest how new technologies help manufacturers effect better marketing-production integration:

- Clearly, American Hospital Supply built its preeminent position by placing computer terminals right in its customers' premises and providing updated software to help users control inventories, access catalogs, and order American Hospital Supply products directly from the terminals. Similarly, by affixing uniform product code (UPC) bar-code labels on its products at the factory and offering electronic ordering systems, Levi Strauss is now able to provide extremely fast replenishment of stock for retailers who have installed bar-code scanners at their checkout registers —a practice long used by supermarkets (*Business Week*, 1987c).
- Some manufacturers are closely tied to retailers that, like The Limited, daily aggregate sales data electronically from their entire retail networks, transmit restocking orders by telecommunications to their manufacturer-suppliers all over the world, and expect accurate deliveries within days to their specific distribution points. "Quick response" reordering systems can allow American textile mills significant competitive advantages in the fashion apparel business, which now constitutes 40 percent of apparel sales. They make it possible to deliver fabric in one-third the time it takes from Taiwan, and CAD/CAM links between cutters on Seventh Avenue and Southeastern mills can halve the time from design to delivery (*Business Month*, 1987).
- When a Stratus Corporation customer's mainframe computer malfunctions, its backup system is designed to take over and send a signal to Stratus specifying the particular part the computer may need. The first time the customer may know a problem exists is when the replacement part arrives with the next morning's express delivery (Davis, 1987, p. 14).

U.S.-based manufacturers able to link flexible production systems directly to their customers' market intelligence networks should have very real competitive advantages (in both timing and transportation costs) over foreign producers. In the early 1970s Continental Can locked in its customers by putting its Conoweld lines directly into its customers' premises offering the ultimate in supplier-customer integration. Similarly, in heavy, bulky, or complex products such as automobiles, it could become impossible for overseas producers of more standardized lines to compete against well-developed, responsive, integrated, U.S. manufacturing-distribution systems. Already foreign firms investing in U.S.-based manufacturing capacity to get closer to the huge, highly variable U.S. marketplace are helping to generate a "remanufacturing" of the United States. This is noticeably under way in the automotive industry.

- Such considerations seem an important factor in Honda's rapid proliferation of auto models and options, its increased emphasis on U.S. production, and the importation of its own suppliers to the United States (when U.S. manufacturers proved unwilling or unable to meet Honda's service and quality demands). Honda now offers a much wider variety of styles and options and faster delivery to the U.S. marketplace than it ever could have from a solely Japanese base. Increasingly, it will have to source more of its high value-added components and subassemblies from nearby U.S.-based producers—to obtain the response time it needs without escalating inventory costs beyond economic levels. In addition, Honda will shortly be helping reduce the U.S. merchandise trade deficit by exporting American-built Accords to Europe and Japan (Business Week, 1988b).

Manufacturers Become Services Producers

In a variety of other ways, services technologies and their extensions as new "products" for exploitation are becoming vital competitive weapons for large "manufacturing" companies. For example:

- General Motors' or Exxon's competitive positions are largely determined by their capacities to manage information worldwide—about suppliers, new technologies, exchange rates, swap potentials, or the changing political or market sensitivities in key countries. With crude oil resources primarily in the hands of sovereign nations, Exxon's profits depend ever more on its ability to find, track, deploy, trade, transport, finance, and distribute energy efficiently. All these "services" activities are technology driven. In addition to such logistics activities (especially in worldwide sourcing), General Motors has found financial services to be an indispensable competitive weapon in the marketplace. General Motors Acceptance Corporation (GMAC) now manages over $75 billion in private consumer debt, has provided 1/3 to 1/2 of GM's profitability in recent years, and has allowed GM to offer 1.9 percent financing as a pricing strategy to offset competition on a "features" or production cost basis.
- At IBM, software has always been a key to success, especially in the early years when it was "bundled" into product and rental prices. Now, with the hardware aspects of computing becoming extremely low cost and competitive, IBM has been

shifting its focus even more toward software, networks, and communication linkages (services) as its basis for improving value added and profits (*Business Week*, 1987b).

Some manufacturers have initially ventured into services to support their main product offerings, then found that the service itself can become an attractive new product line. For example:

- Because of the very high costs its customers incurred if their equipment went down for repair, Caterpillar Co. developed one of the largest, fastest response parts systems in existence—with 23 distribution centers, 10 million square feet of storage space in 11 countries, and 340,000 stock keeping units (SKUs) available to customers on a 24-hour-turnaround basis for most parts. Recognizing the success of this system, around 1980 some of Caterpillars' customers began asking it to help them manage aspects of their parts control and distribution systems. Caterpillar responded and now has a lively business (Caterpillar Logistics Services) providing logistics and transportation services to large manufacturers (such as Land Rover) interested in efficient warehousing and delivery of their own parts on a worldwide basis (*Distribution*, 1987a).

Similarly, foods companies, such as Pillsbury, have found that their large scale distribution and support systems (such as Davmor and Distron servicing Pillsbury's fast food chains) can provide potential new growth opportunities in industrial and institutional feeding. Manufacturers of steel and glass run short-haul railroads that deliver bulky materials to their own and other manufacturers' plants. One manufacturer of measurement and control devices used in nuclear power plants runs a personnel services firm—with revenues of more than half the parents'—supplying trained personnel to operate the complicated measurement and control devices in nuclear facilities. Creative expansions of their own services capabilities can offer other manufacturers new opportunities to share in the growing services market directly.

Services Lower Costs and Increase Product Value

Many aspects of a manufacturer's cost competitiveness depend intimately on services. Greater efficiencies in communications, transportation, financing, distribution, health care, or waste handling (services industries) can markedly affect a manufacturer's direct costs. To the extent that these services also improve the quality of life or lower living costs in the United States, they indirectly create downward pressures on U.S. wages relative to foreign nations. That is, for any given wage level U.S. manufacturing employees can live better than foreign workers in countries with less efficient services. And in most areas the efficiency of U.S. services industries has been more than competitive, even compared with Japan (see Table 2). In fact Japan has usually suffered a large negative trade balance in services (Tanaka, 1984).

It should be noted that Japan's services industries have not been standing still since 1980; the application of new technologies, deregulation, and rising consumerism have no doubt created restructurings, innovations, and improved efficiencies. The following attest to that trend: (1) the rapid rise in catalog sales and the increasing use of telemarketing techniques; (2) the innovations in the household-moving business led by Art Moving and the boom in just-in-time trucking and parcel courier services; and (3) the installation by deregulated insurance companies of expanded nationwide computer networks designed to increase the level of automation in field offices and to improve their headquarters' ability to develop new products. Given Japan's emphasis on the information industries and its proved capability to apply new technologies, one would expect the trend to grow even stronger in the future.

TABLE 2 U.S. and Japanese Productivity in Servicesa

	Japan		U.S.		U.S. Output per Hour as a % of Japanese Output per Hour,
	1970	1980	1970	1980	1980
Private domestic business	3.59	6.01	9.40	10.06	167
Agriculture	1.37	2.38	16.53	18.36	771
Selected services	3.86	5.66	9.29	13.14	232
Transportation and communications					
Electricity, gas, water	14.01	19.74	21.98	25.38	129
Trade	2.88	4.53	6.88	7.92	175
Finance and insurance	6.69	12.03	8.21	8.20	68
Business services	2.29	3.60	7.69	7.59	211
Manufacturing	3.91	8.00	7.92	10.17	127

[a] Measured as output per hour in 1975 dollars.
SOURCE: UNIPUB, 1984.

Within manufacturing some 75 percent of all costs—and a much higher percentage of value added—is generally due to services activities (Office of the United States Trade Representative, 1983; Vollman, 1986, p. 149). The major value added to a product is typically less due to its basic commodity value (i.e., producing the "body in white" for an automobile or the grain and vegetables in a processed food) than to the styling features, perceived quality, subjective taste, distribution, marketing presentation added by "services" activities inside or outside the producing company. Except in continuous process industries (such as oil refining or chemicals manufacturing), planning, accounting, inventory, quality assurance, transportation, design, advertising, and distribution costs generally outweigh direct labor costs by factors of between 3 to I and 10 to 1.

Most companies have diligently driven out direct labor costs while only cautiously attacking the greater costs and value added potentials of their services functions. Aggressively managing services activities within manufacturing enterprises to both improve quality and lower costs can provide a major attack point for improving competitiveness. Successful quality programs start by analyzing quality from the customers' viewpoint and then emphasize quality considerations in all functions from design through final shipping and presentation, mostly services activities. Two examples will suggest how some large companies have approached this complex problem.

- In the automobile industry, designing and introducing a new car line are major components of overhead costs and significant contributors to the value the customer perceives in the product. At Ford this design cycle had taken 5 to 6 years and could cost several billion dollars. In 1980, Ford's chief executive officer Philip Caldwell and president Donald Petersen decided to undertake a new "simultaneous design" process for their new upper middle line, which became the Taurus/Sable cars. To integrate design from research to marketing, they established a core group, called Team Taurus, which had a full-time representative from all of the critical groups within Ford—product design, component engineering, manufacturing, sales, marketing, purchasing, service, legal, environmental and safety engineering, and corporate headquarters (virtually all "services operations"). Instead of moving the design linearly along these groups to the marketplace—with numerous conflicts and design reworks—Team Taurus involved everyone from suppliers to workers, distributors, repairmen, customers, and even insurance and media representatives from the outset. The result: Taurus/Sable's introduction cost some $250 million less than its design budget, and the cars were exceptionally successful in the marketplace. Future cars will be designed with the same process in only 3 or 4 years—potentially saving Ford hundreds of millions of dollars more and ensuring that manufacturing design is much closer to customer desires.
- Black and Decker (B&D), faced by both U.S. inflation and strong cost competition from overseas manufacturers, decided it had to radically redesign its products for greater value added and lower cost. It focused on (1) simplifying its product offering, (2) developing a "family" look for the 122 basic tools (with hundreds of variations) in its line, (3) standardizing materials and components, and (4) creating new product features specifically adapting the products to meet worldwide (non-U.S.) product specifications. In a program that took 7 years to break even, B&D drove labor costs down in each component to such "trivial" levels that no competitor could obtain a meaningful competitive edge by simply having lower labor costs. Interestingly, however, B&D found that its simplification and automation programs drove out about $3 of overhead for each $1 of labor saved, and its standardization programs substantially improved the rapidity, efficiency, and effectiveness of its future product development capabilities (Lehnerd, 1987).

The Strategic Planning Institute's Profit Impact of Management Strategy (PIMS) data base clearly shows that financial performance is directly related to the perceived quality of a company's goods and services. Not surprisingly, one of the biggest determinants of perceptions of overall quality is customer service. As the two examples above illustrate, greater attention to product design which ultimately shows up as customer service is a significant component of improving quality. Using technology more directly, Cadillac has vastly increased its service ratings by linking its dealers' mechanics with a corps of 10 engineers who use a computerized data base of engineering drawings and service bulletins to walk them through difficult repairs so that a Cadillac gets repaired right the first time (Uttal, 1987, p. 116). In an effort to improve the product support so essential to sales of its aircraft engines, Pratt & Whitney—following the dictates of SAS's Jan Carlzon—has slashed layers of bureaucracy and pushed decision making down the chain of command to the people who actually deal with the customers so they can fix or replace a part and worry about whether it is covered in the customer's support contract later (*Air Transport World*, 1988).

Services Support International Manufacturing Operations

One of the areas where services technologies affect manufacturing most markedly is in international operations. Telecommunications, air transport, and improved surface cargo handling technologies have forced virtually all manufacturers to consider their supply sources, markets, and competition on a worldwide scale or to lose their competitive position. Although the figures do not show up in merchandise trade balance statistics, the greatest impact of U.S. manufacturing technology on world markets is probably through multinational companies' operations inside host countries. Approximately one-fifth of the total capital invested in U.S. manufacturing firms is in facilities outside the United States. And some of the largest continuing favorable net balance-of-trade accounts for the United States have been the profits, royalties, and intercorporate sales these multinationals remit to the United States (see Table 3).

Effective coordination of these giant enterprises—and many much smaller companies' international transactions—clearly depends on services technologies and efficiencies.

- IBM, in bringing forth its 360 computer series in 196$1$2$3964, is often credited

TABLE 3 Manufacturing Related Trade Balances ($ Billions)

Category	1970	1975	1982	1985	1987[a]
Royalties and fees, net	2.1	3.8	4.6	5.3	6.8
Direct investments, net	2.6	14.4	18.2	26.6	35.3
Merchandise, net[b]	2.6	8.9	(36.4)	(122.1)	(159.2)

[a] Preliminary figures.
[b] Excluding military.
SOURCE: Bureau of Economic Analysis, U.S. International Transactions.

- with the first truly simultaneous international design of a product line. Its various laboratories became permanently linked by telecommunications for daily coordination of the design process. Ford Motor Company has recognized the huge advantages it can achieve by designing and engineering different sized platforms for its cars only once (at specified "centers of excellence") to get maximum economies of scale, then doing the styling and feature designs for its different markets within its subsidiaries in individual countries. The corporation later coordinates cost comparisons for various features, sourcing performance, market information, parts inventories, and customer service activities through its large transborder data network (*Business Week*, 1987a).
- Global information systems and data bases dealing with exchange rate fluctuations, transportation alternatives, and sourcing performance have become crucial competitive weapons in manufacturing. While labor, materials, and plant level overhead costs tend to drop markedly with offshore manufacture, logistics costs tend to increase by a factor of about 20 percent and tariffs can become a significant factor. So important are these costs that before an experienced Japanese automobile manufacturer would begin its joint venture in the United States, it set as its target having "the most cost efficient inbound-outbound logistics system in the world." Since most of the other costs of manufacture here and in Japan were essentially fixed, logistics control technologies became the key strategic variables (*Distribution*, 1987b).

Recognizing the importance that cross-border data and services flows have in foreign companies' competitive postures, many host countries are trying to tax or regulate such activities. Since the real economies of scale multinational manufacturing companies most often enjoy are due to their services capabilities—i.e., technology transfer, marketing skills, financial services, logistics—rather than plant scale economies, such restrictions on data and services flows are as significant to manufacturers as they are to international services providers. Worldwide integration of manufacturing operations through services technologies is essential for producers seeking competitive advantage in today's world.

Manufacturers Benefit from External Services Innovations

As technology creates new capabilities, economies of scale, or economies of scope for services, manufacturers will constantly have to reassess when to produce their own services or buy them "out of house." Many have found that specialized services companies can handle their accounting, legal, payroll, benefits, maintenance, repair, or even research and design functions much more effectively than they can "in house." This is one of the factors contributing to the rapid growth of the business services industry in recent years—in 1982 business services accounted for $90.7 million of GNP and employed 3.4 million people; by 1986 it provided $162.8 million of GNP and 4.9 million jobs (Kutscher, 1988; Tschetter, 1987).

More than this, services companies have become major innovators on their own. Focused and imaginative technology investments by many business and professional services providers have greatly improved the quality, range, and flexibility of their services offerings. Although these services companies often cannot translate such product improvements into higher margins, manufacturers and other customers benefit directly. Unless manufacturers systematically and continuously cultivate the innovations services suppliers can provide, they will miss out on important sources of competitive advantage. For example:

- Federal Express has been a leader in the development of package sorting, handling, and control equipment and techniques. Based upon this expertise, Federal Express now offers its package tracking and supporting systems as a part of a service to help producers automate their own shipping docks and package handling activities.
- A freighter waiting at the Suez Canal or at a customs wharf in France will often have to pay all required fees before it can pass through or unload. As an alternative to paying an 8-day float on funds cleared through various correspondent banks and $10,000 per day of waiting cost for the ship, Citicorp offers a service for instant transfer of funds to a local (Port Said or Marseilles) correspondent bank as soon as the fee is known—thus saving manufacturing customers unnecessary costs on the (otherwise) delayed cargo (Davis, 1987, p. 18).
- Michigan Bell has created innovative new "facilitating" services—such as those which improve communications among the Big Three automakers and their suppliers-coordinating hardware and software vendors in developing and exchanging engineering drawings and specifications between job shops and factories. Although Michigan Bell started with such large company networking projects, it should be able to roll out this experience to other areas. It can offer a shared service (with a standard interface system) providing state of the art design and specification coordination to large and small manufacturers alike (*Managing Automation*, 1987).
- Finally, services technologies offer a rich new array of channels through which manufacturers can reach specialized segments of their markets. Electronic home shopping and interactive video terminals located in banks, airports, hotels, airplanes, and shopping malls allow manufacturers to contact whole new groupings of customers in a psychological situation where they are likely to buy. Similar technologies in retail showrooms allow customers to see a full array of a manufacturer's product line when the retailer—for example a shoe or furniture outlet—could not carry all possible sizes and variations of the product.

With services enterprises growing rapidly in scale and technical sophistication, the opportunities are expanding for manufacturers to leverage their own internal innovation programs by exploiting those of services suppliers, just as they do those of product vendors. Although 75 percent of a manufacturer's costs are in "services" activities, flexible manufacturing systems (FMS) and computer integrated manufacturing (CIM) are likely to drive variable costs even lower—and hence the percentage of costs represented by services support and other fixed costs higher. Consequently, there should be enhanced incentives for more effective use of services suppliers' innovative capabilities. In fact, FMS and CIM installation themselves are unlikely to be effective if proper attention is not given to services support arrangements such as training, cost accounting, personnel practices, organization, communication, and control systems. Because internal groups are likely to be inexperienced in such matters or biased by old ways of doing things, the required expertise for such changeovers may often be found only in external services groups.

MANUFACTURING'S CHANGING STRATEGIC ENVIRONMENT

Perhaps the most important structural change in international manufacturing competition stems from the continuing integration (through electronics) of the world's financial centers into a single world financial marketplace. World financial flows have already become largely disconnected from trade flows.[2] Although world trade in goods and services aggregates only $3–4 trillion annually, financial transactions by the Clearing House for International Payments (CHIPS) alone totaled $105 trillion in 1986—and early 1987 transactions were running more than $200 trillion on an annualized basis. Turnover on the 1986 Eurodollar market also amounted to $75 trillion. Any of these sums dwarfs world merchandise trade, just as Fedwire's $125 trillion of 1986 domestic transactions dwarfed the $4 trillion GNP of the United States. Instead of following trade in goods, money now flows toward the highest available real interest rates or returns in safer, more stable economic situations.

Exchange Rates Determine Manufacturing Costs

As a result, exchange rates have fluctuated approximately ± 50 percent among major trading partners within a few years, principally because of fiscal or monetary—not trade or management—decisions. Comparative costs for an international competitor are often more a function of exchange rates than of productivity or competitive managerial decisions. Even skillful Japanese manufacturers have found it difficult to compete when the rising value of the yen pushes their relative costs up by 48 percent in 18 months against competitors paying for wages and materials in U.S. dollars. But they appear to be responding with dramatic productivity increases and transfers of operations offshore (*Business Week*, 1988a).

As was recently demonstrated, the short-run volatility of securities markets has amplified enormously because of the capacity of large investor groups to move rapidly into or out of world securities and funds markets. The stimuli that disrupt a nation's capital markets and relative cost structures can easily come from outside sources. Consequently, it has become increasingly dif ficult for sovereign nations to control their economies in the short run or to fine tune them through traditional fiscal or monetary interventions. And it has become impossible for major manufacturers to operate effectively without global plant location, sourcing, logistics, and financial strategies.

With freer access to world capital markets everywhere, it is becoming very difficult for a single nation to maintain differentially low capital costs as a policy—as Japan has in recent years and as the United States once did—in support of aggressive economic development or trade objectives. As capital costs among nations are leveled by globalized financial markets, countries such as the United States and Japan—with high domestic labor and materials costs—will be under ever greater pressures to move manufacturing overseas. Among the few alternatives available is innovation, in both technical and structural terms, at a rate others cannot match. A key element will be innovative utilization of the full potentials of services technologies both internally and externally. Simply targeting more reductions in direct labor and materials costs will not be sufficient (Bleeke and Bryan, 1988).

New Power Relationships with Services Groups

The entire power relationship between manufacturing and services groups is changing profoundly. For example:

- As the number of major airlines (services providers) declined through consolidation due to deregulation and the introduction of wide-bodied jets, aircraft manufacturers had to modify their strategies. The number of U.S.-lead customers for a firm such as Boeing has shrunk to a handful of large "megacarriers," each with a more powerful bargaining position than before. The military no longer can provide the "first order" volumes necessary to launch a commercial aircraft. Consequently, to obtain essential economies of scale, yet satisfy each customer, companies such as Boeing have invested heavily (more than $800 million) in flexible automation, allowing them to modify their aircrafts' internal seating, baggage storage, maintenance system, and customer convenience patterns to suit virtually any requirement their major customers might conceive over the useful life of the aircraft. Boeing is even proposing to restructure its overall cabin configuration (in its 7J7 aircraft) in response to the recommendation of SAS's President J. Carlzon that airline seating patterns would be optimized if the cabin cross section were a horizontal ellipse rather than a circle or vertical ellipse. The coming deregulation of Europe's airlines will only lead to further concentration in the world's airline industry.
- Even when buyers have not consolidated into large individual units, services technologies allow them to achieve similar bargaining power. For example, a team of airline and manufacturer-supplier experts, coordinated by the Air Transport Association, has created an automated parts procurement system (Spec 2000) that is expected to save millions of dollars and can be accessed using a personal computer. Spec 2000 is a computerized compendium of manufacturers' catalogs including more than a million part numbers and an automated computer-to-computer ordering system.

Not only does the system allow an airline to get comparative specification and quotation data, but it also gives information about how vendors meet their delivery and performance criteria, thus increasing the information leverages of all potential purchasers (*Air Transport World*, 1987). Retailers with electronic point of sale systems and those that use McKinsey's Direct Product Profitability (DPP) analyses have also found that they now have the best information and thus increased bargaining power over their product suppliers. In response, Procter & Gamble has reorganized itself and changed many of its long-established attitudes and ways of doing business.

- Large retailers, such as Sears (or IKEA), can offer instant distribution for any manufacturer's consumer product (or furniture) innovations, with powerful presentation and financial support for any products they accept. Thus, overseas producers have immediate access to U.S. (or European) markets without creating the costly distribution channels they once had to build. Access to large retail electronics chains has provided rapid penetration for many Asian "clones" competing with U.S. manufacturers. Along with the capacity of the retail and distribution system to help manufacturers target and design their products better, such potentials call for much more attention to carefully developed "downstream alliances" in manufacturers' strategies.
- Services suppliers' cost leverages have also increased. Many manufacturers find that their medical care or insurance outlays for employees are higher than their own profits. Hence, new strategies creating "coalitions" with providers and insurers have emerged as key elements in cost control.3 Deregulation has created more powerful transportation companies with intermodal handling capabilities that can increase shipping efficiencies enormously, but at the cost of much more bargaining power against manufacturers or other shippers. And large money center banks have the information, instant capital access, and worldwide connections to manage a company's financial assets with a knowledge base and leverage few manufacturers can equal.

Disintermediation Builds New Industries

Disintermediation of service activities through product substitutions offers powerful new strategic options. All "services" or "goods" are really just means for providing satisfaction to customers. Thus, the boundary between services and manufactures is very fluid and varies widely over time. Recognizing this, manufacturers can greatly extend the scope of their operations, lower costs, and expand their margins by eliminating or taking over adjacent services functions—sometimes opening up entire new support industries. For example:

- In the early days of computers, a high priesthood controlled access to huge machines that ran only in their own special, environmentally controlled temples. Programmers or others "submitted" punched cards or tapes for processing, often at inconvenient times determined by the priesthood. By designing their machines and software to be more "user friendly," manufacturers slowly eliminated the priesthood and expanded their markets in the process. The success of Apple computers and the

IBM PC was essentially due to service disintermediation, which (oddly enough) later opened more services and product opportunities through clever use of the computers and design of software for them.

- "Permanent press" and crease resistant fabrics, textile and fibers manufacturing innovations, have significantly restructured the white goods, apparel, and personal services industries. In the past, those who could afford it sent their tablecloths, bed linens, and dress clothing (especially men's trousers and shirts) to the laundry, getting them back neatly ironed in a week or so. Others spent hours tediously ironing in their own homes. Permanent press and "wash and wear" fabrics shrank the commercial laundry business, increased the attractiveness of home washers, dryers, and "touch up" irons, and freed domestic workers and homemakers for other tasks. To fully exploit permanent press possibilities (a services disintermediation), new detergents, whiteners, and home laundry equipment designs soon appeared, changing the nature of competition in those fields as well.

The potentials for such substitutions of products for services abound. Automated communications systems have eliminated tiers of banking, brokerage, and insurance personnel. Increased "quality" in home appliances and automobiles has decreased the personal time, cost, and frustrations consumers expend on maintenance and repair. Prepared gourmet foods, convenience mixes, and microwave dinners substitute for restaurant sales and home cooking time. And so on.

CONCLUSIONS: SERVICES INNOVATION

Discussions concerning U.S. competitiveness have tended to overlook two major areas for innovation and productivity improvement—services and the manufacturing-services interface. Technological changes in services have radically altered the economy and the strategic environment in ways that offer significant new opportunities and threats for technologists and investors. In addition to providing major innovative opportunities in their own right, services industries have become some of the most important customers, suppliers, and coalition partners for manufacturing concerns. In these roles, U.S. services enterprises both are near at hand and are generally among the most efficient performers in the world. In addition, services technologies offer a variety of ways for manufacturers to add value or lower costs within their own operations. This paper has merely attempted to offer a few useful case vignettes; other cases in this volume will develop some of these in detail.

The United States painfully learned that its goods-producing industries were vulnerable to challenges from abroad. So too are the services on which much of the nation's economic wealth depends. It is difficult for even well-run services establishments to maintain their competitive advantages when everyone can buy the same hardware and software, connect into the same networks, and exploit the offerings of suppliers with strong incentives to sell their products as quickly and widely as possible throughout the world. It is becoming ever more important that services companies develop their own internal technological capacities and strategically leverage these in conjunction with manufacturing concerns and other services companies.

It will take hard and dedicated work not to dissipate the broad-based lead that the United States enjoys in services. Seemingly secure markets may be easy to invade. On the one hand, U.S. policies of deregulation in communications and financial markets may have given the United States a lead that, some say, other countries' more regulated sectors will find difficult ever to close. But those countries are deregulating, too, and there is little room for complacency. Deregulation also opens domestic U.S. markets to foreign competition. And many countries and foreign companies have proved that their skills in managing services enterprises are formidable indeed.

Unfortunately, we daily encounter the same inattention to quality, emphasis on scale economies rather than customers' concerns, and short-term financial orientation that earlier injured manufacturing. Too many services companies have been slow to invest in the new market opportunities and flexible technologies available to them. They have stayed with their old concepts too long and have concentrated on cost-cutting efficiencies they can quantify, rather than on adding to their product value by listening carefully and flexibly providing the services their customers genuinely want. The cases and concepts in this book are intended to help others avoid these errors and understand some of the opportunities and complexities of introducing new technologies in the services and services-manufacturing areas. The strengths of manufacturing and services in the United States are intimately intertwined. Increasingly, success in both services and manufacturing will go to those who understand and successfully combine the new potentials of services and manufacturing technologies.

ACKNOWLEDGMENTS

The author gratefully acknowledges the important contributions of Dr. Jordan J. Baruch and Penny C. Paquette in developing this paper, as well as the generosity of Bankers Trust Company, Bell & Howell Company, Royal Bank of Canada, Braxton Associates, and Bell Atlanticom Co. in supporting this research.

NOTES

1. The selected business services category includes travel, passenger fares, and other transportation—areas where large negative balances exist today—as well as fees and royalties where positive balances have continued to grow and other private services where balances have changed little over the 5-year period from 1982 to 1987. Statistics are taken from the U.S. International Transactions series published in *Survey of Current Business*.

2. S. Bell and B. Kettell estimate that 95 percent of the daily volume in foreign exchange markets in 1983 was not direct commercial business but trading between the foreign exchange dealers of the world's international banks (*Foreign Exchange Handbook*, Quorum Books, Westport, Conn., 1983, p. 3).

3. At the national level, The Washington Business Group on Health has been attempting to coordinate provider, payer, and government agency groups that need to cooperate on this issue and started publishing *Business and Health* to bring the views of both private and public authorities to the fore.

REFERENCES

Air Transport World. January 1987. Spec 2000 weds airline procurement systems with automatic technology. P. 66.
Air Transport World. February 1988. Product support: Promises, promises. Pp. 25–31.
Bell, S., and B. Kettell. 1983. Foreign Exchange Handbook. Westport, Conn.: Quorum Books.
Bleeke, J., and L. Bryan. 1988. The globalization of financial markets. McKinsey Quarterly (Winter): 17–38.
Booz Allen & Hamilton. Undated. From strategic planning to strategic performance: Closing the achievement gap. Outlook. Number 4.
Business Month. April 1987. Taking Control in the Rag Trade. P. 49.
Business Week. September 28, 1987(a). Can Ford stay on top. Pp. 78–86.
Business Week. November 30, 1987(b). Computers: The new look. Pp. 112–122.
Business Week. December 21, 1987(c). Power retailers. P. 92.
Business Week. January 18, 1988(a). Japan's latest triumph: Hurdling the high yen. Pp. 36–41.
Business Week. February 29, 1988(b). U.S. exporters that aren't American. Pp. 70–71.
Carlzon, J. 1987. Moments of Truth. Cambridge, Mass.: Ballinger Publishing Co.
Davis, S. 1987. Future Perfect. Reading, Mass.: Addison-Wesley.
Distribution. September 1987(a). CLS explores new logistics frontiers. Pp. 52–54.
Distribution. October 1987(b). International logistics management. Pp. 13–20.
The Economist. July 6, 1985. The other dimension: Technology and the City of London. A survey. p. 6.
The Economist. June 27, 1987 (a). British food retailers: Checking out. P. 73.
The Economist. December 5, 1987 (b). Purveyors to all nations. P. 16.
Kutscher, R. 1988. Growth of service employment in the United States. In Technology in Services: Policies for Growth, Trade, and Employment, B. R. Guile and J. B. Quinn, eds. Washington, D.C.: National Academy Press.
Lehnerd, A. P. 1987. Revitalizing the manufacture and design of mature global products. Pp. 49–64 in Technology and Global Industry, B. R. Guile and H. Brooks, eds. Washington, D.C.: National Academy Press.
Managing Automation. October 1987. Ameritech has promises for future manufacturing. Pp. 48–49.
Office of the United States Trade Representative. December 1983. U.S. National Study on Trade in Services. Washington, D.C.
Porter, M. 1985. Competitive Advantage. New York: Macmillan & Co
Roach S. 1988. Technology and the service sector: America's hidden competitive challenge. In Technology in Services: Policies for Growth, Trade, and Employment, B. R. Guile and J. B. Quinn, eds. Washington, D.C.: National Academy Press.
Tanaka, M. 1984. Moving beyond merchandise: Japan's trade in services. Journal of Japanese Trade and Industry 4:12–15.

Tschetter, J. 1987. Producer services industries: Why are they growing so rapidly. Monthly Labor Review (December): 31–40.

UNIPUB. 1984. Measuring Productivity: Trends and Comparisons from the First International Productivity Symposium, Tokyo, Japan, 1983. New York: UNIPUB.

Uttal, B. 1987. Companies that serve you best. Fortune (December 7): 98–116.

Vollman, T. 1986. The effect of zero inventories on cost (just in time). Pp. 141–164 in Cost Accounting for the '90s: The Challenge of Technological Change. Montvale, N.J.: National Association of Accountants.

Pictures and Parts: Delivering an Automated Automotive Parts Catalog

Frederick A. Fellowes and Donald N. Frey

This case study reviews the experiences of Bell & Howell's Publication Systems Division (PSD) from the late 1970s until the present in successfully developing an automated parts catalog lookup system. * It examines the transformation of a microfiche publishing service business in applying new and unique electronic technology to meet its customers' information retrieval requirements.

The case also illustrates the critical success factors of continued and intense focus on customer needs; application of new technology to meet customer needs; persistence in spite of failures of several products; the necessity of active, ongoing corporate support; and, finally, strong project and program management of new technology for PSD.

EARLY HISTORY

One of the world's largest micropublishers, PSD converts thousands of catalogs, manuals, and price lists to microfiche. Organizations responsible for high-volume communications with hundreds of dealerships, branches, and field-support facilities often rely on micropublishing to reduce communications to a manageable size and to controllable costs. By dramatically reducing printing and mailing costs, micropublishing enables large documents to be economically distributed frequently, providing dealers and service operations with more current information. Producing more than 10 million

* Prior to March 11, 1985, the name of the Publication Systems Division was the Micro Photo Division. For purposes of this case study the Division will be referred to as PSD.

microfiche a month—the equivalent of 2.5 billion pages of printed information—PSD's business mission is to continue as the market share and quality leader of technical document republishing.

Parts Catalogs and Service Manuals

Most manufacturers of vehicles and heavy equipment use a mix of paper and microform documentation to support their own or their dealers' repair and parts businesses. Documentation is essential to parts sales and customer satisfaction. Parts sales top $1 billion annually for several automakers, and for many farm equipment dealers, parts and service revenues exceed new product sales.

One set of parts catalogs for General Motors cars requires 16 feet of shelf space and larger dealers may subscribe to 10 or more sets. Each set consists of more than 20,000 pages of illustrations, 15,000 pages of charts, and nearly one-half gigabyte of ASCII text (filling additional tens of thousands of pages). Larger data bases are common—the parts catalogs for Caterpillar contain nearly five times as much documentation.

New editions of a parts catalog are published, on average, every four months and mailed to the dealership. In paper form catalog distribution (mailing) costs are substantial. Between editions, corrections and other changes are sent as bulletins, and the number of bulletins can be very large.

Manufacturers choose to publish their parts catalogs on microfiche rather than paper mainly to save costs. End-users' attitudes range from neutrality to dislike of microfiche as compared with paper (owing to legibility and convenience factors).

The Creation of Parts Catalogs

The process of creating new parts catalog editions evolved simultaneously with data processing. By the early 1970s computerized photocomposition for parts catalogs was popular. Changes to the text were identified and entered for each individual catalog; then photocomposition tapes were written which, in turn, were used to print the catalog. In this form, microfiche versions of the catalog could be produced directly, but the relationships between data elements necessary for an "electronic parts catalog" did not exist.

As data processing techniques for relational data base management spread to publishing in the mid and late 1970s, a parts catalog data base became an important way of lessening the time and labor content required to update a set of catalogs, as well as a way of improving accuracy and consistency. The basic data base element is the parts record containing all applications of the part across all catalogs. A part change or addition, although affecting many catalogs, had to be entered only once. Essential to a commercially viable electronic parts catalog, the existence of a parts catalog data base is referred to as an "automated" catalog.

The processes of creating a publication edition are known as "prepress" activities because they are the steps taken before the printing press makes copies or the micropublisher (such as PSD) generates copies on microform.

Micropublishing Processing

The data processing for parts catalogs is very different from typical management information system operations. Data bases typically contain 100 to 500 megabytes of ASCII data. The data base also includes thousands or tens of thousands of images—a data type foreign to nearly all data processing operations. Even the control and format encoding of text is different—" photocomposition" files have their own unique rules.

Text data are received on reels of standard 1/2-inch, 9-track magnetic tape. The text data are reformatted, resequenced, and indexed. Art is received in three forms: paper, raster data on magnetic tape, and vector data on magnetic tape. Paper copies are photographed, and image data records from magnetic tape are scaled and sequenced. The art and text are then combined and fiche masters are produced. Masters are then used to produce duplicate fiche for distribution. Silver halide-based film duplicating is capital intensive, whereas diazo and vesicular film duplicating are not.

In the 1970s PSD was regarded as the "Cadillac" of micropublishing service bureaus in terms of product quality, service, and indexing. PSD was one of the few service bureaus to use all three microfilm technologies (vesicular, diazo, and silver) to meet customers' technical documentation duplicating needs. Recognized as the best technical document micropublisher, PSD had acquired a substantial market share. Its customers were primarily automobile, construction, agricultural, and computer equipment manufacturers.

The 1632 Project—GM I

In the mid-1970s, through the efforts of Bell & Howell's Document Management Products Division (DMPD), PSD became aware of GM's desire to improve parts catalog lookups. GM was interested in improving its automobile dealerships' ability to retrieve parts catalog information.

John Marken, the division president at the time, recalls:

> The general sales manager at GM wanted a fast automatic frame selection microfilm system. When we learned about the situation from DMPD, a competitor was far ahead of us. They had already proposed a cartridge microfilm system. Dick Miller,*

* Dick Miller, past president of DMPD, started Bell & Howell's parts catalog micropublishing business (when it was part of DMPD) in the 1960s and was well known to GM.

> from DMPD, did some superb selling to slow down the process and provide us with time to develop a concept. In 1976 the GM 1632 project was assigned to PSD from DMPD. We proposed to GM management that we would introduce more automation. We would use a "roll" of microfiche, for the entire GM parts catalog, giving them more information, and a way to get to the image faster than with a multitude of microfilm cartridges. We convinced them of the concept. We built 125 units and did a field test. The result was favorable.

In 1978 the 1632 rollfiche reader concept was presented to GM's Dealer Council in Brownsville, Texas. John Marken remembers dealers' reactions as, "Can't you automate this more? We want to access the information immediately. It wasn't that they were against microfilm—it was just that they wanted even more automation."

In an effort related to the 1632 program, GM planned to create an automated parts catalog database. PSD was awarded a $1 million contract to design the automated parts catalog data base for GM. The project was completed successfully and it generated a small profit.

A meeting was held in 1980 between GM and PSD in Flint, Michigan, to conclude the negotiations for the 1632 rollfiche reader project. At that meeting, GM's management told PSD management that the automotive industry had suffered a serious decline and that no more contracts could be signed. Those close to the project noted that technical problems with the 1632 and the fact that it did not meet users' needs were as critical as the recession to the program's cancellation. Write-offs on the program's cancellation totaled $3 million and total program losses exceeded $5 million. In addition to this, during the early 1980s several events occurred that had a serious impact on PSD's business.

Silver Crisis

The first shock to current business practices was the silver crisis. Silver was the predominant film for parts catalog microfiche systems. Owing to its higher quality, silver fiche could be left on a microfiche reader for extended periods of time with no deterioration. When silver prices soared, PSD's silver film business was dramatically affected. As the prices of silver kept rising, so did silver film prices. The price of silver in August 1979 was $10 per ounce. By December 1979 it had risen to $34 per ounce and reached an all-time high of more than $50 per ounce by January 1980. John Marken recalls, "the silver crisis drove us and our customers to diazo film."

Dave Gump, who was then general manager, remembers, "The profit margin on silver was higher, and we had few competitors in silver microfiche. The profit margins went down significantly with the switch to diazo film."

Don Prince, who is now vice president of operations, was production manager at the time, and notes that with the switch from silver to diazo:

Many companies saw ours as an attractive business and for relatively small capital investments were able to start small diazo microfilm businesses. That's not to say that they could compete with us in terms of our copy preparation and film expertise, but these smaller companies were capable of nibbling away at our business.

The result was that the microfiche service bureau business began to deteriorate into a commodity business.

Search for High Tech

Don Gardner, application development manager, recalls that during this period, a second factor began affecting PSD's business:

> Our customers were increasing the sophistication of their publishing. Industry was waiting anxiously for the birth of office automation. The panacea of the future was in a black box—new tech was coming and our customers believed it was going to make our microfilm outdated. This was making our business extremely difficult. We tried to prepare to meet the technological opportunity.

As Dave Gump recalls, "the business was experiencing declining margins—we were the only source of new business for our competitors. The company had to find a way to change the product or expand the business."

In April 1981 PSD ventured into prepress services by acquiring DDSI, which offered high-speed plotting of computer-generated illustrations, highspeed typesetting, and data base publishing services. PSD's goal was to vertically integrate its publishing services by moving upstream into processes done before micropublishing. Dave Gump's comments on the DDSI acquisition:

> Automated technical publishing (ATP) was a very complex business. Our microfiche customers wanted in-house capability, not a service bureau. We thought the Department of Defense would be an excellent market for ATP. As we found out, there was no incentive for DOD contractors to reduce their labor since they were cost-plus. In September 1982 we closed the California office (losses totaled $800,000) and brought ATP back to our Wooster, Ohio, facility. We are still selling ATP services.
>
> Included in the acquisition was a very high resolution photocomposition system that output paper and microfilm. To utilize the equipment, we learned to create programs to handle customers' photocomposition files (before this, with the exception of GM's photocomp files for the 1632 program, PSD processed only standard data processing files). We became much better at talking with publishing departments. This equipment made a new venture with GM possible.

A FRESH START

Microfiche Catalog—GM II

In February 1981, PSD started development of a standard 105-mm microfiche parts-catalog system designed for small-to-medium-sized GM dealer ships. In 1982, PSD secured a contractual arrangement with GM for the microfiche parts catalog program, because of the savings to GM in switching from paper to fiche—the traditional microfilm benefits. Don Gardner explains:

> We were able to design some user-oriented indices to make it acceptable. Our system design approach has made us unique in this marketplace. We didn't just sell microfiche, we designed information retrieval systems. We even designed and redesigned fiche readers to meet the applications.

The microfiche program was presented to select GM dealerships in Houston in April 1982 (see Figure 1). These dealers indicated that the microfiche system was acceptable, but they strongly preferred an automated solution. They asked for "push button" access to the GM factory parts data; 67 percent said that they would wait for a more automated solution. But 25 percent were willing to switch to fiche from paper—enough for the program to proceed.

John Ramagli, then vice president of sales, recalls:

> In April of 1983, the first GM parts catalog microfiche publication went out to the field. We had reasonable support from the GM dealers. The GM Fiche Program is ongoing today with 1,500 to 2,000 of GM's 10,000 dealers. At the same time, however, GM's management was strongly urging us to do something else.

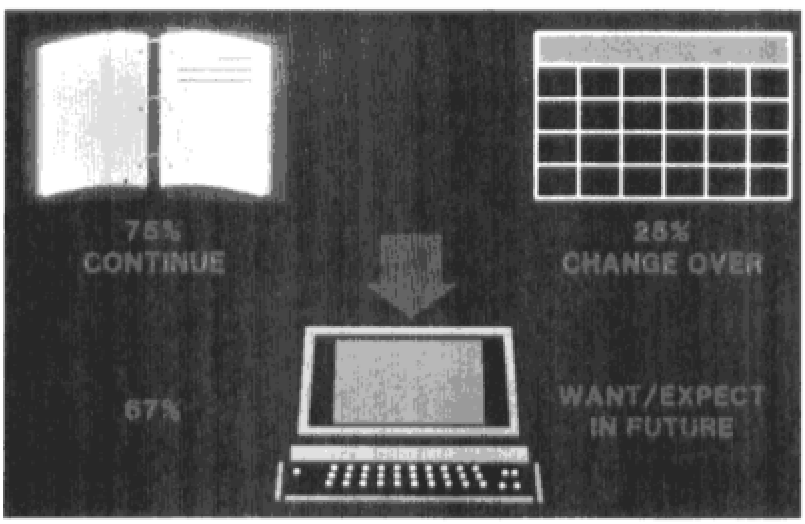

Figure 1 A 1982 presentation to GM dealerships yielded information about dealers' preferences on automation of catalog access.

A Hybrid Solution Pursued—GM III

By this time, PSD had made several attempts to improve its business. Acquisition of DDSI, increased marketing efforts directed at traditional business, and internal productivity incentive programs were all undertaken.

PSD's limited market had matured. New products to speed catalog access were seen as necessary to preserve the customer base and ensure growth. PSD marketing continued to hear parts catalog users demand "automated access." At the same time, PSD's customers were in various stages of automating their parts catalog data bases.

PSD marketing proposed development of a system that would use microfilm for graphics and a PC for text. A hybrid solution, using microfilm for parts illustrations, was proposed rather than an all electronic approach that the market seemed to be seeking because (1) no computer solution that had adequate storage space for all the parts catalog illustrations was available; and (2) the cost per image in microfilm was stunningly cheap compared to the alternatives.

John Marken remembers:

> The standard fiche program was very important in keeping the door open to future more automated solutions for the parts catalog. We kept talking with GM, updating them. Within PSD, John Ramagli developed an idea for a more automated system for GM dealers. The concept was to employ cartridges containing high-reduction images of parts illustrations printed on high-density 35-mm strips. The PC would assist in selecting the correct cartridge, strip and frame as well as providing access to text portions of the catalog.

The concept was presented in April 1983 at the annual Bell & Howell Technology Conference, an interdivisional forum to promote discussion of technology trends and to present new product concepts. John Marken remembers Donald Frey, the chairman of Bell & Howell, suggesting that PSD should pursue an all electronic solution because dealers would not accept a mixed microform/PC system. Dave Gump recalls that it was also at this conference that he first learned about low-cost compact disk read-only memory (CD-ROM) drives. At this point CD-ROM technology was evolving and interest in it was growing.

In response to PSD Marketing's desire to offer a more automated parts catalog system to GM dealers, Don Gardner became involved. He explained:

> In the summer of 1983 Dave Gump called me to discuss Marketing's proposal for a system to improve the productivity of our GM fiche users. In order to prove productivity improvement, I needed to know the current productivity. In July of 1983 we proposed that we would have a survey done. I hired an industrial engineer and designed the measurement methods and in September we proceeded with on-site studies of dealers in the truck/transportation, agricultural, heavy equipment, and automotive industries. The focus was on the parts department and particularly on

> parts catalog lookups. We took into consideration the size of dealerships. We did time studies and analysis of their operations (including profit and loss and management records) to understand the parts business and its impact on the dealership (by industry).

The Electronic Parts Catalog (EPC) Concept—GM IV

The research verified a product need, but not PSD's product concept. Dealers would not realize sufficient benefits to justify a partially automated solution in which a cartridge microfilm reader driven by a PC was priced at $5,000 per workstation. The concept of PC-aided-retrieval of microfilm for parts catalogs would no longer be pursued—dealers insisted on a completely electronic system. The question became: could the likely price of an all electronic parts catalog (EPC) be supported?

Results from this initial study led Don Gardner to recommend that the best opportunity for developing an EPC was GM because (1) the automotive industry represented the best single market segment; (2) within the automotive industry, GM had the largest dealer network; (3) GM had an automated parts database that PSD had designed and implemented; and, (4) PSD had also implemented a catalog fiche system for GM (gaining experience in processing GM's photocomp tapes) and was already familiar with their parts lookup application.

In November 1983 Don Gardner began to concentrate application research on the GM parts catalogs at GM dealerships. Detailed data on their parts lookup activities were recorded. The research verified the need to automate the parts counterperson's catalog lookups.

However, from Don Gardner's research (completed in March 1984) it was determined that a parts counterperson uses the catalog in just under half of all transactions. Actual use of the catalog accounts for approximately 10 percent of a parts counterperson's day. Just over a third of parts catalog lookups involve the illustrations.

Dave Gump notes:

> Reducing the parts department's labor cost for catalog lookups was what we expected would provide the cost justification for an electronic catalog. But when Don discovered that only 10 percent of a parts counterman's time is spent using the catalog we almost killed the project—there could be no way to justify the product. But we also learned that 16 percent of the counterman's time was consumed in transcribing part numbers and entering them in the parts inventory system.

By interfacing an EPC with the parts inventory system nearly all the transcribing and key entry could be eliminated—possibly enough productivity improvement to justify purchasing the product. The design specification was modified to include a hardware interface to dealer computer systems and software to communicate with the parts inventory program.

GM Dealer Business Environment

At a dealership the parts business consists of three components: retail, wholesale, and service. The retail parts business is usually small. The wholesale business is the sale of parts to repair shops, garages, and other dealerships. The service business is parts consumption by the dealer's own service operation.

GM has 50 parts catalogs covering 1976–1987 model cars and light trucks. Each catalog is two catalogs in one: the white pages contain part name, part number, and usage data; the green pages have illustrations and references to the white pages. It generally takes 18 months for a parts counterperson to become proficient at using the parts catalogs. Experienced personnel frequently are very well compensated, but despite that, parts counter personnel turnover is high.

Parts lookup is complicated by the infrequency that individual catalogs are updated and the flood of loose corrections, notes, and bulletins that are received between new editions of the catalogs. The process of inserting the bulletins and finding the notes associated with particular parts often is too cumbersome to actually be done.

It is not unusual for a dealership to return to GM each year more than 5 percent of its ordered parts, usually owing to mistaken part numbers. GM offers a program of credits to encourage dealers to reduce returns.

Nearly half of all GM dealers have computerized inventory control systems, and most of the larger dealers have computerized inventory control, usually as part of a "dealer services" system sold by a data processing equipment value-added reseller. A value-added reseller typically combines industry-specific software it has developed with purchased hardware and sells the package directly to that industry.

The Challenge

PSD was confronted by a major task: to find a technical solution to automating the parts counterpersons' access to the GM parts catalogs in a manner producing sufficient economic payoff. From Don Gardner's research, the dealers had identified what they wanted in an automated system:

1. To access information in 1 second.
2. Illustrations and photographs to be clear and easy to see.
3. To integrate with existing dealer systems and be able to communicate text data to their currently installed inventory management systems.

PSD was determined to find a solution for GM. John Marken explains, "General Motors was the biggest company in our biggest market and they had the biggest data base. We could see the opportunity. We had to hold the focus on that one product."

At this point, the "beta" project team consisted of only Dave Gump and Don Gardner. Don Gardner recalls, "In 1984 we began a year of 'airplane research.'" They traveled around the country learning of the alternatives available in CRT technology, high-volume electronic storage, data base management software, and image scanning.

Dave Gump explains his strategy for acquiring technology:

> We didn't want to replicate what was already there. We wanted to use as many off-the-shelf components as were available. We were also interested in finding solutions to avoid sole source situations.

There were three initial technical issues facing the PSD "beta" project team: data volume, response time, and electronic image display quality. Each of these issues had to be resolved to design an effective electronic lookup system. GM parts catalogs were extremely large paper-based data bases. Each GM car line had several thousand illustrations and a similar number of charts. In addition, a large single car line requires approximately 100 megabytes of storage for text data.

Response time, the second technical issue, was affected by data volume and image resolution. Catalog users insisted on I-second access and display. Figure 2 shows the image data transfer time and the relationship between resolution in dots per inch (DPI) and the seconds to display. As shown, the higher the resolution, the greater the time for image display. Response time considerations excluded standard approaches to networking and data decompression.

The third technical issue, image display quality, centered on the counter-person's desire to see detailed illustration drawings and to easily identify the

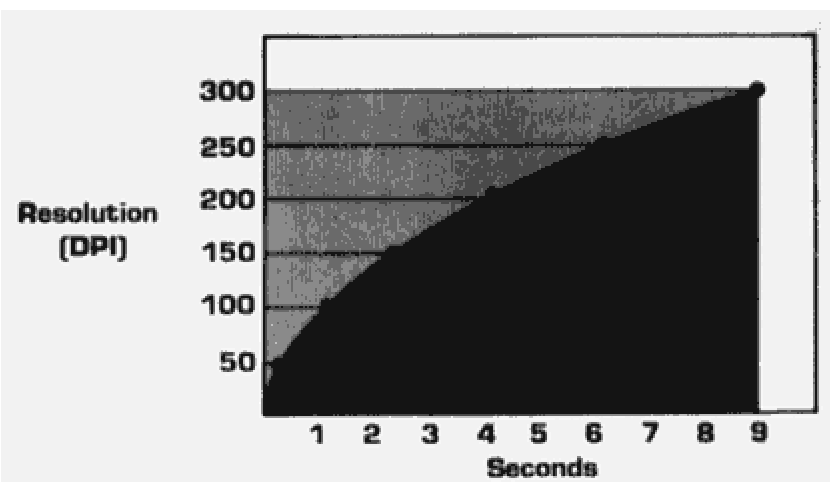

Figure 2 Image data transfer time versus resolution (via RS-232 link).

part call-out number without "pan and zoom." Many illustrations in GM parts catalogs contain very fine, complex detail. Also, because of PSD's expertise in designing microfiche retrieval systems for service applications, the project team was aware that service and repair documentation contained photographs with very small print describing repair procedures (see Figures 3 and 4).

Sources of Information

In search of a technical solution for the "beta" project, Dave Gump and Don Gardner began to investigate and evaluate the options available. They

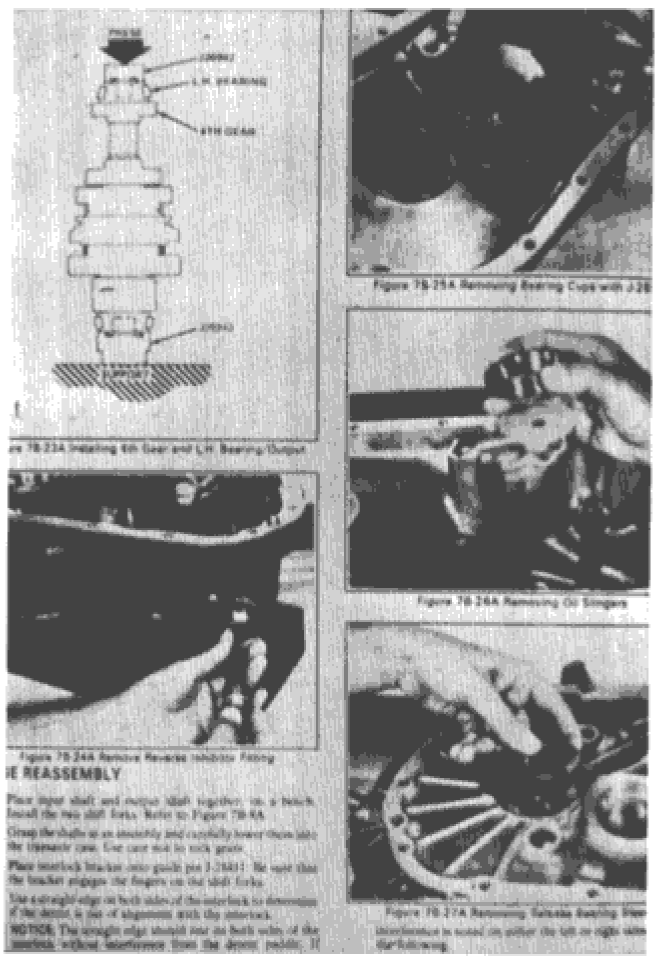

Figure 3 Illustration from GM parts catalog.

carried sample graphics from one electronics company to another to be scanned and displayed. They experimented with a number of 200 dots-per-inch (bitonal) monitors, but the display quality was not acceptable. In early 1984 they met with a consultant affiliated with Massachusetts Institute of Technology (MIT), who was familiar with medical imaging and scanning systems and found a satisfactory gray-scale imaging and display system. Other technology decisions were made. The specific relational data base was selected. Evaluations and decisions on optical disk drives and high-capacity Winchester disk drives were made. Incorporating a touch screen was considered, and a source was selected.

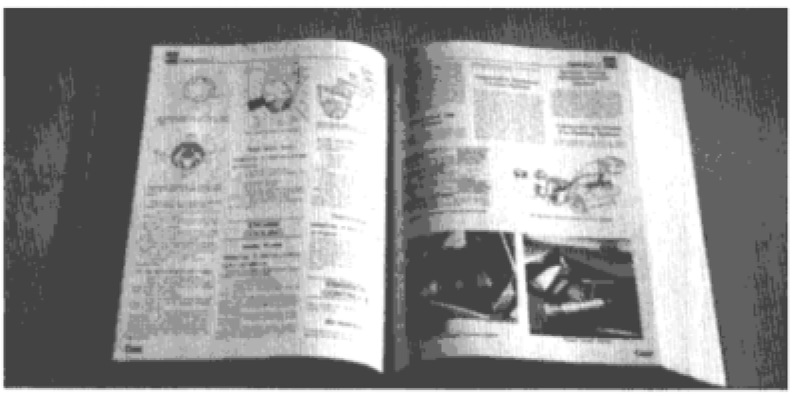

Figure 4 Photograph of service manual.

By the spring of 1984 the team had defined what to do. In April 1984 the project team presented its marketing and technical research along with its product concept at the annual Bell & Howell Technology Conference. It proposed building a prototype system. Subsequent discussions with senior management focused on optical disk as the medium of the future, and thus the direction for "beta" was set.

Dave Gump explains his thoughts on the importance of the Bell & Howell Technology Conferences:

> Not only did these conferences foster the sharing of technical research among the Bell & Howell divisions, they also provided annual reinforcement and corporate support for technological change. The conferences afforded us the opportunity to present new product concepts which could easily have been stifled.

Organization Structure

At the time, Dave Gump was vice president and general manager of PSD. Recognizing that the "beta" development program was going to be as de manding as starting a new business, requiring full-time supervision, Dave recommended an organizational change that John Marken approved. In May 1984 Dave Gump became the vice president of development responsible for the program's research and development on a full-time basis.

Gray Scale

One key technology issue was image display quality and the decision to use gray scale. Dave Gump explains:

> We saw a demonstration that visually confirmed that gray scale solved our image quality problem at an affordable cost. Our contact at a Boston consulting firm explained that there was scientific support for gray scale and referred us to research being done at MIT. This reassured us we were following a reasonable path.
>
> Gray scale versus bitonal was an intense issue among the engineers in different divisions within Bell & Howell. To prove that gray scale was the correct approach to PSD's parts and service applications, we hired an outside consultant, Battelle Institute, to research this issue.
>
> Battelle confirmed that gray scale was the correct approach for our application.

Selection of gray scale (rather than bitonal image data) had the added advantage of requiring only half the storage because Image Data Base (IDB) used fewer lines of resolution. Despite these benefits and user preference for gray scale, opposition to gray scale continued from engineers in divisions that had adopted the facsimile (bitonal) standard for office imaging products.

Prototype Development

Don Gardner recalls the activities in 1984:

> Interspersed throughout 1984 were meetings with GM dealer managers. 1984 was not just a year of identifying hardware, it was a time to solidify the perception of the system . . . redefine how the product would function, try our ideas of what it could do, as well as listen to users' ideas on what they wanted, and then determine what hardware was needed.

The expected benefits and estimated value of the IDB were constantly being refined by market research. In addition to labor savings at the parts counter identified earlier, it became clear that another benefit expected from an electronic parts catalog was reduced part number errors. Error reductions were expected because transcription could be eliminated, information was more easily accessible, and publishing could become more frequent. As a result of fewer part number errors, customer satisfaction would be improved and restocking fees could be much lower. It was also believed that IDB's ease of use would reduce the lost time and productivity of new employees learning the GM parts catalog.

By July 1984 the PSD project team had developed a full proposal on the costs to design and build a prototype system. In August 1984 this proposal was presented to the chairman on his annual visit. The funding necessary for prototype development was estimated to be $425,000. "How could dealers justify purchasing a $50,000 electronic catalog system if a $5,000 1632 reader could not be justified?" This was the central question posed by the chairman. An economic justification based on parts counter productivity gains, reduction in part order error rates, and enhanced customer satisfaction revealed that at $10,000 per workstation, the EPC was justifiable whereas the 1632 was not. The chairman agreed in principle, and remembers that throughout August, PSD management proceeded to push the proposal up through channels for funding approval "with acceptance at best neutral from outside the division."

On September 6, 1984, an outside consultant was contracted to develop the prototype. During the same time, the project team increased to five people. In response to the growing need for prototype hardware development to be carefully supervised, Dave Fehr was hired as manager of Systems Integration in October. Two application programmers were hired to begin the software design.

Throughout the fall and winter of 1984, Dave Fehr supervised "beta" prototype development, including design and coding of publishing software and workstation "retrieval" software. The software engineering effort benefited substantially from PSD's participation (1978–1979) in the development of GM parts catalog data base. The two "beta" programmers were assisted by the original data base designers. The "beta" publishing software was able to incorporate files and code from the 105-mm fiche publishing software for GM parts catalogs. Months of design, surprises, and redesign owing to complexities in the data were avoided as a result of experience from the 1632 and 105-mm fiche programs. By December the prototype system, consisting of a display workstation and scanning system, was up and running in the 1984 GM Cavalier parts catalog.

In January 1985 Don Gardner developed a slide presentation that explained how the system and parts lookup application would work, how it would be used, and what would be delivered. Don explains:

> We had our first monitor, and the first images displayed on the monitor were so good, I was really impressed. We had a photographer develop a series of slides of the display. We took our slide presentation to the GM dealers in Houston—the same dealers who back in 1982 had said they would wait for a fully automated system. We got absolute confirmation that functionally the system was going to do more for them than they had asked.

John Ramagli recalls:

> On February 28, 1985, the first people to see a demonstration of the parts catalog lookup application on the "beta" system were the managers from General Motors. We flew them from Detroit into Wooster to see the prototype.

> At this point all we had on the system was the 1984 Cavalier parts catalog. We showed them a demo of the system. They loved it, and urged that PSD take it to pilot. (GM provided no funding nor even a commitment to allow PSD to publish the catalog data electronically on a commercial basis.)
> In February and March, the prototype system was demonstrated to Chrysler and Ford. Chrysler showed a reasonable level of interest. Ford's interest was modest.

During June and July the prototype (with only the 1984 Cavalier) was demonstrated to GM dealers throughout the country. The PSD project team confirmed that they had a viable product and through research were able to establish their goals—a $10,000 workstation. In August PSD received funding for full development of the project. During the remainder of 1985 the production and software development team built the data base for three GM car lines: Chevrolet, Oldsmobile, and Light Truck. Marketing identified the activities that would be required to bring the product to market, and the "beta" product was named the Image Data Base 2000 (IDB2000). In addition to development activities on the GM parts catalog application, during 1985 research projects on GM dealers' service and repair procedures and Chrysler dealers' parts operations were started.

Field Tests

In January 1986 the IDB2000 prototype system was demonstrated at the National Automotive Dealers Association (NADA) trade show in the booth of a parts inventory system vendor specializing in very large "mega" dealers. PSD's Marketing team successfully solicited pilot dealerships and support from GM management. Several small companies demonstrated videodisk-based systems with sample pages from parts catalogs. The image quality was poor; vital reference numbers ("call outs") were illegible. Computerized microfilm systems for parts catalogs were shown by a major vendor of data processing systems to dealers and by others. The strongly positive reaction to IDB convinced PSD to further accelerate the IDB program.

In March PSD and GM management held a review meeting to discuss the criteria for a successful field test and the strategic implications of a good pilot program. Lustine, the largest GM parts dealership, and Hutton Chevrolet, the second largest, were selected to be part of the pilot program. GM requested that the pilot be expanded to include a medium-sized dealership, Bell Dean, and that a system also be installed at the GM Technical Center.

In April field tests of the IDB2000 parts catalog application began. The first field test location had only three workstations, and yet the users complained about the system's response time. Internal testing with many more workstations (large customers would need 15 to 25) revealed that communications completely bogged down in the CPU, largely because of the combined overhead and interaction of the network software, the relational data base, and the operating system. It was a major setback. In June after much consideration, the PSD development team determined that IDB2000 design changes would be necessary. Dave Gump explains:

> The number of users that we had to support and the performance requirements indicated that it would require a very large computer. Our target price couldn't support that size computer. We chose to put the application into the user-level computer, eliminate the CPU, and switch to a network server for data base storage.

John Ramagli remembers his discussions with GM management about the design revision, which had been named the "Jade" project:

> It was a crisis. At this point we had growing concerns about the speed implications of using the MicroVAX II computer to run the application on a multi-user workstation system. I met with the director at General Motors to explain to him that our decision to design and build our own hardware would significantly reduce the cost and increase the performance.

GM agreed to have PSD continue with the pilot tests, and, as part of the program, PSD agreed to replace each of the pilot sites with Jade systems. In July 1986 the design revision was funded and the Jade project began. The first Jade system was available in October and the system was installed at the Lustine dealership. Lustine's management was so pleased with its improved performance that it offered to buy the system. On November 19, 1986, after completing his pilot test research, Don Gardner returned from Lustine with the first purchase order.

As Jade systems were installed at the remaining GM pilot sites, the marketing group was developing the product introduction plans for the IDB2000 EPC and working with GM to complete the negotiations for rights to the parts catalog data base. On December 18 the contract providing Bell & Howell with rights to the data base was approved by GM. PSD worked with several units of GM including, for data processing, Electronic Data Systems, a billion dollar provider of data processing services and automation acquired by GM in October 1984. The diverse interests of each unit made securing a commitment from GM an exacting process.

During the field test many recommendations were made by the users for improvements in the application software, but of even more significance was how the system was used. One dealer placed a workstation in the service area (rather than placing all at the parts counter), and discovered that mechanics could find their own parts and have the list of requested parts printed out in the stock room. Each mechanic was thus able to increase billable time an average of 5 hours per week. The additional service work provided more financial justification than parts counter labor savings, reduced part number errors, or any other quantified factor. Table 1 summarizes the measured productivity improvements owing to IDB installation.

Although none of the four dealers that field tested the Jade system was interfaced to inventory control systems, all four purchased systems. It appeared that interfacing to the inventory control system was desirable but not required by the market. Lack of cooperation from the two major inventory control system vendors was likely to turn into opposition and competition at product launch.

TABLE 1 Benefits of Image Data Base for Dealers

MEASURED PRODUCTIVITY IMPROVEMENTS
Wholesale parts orders/day: Up 20% after image data base installed.
Service revenues: 10% increase if mechanics order parts; 2% increase otherwise.
Parts returns: Reduced from 7 to 6% without emulation. Reduced from 7 to 4% with emulation (dollar-for-dollar credit for reductions paid by GM to dealer).
DEALER CHARACTERISTICS[a]
Annual wholesale parts: $1,940,000 (gross margin 16%)
Annual parts sales to service: $669,327 (gross margin 22%)
Annual service labor revenues: $2,430,000 (gross margin 54%)

[a] Prospect for 10-workstation ($100K plus 20% per year operating costs) system. A dealer of half the size would need a 5-workstation system. This "ideal" dealer is evenly split between wholesale parts and service; most divisions are skewed in favor of one or the other.

Product Launch

Early in 1987, PSD's Marketing and Sales Management was completing the plans for the IDB2000 product introduction and adding the marketing and sales staff required to successfully launch and actively sell the system to GM dealers. A training department was established and customer support specialists were hired to be prepared to train IDB2000 customers. Sales managers and a national sales force were hired and trained. The Bell & Howell national service force was trained on installation and repair procedures for the IDB2000.

In February 1987 the IDB2000 EPC was introduced at the NADA trade show. Figure 5 shows the terminal and the display quality in the commercially available EPC. The IDB was in five booths: GM, Chrysler, and three vendors of dealer data processing equipment. The IDB and competitive EPCs drew large crowds. The two largest data processing equipment vendors and IBM demonstrated PC-based EPCs that did not use gray scale. Several publishers that were developing CD-ROM products for libraries presented parts catalog concepts at other shows the next month. Computerized film retrieval and videodisk parts catalogs were absent.

In March sales of the IDB2000 began, and during this month the first "nonpilot" system was sold. Although no competitive electronic GM catalogs were on the market, a vendor of data processing equipment to dealers was field testing an Acura EPC with Honda dealers. The two largest data processing equipment vendors were advising dealers not to buy the IDB, but sales continued to be strong—surpassing one large system sold per day within 3 months of launch. During May the Chrysler Parts and Service Technical Information System pilot program began that included pricing, supersedence, and service bulletin data bases not included in the GM application.

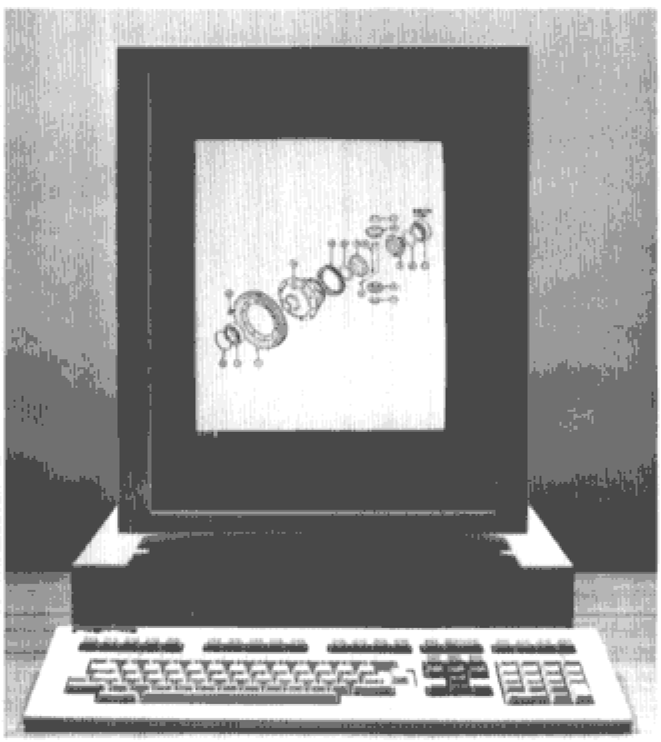

Figure 5 Terminal for IDB2000 electronic parts catalog.

By the March launch PSD had spent $11 million on development (both engineering and marketing) of the IDB System. When added to the 1632 project, the loss on DDSI, and the 35-mm cartridge film/PC study, total investment approached $18 million. This was also the amount budgeted for IDB Sales in its first year and nearly the annual sales for the entire division in past years for micropublishing.

From the point of launch, major cost reduction redesigns were under way to expand the target market to include medium-and smaller-sized dealerships.

What brought the IDB2000 this far? Dave Gump comments:

First, the IDB2000 is really an extension of current business. We built on what we

> know. Our forte is application expertise and the design of information retrieval systems for technical documentation users. We have combined technology in an innovative way to meet our customers' needs. Bell & Howell Company has always had a desire and an attitude to move the company technologically forward.

John Marken attributes IDB2000's success to perseverance. He recalls:

> We could have easily walked away from pursuing "beta" but we were determined, and the risk was worth the effort. Back as far as the late 1970s, the dealers knew what they wanted. The technology just didn't exist to provide them with fast access to pictorial part information at the right price.
>
> We have been successful because we have been able to combine talent and knowledge and have always been focused on what our customers wanted. It took guts and sponsorship. Everybody we reported to wasn't necessarily in favor of it. But, we had a sense of customer need and corporate support.
>
> By dwelling on the needs of our customers and being persistent in our search for a solution we have met with success in the IDB2000. At all times our strength was our focus on the application, the dealers' needs, and the ability to include the right technology to meet their needs. It is no question that we were living on the forward edge of technology and continue to do so.

April to October 1987

During the spring and summer of 1987 a steady stream of software changes that improved access, saved storage space, or made system usage easier (particularly updates) were successfully released to users. The fine-tuned application software incorporated an understanding of GM's data and its use that PSD had gained over the past decade. Also, a substantially lower cost "pedestal" was developed to replace the data storage "rack." The new device allowed affordable pricing of smaller systems (two to four workstations).

The major dealer computer systems (DCS) vendors' efforts to discourage IDB sales became more effective over time. By midsummer IDB sales plateaued as the DCS vendors promised full integration of their future EPCs to existing packages (such as inventory control packages), and demand for interfacing the IDB grew. In August a nonexclusive, sole distributorship agreement was reached between PSD and the largest DCS vendor, Reynolds & Reynolds. The second largest DCS vendor announced a competitive product and acquired access to the GM parts catalog data.

PSD developed simple interfaces to the more popular DCS packages. But without support from the DCS vendor the simple interfaces were awkward at best and often unworkable. In response PSD developed "emulation" software. This software allowed the IDB2000 workstations, when interconnected to the DCS system, to behave exactly as the DCS vendor's terminals did (i.e., emulation). By October the emulation package had become very popular and was displacing DCS terminals in dealership parts departments. Emulation was expected to create more value to the dealers than simple interfacing. In addition to increased speed and accuracy—no keypunch or transcription errors—provided by simple interfacing, emulation was expected to be more convenient and save counter space, equipment cost, and time spent moving from IDB workstation to DCS terminal. By November nearly all IDB System sales included purchase of emulation software.

At the end of October 1987, PSD's fiscal years revenues were forecast to achieve 100 percent of budget and 1988 IDB revenues were budgeted at approximately twice those of 1987.

CONCLUSIONS

This case chronicles the efforts of Bell & Howell PSD to meet its customers' desire for an "automated parts catalog lookup system." It describes the transformation of a micropublishing service bureau through the search for a technical solution to meet its customers' unique information retrieval requirements.

The EPC system developed is a back-room system for GM auto dealers. It is one part of the dealer's strategic information system. Its value is largely owing to its effectiveness rather than its efficiency. Originally envisioned to increase efficiency by reducing parts catalog lookup time, and then to reduce part order errors, the EPC has been found by dealers to increase service billings (labor and parts) up to 15 percent—providing for more profit opportunity than possible from cost savings.

This case presents the trends and issues affecting the micropublishing industry as far back as the 1970s, and identifies specific events that led management to pursue the development of the IDB2000. Five management and policy issues were critical to the success of the IDB2000 and responsible for PSD's ability to achieve technological change.

The first and most important issue is the division's constant focus on its customers' needs. Study of users continued through the field test and into product launch, revealing new insights well after the formal studies ended. Efforts by PSD to understand the parts business and its impact on dealerships provided confirmation of the product need and a clear understanding of the user benefits to be achieved through "automation." By performing detailed time studies, PSD gained a complete understanding of the parts counterperson's daily use of the paper parts catalog. This provided the foundation for PSD to develop the parts lookup application based on how the catalog user wanted to access part description, part number, and part illustration information. Ongoing customer surveys, focus groups, and one-on-one interviews were responsible for the achievement of a system that has received exceptionally positive dealer responses.

The second central issue is PSD's willingness to apply new technologies in developing a solution to market needs. The task of evolving from microfilm to electronics technologies was recognized as difficult and risky, but the alternative was a shrinking business. With a knowledge of the user's requirements for information retrieval, the PSD project team was able to identify the technical issues essential to providing dealers with a fully electronic solution. The team's dedication to creating a solution the market desired, including 1-second image access and display, as well as high-quality display of parts illustrations, resulted in the pursuit of system components (including infrared touch screens, gray-scale monitors, optical-disk storage devices, and networking) that were on the leading edge of technology.

The third factor was management's persistence for 10 years from the ill-fated 1632 project to the successful launch of the IDB2000. Despite several failures and mounting costs, the convictions of a few people carried the program forward. Dedication to the final goal was coupled with an open mind toward implementation. Engineering plans and the perceived value to the user both changed substantially more than once. Management was dedicated to success but flexible enough to change quickly in response to constant market feedback and engineering experience.

The fourth factor in the success of the IDB2000 project was the corporate desire to move the company forward. A cultural direction favoring change was emerging. Sponsorship of annual interdivisional Technology Conferences provided a forum to promote the sharing of technology trends and presentation of new product concepts. Through attendance at these conferences, PSD's management gained valuable information about new technologies, was also able to present its product ideas, and gained ongoing corporate support and funding for the IDB2000 project.

Last, strong project and program management was critical to the successful development of an EPC. PSD had little experience with developing electronic equipment, and the technical resource requirements were many times larger than previous division staffing. PSD reorganized for a program of these proportions, locating the entire team in nearby but separate facilities. Classic project management processes were implemented and refined to specific needs. As a result, the program manager could more effectively manage and control PSD's efforts to develop an EPC while effectively and credibly reporting progress and problems to management outside PSD.

Custodial Package Tracking at Federal Express

Carl Nehls

Federal Express is a true American success story. Started a mere 15 years ago, it created the door-to-door, overnight express package delivery business and forever changed the way the United States and the world do business.

Founded by Fred Smith in 1973 with a fleet of 14 Falcon jets, Federal Express had grown, by 1987, into an international company with $3.2 billion in yearly revenues and more than 50,000 employees worldwide. Federal Express picks up and delivers more than 850,000 packages each day using a fleet of 60 Boeing 727s, 20 DC-10s, and approximately 100 Cessna 208 aircraft.

Federal Express's phenomenal growth, and the resulting articles, business seminars, and case studies have made Fred Smith a famous man. However, a frequently overlooked "key element" to Federal Express's success was Fred Smith's realization that being able to provide information about a package's location, status, and movement was just as important as actually picking up and delivering the package. If the package is so important to the customer that he or she needs it delivered overnight, the customer is almost certain to want to know that everything has gone according to plan.

Federal Express's information management and positive package tracking systems were started early and quickly grew in importance as thousands of people called to have their packages picked up and many called again the next day to ask, "Has my package been delivered?" COSMOS IIB is the latest in a series of information system developments that Federal Express has undertaken to add value to its product while simultaneously improving its operations.

BACKGROUND

The core of Federal Express's information system is COSMOS (Customers, Operations, and Services Master On-line System). Originally developed in 1977, the system was rewritten to run on IBM computers in 1979. These mainframe computers located in the Memphis, Tennessee, corporate headquarters were used to store customer information and rate tables. Call Center agents accessed this information when customers called to request a pickup, and the requests were sent to a printer in the appropriate station. Initially, couriers used pay telephones to call in and receive their dispatches; however, radios were soon installed in the Federal Express vans for quicker and easier communications.

During the next several years many more functions were added to COSMOS to facilitate the collection and distribution of customer, ZIP Code, and operations information. By the late 1970s software applications totaling nearly I million lines of code were running on COSMOS. The network, comprising a cluster of IBM 3090-class machines with thousands of terminals located throughout the world, continues to be one of the largest IBM networks in existence.

Additional computer systems such as FAMIS (Field Activities Management Information System) and ORBIT (On-line Revenue Billing & Invoicing Technology) were developed to handle the company's payroll reporting, field reporting, and billing functions. Although these systems received limited amounts of information from COSMOS, each system remained essentially independent. Some standard information, such as ZIP Code and location data, was maintained separately on each system, and in some cases significantly different conventions were followed. For example, COSMOS kept track of stations by the three-letter abbreviations for the local airport (MIA for Miami, Florida), whereas FAMIS used a number to identify a location. Keeping these multiple systems synchronized was a constant problem and would later pose a systems integration problem for COSMOS IIB.

Providing information to customers about their packages was becoming a significant challenge as the package volume grew. Exception packages—those held for pickup by the recipient, those delayed because of weather or other problems—had information key entered into COSMOS. However, if a customer wished to know the status of a nonexception package, the station had to be contacted by telephone or printer message, at which time local employees researched the package and phoned the information back.

By 1977 bar-code technology was emerging as a reliable way of capturing information in a very efficient manner. Federal Express began to investigate the idea of bar coding the package number on each airbill. The concept was to record the number of each package each time it changed hands. If that information was provided to COSMOS, a customer could be told the exact status of his or her package while on the telephone. To further improve the tracking system, additional information could be collected for each package at certain points. For example, as the package is being delivered, the name of the person signing for the package could be entered.

By early 1979 Federal Express was outstripping the supply of technical people in Memphis and elected to construct a Technology Center in Colorado Springs to help attract talented people to the company. One of the first people in Colorado Springs was Harry Dalton, who had spent his early days with the company as a station manager in Philadelphia and later as a district director. Dalton, then a director reporting directly to the president, was sent to Colorado Springs and matrixed to the Advanced Systems Group (that is, he was assigned to the group, but he did not report through the group's chain of command). Dalton explained his unusual placement in Colorado Springs as a desire of executive management to "inject a dose of reality into the bits and bytes group." In mid-1979 Dalton was given the go-ahead to conduct a field test of the available bar-code scanning equipment.

Federal Express's first survey of the hand-held terminal technology market revealed a small selection of available hardware. Most of the devices had very limited keyboards for data entry and were capable of scanning just a few types of bar code. Norand Corporation was just preparing to release its first programmable terminal with a numeric keypad (see Figure 1). Federal Express acquired some of the first terminals Norand could make available and began testing these models as well as terminals from other vendors.

The plan, dubbed COSMOS II, was to supply each courier with a scanning device for use in the station and on the road. The important tests were to determine if hand-held devices were durable enough to use on-road, to see whether large quantities of sequential numbers could be printed in bar code, to determine the impact of scanning on field people's morale, and to provide data that could be used to predict the impact of package scanning on the entire company. The scanners would be used to scan the package-tracking number of each package every time it changed hands from the point of pickup to the point of delivery. The information collected by these scans would be transmitted from bases in the station to COSMOS, where a customer service agent could access the information from a terminal.

These field tests showed that scanning the packages as they changed hands was a feasible plan; however, the equipment was not sturdy enough to withstand the rigors of a delivery route. As a compromise the project was broken into two parts, COSMOS IIA and COSMOS IIB. COSMOS IIA would only involve scanning packages in the stations where the equipment could survive. More durable equipment had to be found before the on-road scans of COSMOS IIB could be implemented. This type of compromise—not allowing "best" to get in the way of "better"—was to become a common approach to many technical challenges at Federal Express.

Figure 1 Norand 101XL.

The equipment selected for the in-station scans was a Norand device that was modified to incorporate an alphabetic keyboard. This special device, later adopted as a Norand standard, was referred to as an HHT, or hand-held terminal (see Figure 2). The software for this unit was specified by Federal Express, written by Norand, and stored in PROMs (Programmable Read Only Memory).

In early 1981 COSMOS IIA implementation was begun. A group of six people crisscrossed the country installing 450 HHTs and training personnel in 150 stations on how to scan the 100,000 nightly packages. By the summer of 1981 Dalton had moved to a director's position responsible for Corporate Industrial Engineering in Memphis, and began leveraging technology applications through this department. Eventually, most of the work with bar codes and scanning equipment was shifted to others in the group. Federal Express engineers, particularly David Dietzel and Jim Turpin, continued to watch for and test new scanning equipment as it hit the market, searching for the device that could support COSMOS IIB.

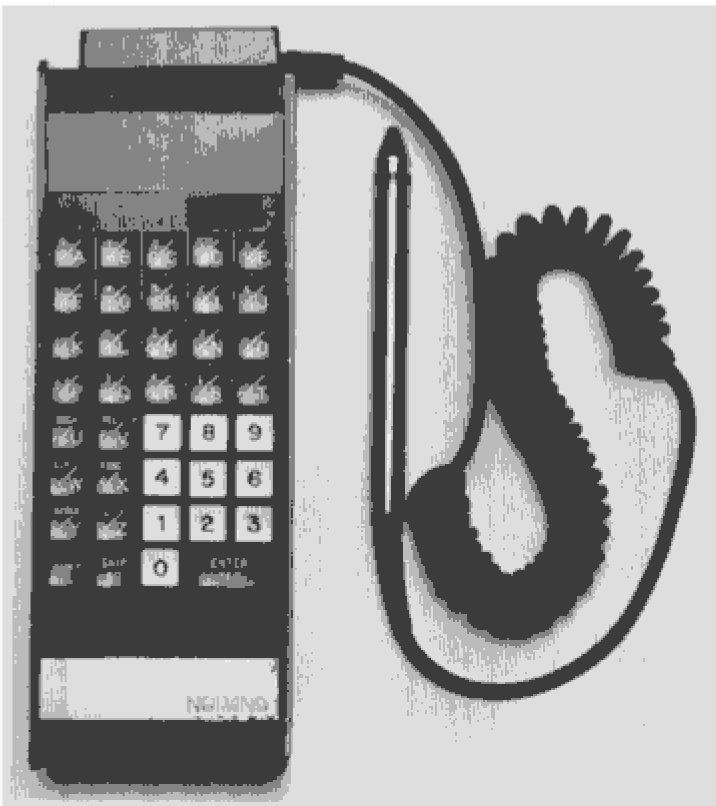

Figure 2 Norand 101XL Alpha 1 (HHT).

The reaction of station management to the implementation of COSMOS IIA was often less than favorable. Extra personnel were required to perform the scanning, and station personnel thought that the scanning hampered the speed of the morning operations. After several months of operations it became apparent many stations were not scanning as diligently as had been hoped. Soon compliance reports comparing the number of scans done by each station to their package delivery counts were developed. Not long after that, the stations' scan compliance was included in the MBO (Management by Objective) system for the station managers, thus supplying even more incentive to make sure the scans were done properly.

Not all the plans for COSMOS IIA were made perfectly. Some of the scans in the HHT were found to be impractical, and the need for new scans soon developed. It was frustrating to Dalton's engineers that the software in the HHTs could not be easily adjusted to meet the changing needs of the field operations. Software changes to the HHT were difficult because not only did Norand have to be contracted to make the changes, but there were also significant logistical problems in locating all the terminals and performing a PROM swap without disrupting operations.

Despite these problems and frustrations, COSMOS IIA scanning became a natural part of station operations and the core of Federal Express's package tracking and tracing operations. The lessons learned in COSMOS IIA would play a major role in the development of COSMOS IIB.

By 1979 daily pickup operations in some cities had increased to the point that 2 to 3 dispatchers were issuing nearly 3,000 requests for pickups to more than 100 couriers in the larger locations. Many of these dispatches occurred just before local businesses began to close, which made the dispatch function almost impossible to manage. Because radio frequencies were tightly regulated and in short supply, it was impractical to use multiple radio channels. Instead, Federal Express went to the marketplace and had technology created in the form of DADS (Digitally Assisted Dispatch System). The first DADS system was implemented in mid-1982.

DADS was developed with a Canadian company, Mobile Data International, and involved having COSMOS send the request for pickup to a computer in the station instead of to a printer. The dispatcher would confirm which route should receive the request and would then forward the request to a computer terminal mounted in the van (see Figure 3). Sending these short bursts of data instead of conversation on the radio dramatically reduced the load on the radio channel and provided a real time link between the courier and Federal Express's computer system.

The ability to pass digitized data between the van and COSMOS opened up an entirely new field in data communications. Dalton and Dietzel recognized that this real-time link to the computer system would allow scan information to be immediately transmitted to COSMOS rather than waiting until the courier returned to the station. Making the information available to the customer service agent in a near real-time fashion would eliminate even more traces. For example, a trace request at 10:00 A.M. without DADS would still require a call to the station, because couriers would still be on their delivery routes. With the real-time communications of DADS available, the information of a 9:55 A.M. delivery could be provided at 10:00 A.M. and prevent the call to the station.

In the early months of 1982 efforts intensified to find a COSMOS IIB device. By late 1982 it had become quite clear that the manufacturers of barcode scanning equipment did not have plans to develop what the engineers of Corporate Industrial Engineering were looking for. Consequently, a Request for Proposal (RFP) was issued in the spring of 1983 for a device with no keyboard and 4K of RAM (Random Access Memory). Five thousand devices were to be produced, and they were to be accompanied by charging and communications boots. Six companies responded to the RFP. Most of the companies, already established in the scanning equipment industry, responded with proposals that used part of their current line of equipment. One exception to this was a capable start-up company in North Carolina named Hand Held Products (HHP).

Figure 3 DADS Mobile terminal. The "shoe" for the scanner, shown in this photograph, was added after DADS implementation.

During one of the RFP review sessions with HHP, they were told that they must bring back exactly what was requested. Turpin held up a large felt-tipped pen and said the scanner must "look and feel like this." Turpin explained to HHP that (1) the device must read Codabar, (2) the device must fit in a shirt pocket, and (3) HHP must develop a unit for presentation to Harry Dalton and it must work. Representatives from HHP visited Memphis 3 months after the RFP went out and had with them a working prototype. This device was nicknamed "tent peg," because of its shape. HHP had been a dark horse, but suddenly the company had made itself a front runner.

COSMOS IIB

1974–1983 Summary

Harry Dalton and his engineers had finally found a vendor that could and would provide a scanning device the courier could carry in his or her pocket and use to scan packages on the road. The compromise solution of only scanning packages as they arrived at and departed from the station had become standard operating procedure. The ultimate goal, first envisioned in the mid-1970s, of collecting tracking information at each change of possession was still there. By 1983, however, the objectives of COSMOS IIB had been expanded. Not only was tracking required every step of the way, but the information was desired instantly and had to be managed to identify packages that did not appear at the next checkpoint. The DADS system was to be used to transmit the data to COSMOS so that the information was available to customer service agents within seconds. Computers in the stations would generate reports of packages that did not appear at the correct destination, thereby enabling the packages to be recovered and delivered on time.

Initial Financing

Harry Dalton had been selling the project on and off over the years, always looking for ways to enhance the project's value. Many of the benefits were related to the courier force that would be using the device, but Dalton also canvassed the rest of the company to identify other benefits. Marketing, sales, and customer service departments identified savings and profit potential that significantly added to the benefits that could be obtained from labor savings, but most of the benefits still were not considered solid enough for inclusion in a financial justification. Finally, after reviewing the entire project plan in May 1984, senior management saw enough benefits to approve development and test funds. Preliminary implementation plans called for 13,000 scanners to be rolled out by June 1985.

Initial Hardware Development

The Memphis-based engineering group spent quite a lot of time working with HHP to develop the ultimate scanning device while a search for a station processor for generating COSMOS IIB reports was centered in Colorado Springs. Dealing with a start-up company such as Hand Held Products brought with it a unique set of circumstances. As a small, aspiring company it carried the vigor of youth, and the energies its employees poured out produced speedy results and changes. The requirements for the COSMOS IIB device quickly evolved from a simple "tent-peg" scanner to a scanner with a display (see Figure 4), and then to a more advanced unit with the ability to make some elementary keyboard entries (see Figure 5). Working prototypes of all these units were produced by HHP in less than 6 months. By July 1984 a

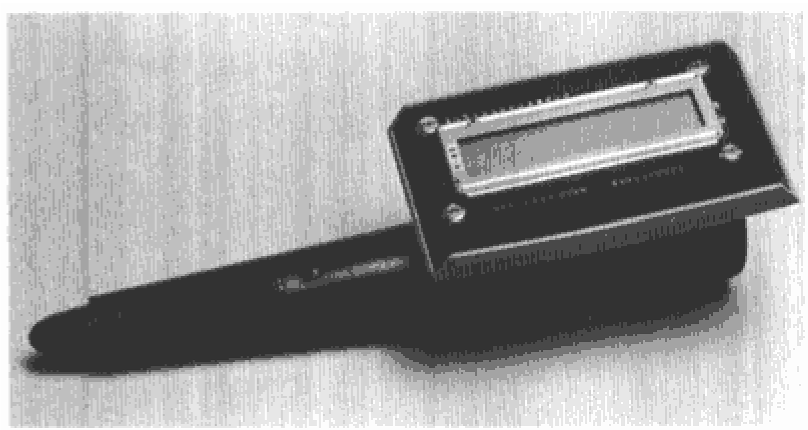

Figure 4 Hand Held Products Prototype II.

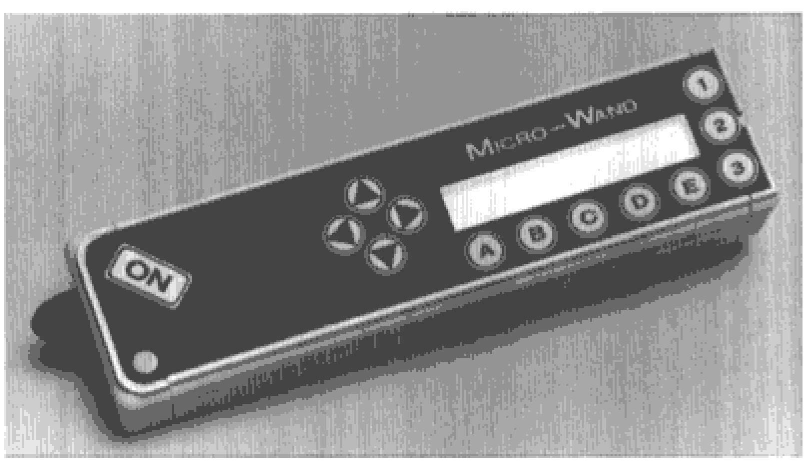

Figure 5 Micro-Wand I (Hand Held Products).

decision was made to manufacture a scanner with a full alphanumeric keyboard, a two-line display, and 64K of memory (see Figure 6). Federal Express eventually purchased 250 of these units for field testing. The development and production of these units were expected to take nearly a year.

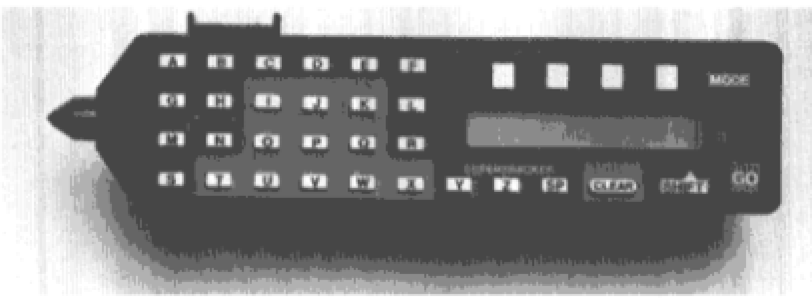

Figure 6 SuperTracker (Hand Held Products Micro-World II), a subsequent model of the Tracker.

Although developing a new piece of equipment had its advantages, it also had its disadvantages. For example, part of the unique design for the Federal Express unit, appropriately named Tracker, was an 8-bit surface-mount processor that had just recently been introduced. The COSMOS IIB project required some of the first batch of these chips to meet its development schedule. Special arrangements had to be made with the manufacturer to secure the order to ensure that the development schedule could be maintained.

In Colorado Springs, bench tests for the station processor had been developed and computers from about a dozen vendors were evaluated. Each station was to receive a processor. All the scan data would be fed through this unit either through DADS, a communications and charging base in the station, or modems connected to telephones that the couriers could call when transmitting their data acoustically. The data would be stored locally for report generation and then, if appropriate, sent to COSMOS for use by the customer service agents. After much evaluation, the newly developed MicroVAX from Digital Equipment Corporation was selected based on its functionality and modularity. David Dietzel was named manager, and Jim Turpin was also promoted at this time and managed a hardware design and development group that became instrumental in the project's success.

Initial Software Development

With the hardware specification complete in July 1984, the staff in Colorado Springs began to write the software for the Tracker, the Station Pro cessor, and COSMOS that would make the system work. Because no production Trackers were available to develop software on, a desktop computer was connected to a prototype Micro-Wand I and used to emulate a Tracker. The initial programs for the Tracker were very straightforward; the courier was forced to answer the same set of questions no matter what type of situation was being documented; no decisional branching was provided. All the software for the Tracker was written in assembler and stored in EPROMs (Erasable Programmable Read Only Memory). The first version of software covering a total of 10 different types of scans required about 24K of the 32K EPROM. The data actually collected by the courier were stored in 32K of RAM.

Programs to extract the information from the Trackers and produce reports were written for the station processor. These programs were written based on the specifications for the Tracker software, but none of the programs could be tested in anything close to a production environment until the first hand-built Trackers were provided in March 1985.

Project Management

Through the end of 1984 all project work for COSMOS IIB had resided entirely within the departments of Harry Dalton and Jim Tollefson, a director in Colorado Springs. As the dates for field testing and, hopefully, implementation grew near, it was apparent that COSMOS IIB would require the involvement of nearly every area of Federal Express to make it a successful project. A meeting was called in January 1985 and representatives from every department in the company were invited; nearly 150 people attended. A briefing of COSMOS IIB was provided and task force committees were formed to cover Field Test, Training, Policies & Procedures, Communications/Video, Software Design & Support, Logistics & Distribution, and Finance.

Communications

Widespread communication within Federal Express became a major element of COSMOS IIB, particularly between Memphis and the field personnel. Regular meetings were held by each task force and once a month the leaders of the task forces would gather to share their progress. Distribution lists on the minutes of meetings grew each month as more and more people became involved.

Articles were written for the employee newspaper every month and the monthly employee and manager's videos almost always had a segment describing the operation, progress, and benefits of COSMOS IIB. A special video showing how COSMOS IIB integrated with the many systems at Federal Express was videotaped using a wooden model of the Tracker.

Any group within the company could obtain a presentation on COSMOS IIB merely by phoning in a request. Almost everyone on the project team made presentations somewhere along the way, but presentations were viewed as such an important function that one individual reporting directly to Dalton undertook them as a major job responsibility.

Special attention was given to communications with the field, because the company believes that it is the commitment of the people actually performing the scanning that determines whether or not the project will be a success. The couriers' attitude and willingness to make a project such as COSMOS IIB succeed is often based on their understanding of the value of the project and how they perceive that their local management views the project. To ensure that the interests of the field personnel were properly represented and that the information about the project was properly communicated to field personnel, the ground operations support staff in Memphis was integrated into the development process of COSMOS IIB, and a director of this group was appointed to represent the interests of the field in general. Executive ground operations management, field engineering groups, and local management of test sites were included as much as possible in the ongoing plans for COSMOS IIB, and several trips were made to each test site to provide pretest briefings.

The field tests also provided valuable communication opportunities. Videos were made during the tests showing the couriers in action, and the comments of the couriers, both positive and negative, were widely publicized.

Testing

Production Trackers became available in April 1985 and the first few units were sent to Search Technology in Atlanta, Georgia, for thorough environmental testing. Additional units, schematics, and software were sent to Georgia Tech Research Institute for an engineering evaluation. Both of these organizations discovered problems with the unit and recommended solutions. It became increasingly obvious at this time that if Federal Express was to be successful in any leading-edge use of technology, it had to become involved in the hardware design. Engineers under Turpin developed, with the help of Dr. Mike Maddox of Search Technology, detailed environmental and operational tests. These tests included heat, humidity, vibration, and drops to concrete from as high as 7 feet.

Hand Held Products made the required changes and the first units were accepted by Federal Express. Within 2 weeks, in June 1985, couriers were being trained and the field tests had begun. These initial tests looked only at the hardware and software that could be operated within the local station environment; COSMOS was considered a much too important system to jeopardize by connecting it with equipment and software that had never been formally tested on the company's quality-assurance systems. Field tests were held in five locations during the next several months and two of those locations continued to use the Trackers even after the tests were complete.

A variety of problems were identified and corrected during the 6 months after field tests were begun. Keyboards on the Trackers came loose and the keys stopped functioning, causing great consternation in the couriers. The charging bases had never been tested with more than a few Trackers, and it turned out the power supply was not charging the batteries properly. Consequently, the Trackers died on the road and, because the charging bases could not be readily fixed, the test teams had to buy all available supplies of 9-volt batteries from local stores to keep the test going. Because the specified material for the battery terminals would have delayed the test by weeks or months, a substitute material was used which lost its springiness, thus allowing the batteries to break contact occasionally. This intermittent loss of power caused strange things to happen to the Trackers, each of which looked like a separate problem. The field environment proved even tougher than the laboratory, with Trackers being dropped repeatedly, being slammed by doors, and even being run over by a vehicle.

Even more important than identifying the hardware problems, the test revealed changes in the software that could make the system dramatically more functional. Couriers were required to answer 12 questions for each delivery exception (such as business closed or package delivered next door), no matter what kind of exception it was. By asking the type of exception and then branching to only the three or four questions required for that type of exception, the software changes could save the courier significant time. Such simple things as making the Tracker ready to scan instead of requiring the courier to turn on the scanner, having the unit always return to the last scan used, and storing information that was often repeated, made the Tracker much easier to use. Many of the suggestions came from the couriers, and after each test a few more changes were incorporated.

The tests also revealed some positive aspects about the project. The couriers enjoyed using the Tracker—it made them look "high-tech." Customers wanted to know more about the device, and couriers from competitive package carriers stopped the Federal Express couriers and asked about the "FedEx computer" their customers were talking about. Not only did the couriers enjoy using the Tracker, but the information they collected allowed the customer service agents to answer customer inquiries on the initial telephone call. The result was fewer package traces for the station and fewer telephone calls from the couriers reporting on exception packages.

Not all of these benefits of COSMOS IIB were free. It took more time for the couriers to run their routes now—not a great deal, and newer software was cutting the time as was the learning curve, but the costs could not be ignored. Even before the tests were over the project team was working on the financial justification for COSMOS IIB. By this time, however, the environment had begun to change.

Financial Environment

By August 1985, ZapMail, Federal Express's 2-hour door-to-door facsimile product, had been in operation for about a year and was performing below plan. The operating losses of ZapMail resulted in a shortage of funds for other projects, and so COSMOS IIB was very closely scrutinized. Jim Barksdale, Chief Operating Officer, requested a complete itemization of all the resources required for full development of the project. The COSMOS IIB team was obliged to take inventory and resell the project to the company.

Field management was very concerned about the productivity costs associated with performing both delivery and pickup scans. Performing the delivery scans eliminated some of the couriers' paperwork, but the pickup scans provided no benefit to the couriers for their efforts. In fall 1985 the senior vice president of ground operations agreed that if the Tracker could provide routing information as part of the pickup scan, ground operations would have no objection to implementing all the COSMOS IIB scans. Adding the routing function would, however, require huge additional memory, and that meant a total rework of the Tracker.

A distinct advantage of an electronic routing guide to replace the familiar paper guide was the ability to have each origin station route a package differently based on destination. The old method required everyone to route a package the same way. Identical routing worked well under the old huband-spoke system of sending all the packages to the Memphis Hub, but volumes in the Northeast Corridor and other areas of the country were large enough that local trucking networks were being implemented, and this practice was expected to grow. The problem with making the trucking network function was recognizing which packages to keep locally and which to send to Memphis. An electronic routing function would provide a code, based on origin location and destination ZIP Code, that would indicate where the package should be sent for sorting.

The combined impetus of the desire to make the courier's job easier and the operational efficiencies and flexibilities inherent in the electronic routing feature were seen as enough reason to push ahead for the next version of the Tracker and delay the full implementation. This more powerful scanning device soon become known as the SuperTracker and was marketed by Hand Held Products as its Micro-Wand II.

Computing Environment

COSMOS IIB required significant amounts of computing power and network capabilities. Management saw some clear business synergies to using the already established and currently underutilized ZapMail network for COSMOS IIB. Consequently, communications and data-processing parts of the project were redesigned based on the ZapMail network. This meant all the programs developed for the MicroVAX station processor had to be rewritten on Tandem computers; consequently, all the applications programmers in Colorado Springs went into intensive Tandem training.

The decision to use the Tandem network did not totally eliminate the need for communicating hardware in the station. Some device had to be in the station to charge the SuperTrackers and poll the devices for uploading of data and downloading of software. This device was termed the Smart Base. The actual computing power and control would reside on the ZapMail network. Considerable discussion occurred over the virtues of designing a unique piece of hardware for these station functions as opposed to using an off-the-shelf computer, but because of the limited functionality required, the cost advantages, and the growing aversion to distributed systems, the unique hardware solution prevailed.

Competitive Environment

Another element weighing in the COSMOS IIB decision was the changing competitive environment. By 1985 all of Federal Express's competitors had adopted the hub-and-spoke distribution scheme, service levels were improving, prices were being pushed down, and at least one competitor was doing selective scanning similar to the COSMOS IIA scans. It was becoming difficult to tell the difference between the various services and, thus, more difficult to justify Federal Express's premium price. Besides providing better information to the customer, collecting valuable data on the operations, and making the distribution system more efficient, COSMOS IIB became the lead project in a campaign to use technology as a way to differentiate Federal Express from its competitors. (Two years later, in 1987, COSMOS, DADS, and the SuperTracker starred in a national Federal Express advertising campaign.)

Implementation Financing

The financial plan for the implementation of COSMOS IIB was revised based on the use of the SuperTracker and the use of the ZapMail network. Every department in the company was polled for its requirements. The financial groups went to great lengths to eliminate duplication of resources and built all the requests into a single package so the total impact of the project could be easily assessed. In January 1986 the entire budget proposal for COSMOS IIB was accepted.

This blanket endorsement of the project and unconditional support from upper management grew from several sources. First and foremost was still Fred Smith's conviction that the information about the package was as important to customer satisfaction as moving the package itself. Now, however, the burden was on the project team. Harry Dalton had been given everything he had requested and there could be no excuses for anything short of a complete and timely implementation.

The implementation plan now called for 27,000 SuperTrackers to cover a nightly volume of 500,000 packages. COSMOS IIB would be implemented in two phases. The first phase would begin in October 1986, would require 14 months for complete rollout, and would include only the delivery scans. The pickup scans would be implemented starting in April and would be completed by December 1987.

The reasoning for this plan was simple. The SuperTrackers would not be available until September 1986. The delivery functions of COSMOS IIB could be rewritten in that amount of time, but more time would be required to redevelop the other functions. It had also become clear in the field tests that the couriers did better if they were trained on only a limited set of scans. Once they were familiar with the scanner, they could easily learn additional functions without as much effect on their daily productivity.

Project Evolution

With the scanner in redesign and the applications in rewrite, the project team's attention turned to Smart Base development, a second vendor for the SuperTracker, and a mini-implementation.

Hand Held Products had taken all the suggestions and findings from the field studies and built them into a finalized production device it was marketing as a Micro-Wand II with the only difference from the Tracker being the keyboard. Federal Express had already developed a method of collecting the scan data and posting it to COSMOS, and the project management felt a great deal could be learned from having 10 to 20 stations doing COSMOS IIB scans while they waited for the new large-memory SuperTrackers. As a result, 1,100 small-memory Trackers were purchased and in January 1986, Federal Express began its rollout of 19 stations dispersed across the country.

The Smart Base began its own development schedule in late 1985. RFPs were distributed and proposals were considered. Norand was finally chosen as the Smart Base vendor. Norand was also chosen as the second source for the SuperTracker.

With 19 stations using COSMOS IIB and implementation plans for the rest of the country in full swing, all parts of the company were pulled into the project. Task-force meetings were held weekly, and monthly project reviews with Ron Ponder, senior vice president of the Information Systems Division, of which Dalton was a part, had to be held in auditorium-size conference rooms to handle the number of people attending. Dalton was regularly required to provide formal briefings to senior management and occasionally to the board of directors. Dietzel's staff was soon doubled to 14 to help monitor the stations already on-line and to manage the continuing development effort.

The monthly review meetings with senior vice president Ron Ponder typically lasted a half day and included reports from all areas of the company. Although the meetings finally became so large that they had to be broken down into submeetings, they served a vital purpose. Federal Express had adopted the practice of developing a separate computer system for each new project, such as billing, payroll reporting, and field reporting. These systems had developed different ways of identifying locations and in some cases had even begun to use the same piece of information in different ways because the exact meaning of the information was not clear. COSMOS IIB was designed to deliver information to all of these systems and each system had to be made consistent with the others. Most of these systems were in the Information Systems Division, which made it easier to see the discrepancies and make sure that they were reconciled.

Ron Ponder was also compelled on several occasions to rally the troops and keep the project moving. As with any project, stumbling blocks were hit, and each group would press for a little more time to make its adjustments before moving on. The pressure to meet the implementation date, however, dictated that the process be kept moving. As problems arose, action plans were created and responsible people were appointed to make sure that the plans were implemented on time. COSMOS IIB became a top priority throughout the Information Systems Division, and Ponder was quick to supply whatever resources were required to keep COSMOS IIB on track.

A New Direction

By mid-1986 it had become evident to Federal Express that problems with the ZapMail network were causing service commitment problems. Many of the communications links were land-based telephone lines that were difficult to get installed on time and were often not available in the capacity required to support the ZapMail data flows. Not being able to provide top quality service defeated the purpose of the product, and the Challenger disaster put the ultimate solution of flying a Federal Express satellite on indefinite hold. Top management decided that it was not in the best interest of the corporation to further burden the network with COSMOS IIB data.

In August 1986 Ron Ponder called a special all-day meeting to review the requirements of changing from the distributed, Tandem-based ZapMail network to a centralized IBM solution using the company's IBM network. All groups were to give a detailed synopsis of the impact to ensure that the rollout would continue as planned. Once it was decided to move away from the ZapMail network, the decision to use the IBM solution was fairly easy to make. One of Federal Express's greatest strengths was the large base of IBM expertise it had developed over the years. Many of the problems COSMOS IIB might experience had been solved before.

Probably the largest impact of the move to the IBM solution was the requirement to again rewrite the software for the COSMOS IIB reports. The software development for the Tandem computers was nearly complete. All this work was scrapped and the application programmers were sent for training on IBM's operating system.

The Smart Base also experienced significant modification because of the decision to move to the IBM/SNA system. One of the primary functions of the Smart Base was to communicate on the network, and the protocol it used for this had to change entirely. Also, the local computing power to actually control the SuperTrackers had been eliminated with the move away from ZapMail and the centralized mainframes could not provide this function. What had started as a power supply with a simple 10-key pad and 2-line display used for hardware polling was turning into a full computer complete with floppy disk drive, 750K of memory, and applications software for loading local parameters (see Figure 7).

At this same time the SuperTracker programmers in Colorado Springs raised an alarm that the ever-increasing demands on the software were stretching the capacity of the SuperTracker EPROM. The changes required to make the SuperTracker easy to use and very functional had grown to over 30K and more changes were seen on the horizon. It was recommended that the software be written to operate in RAM instead of EPROM. This afforded almost limitless room for applications software, but there was a risk that static charges, voltage peaks, or magnetic fields could cause the software in the SuperTracker to disappear. If the software was lost, the SuperTracker would become useless until it could be returned to the Smart Base and have the software reloaded. The benefits of unlimited space for software were considered to vastly outweigh the unlikely occurrence of software loss. Another big advantage to RAM-based software was that the SuperTracker software could be downloaded from the mainframe computers in Memphis to the Smart Bases in the stations, and from there into the SuperTrackers—all automatically. This feature would make changes to the software very easy to make as none of the SuperTrackers had to be specially handled.

Problems and a Delay

There were continuous problems and changes, which could not all be solved by action plans and resources. With great reluctance, the project team petitioned for a one-time delay to the implementation schedule. A new im plementation starting date of January 1987 was agreed to, but the implementation completion dates of December 1987 for Phase I and May 1988 for Phase II were frozen. These completion dates soon became the target. If something needed more time to be developed, then something else had to give up its time so the completion dates could be met.

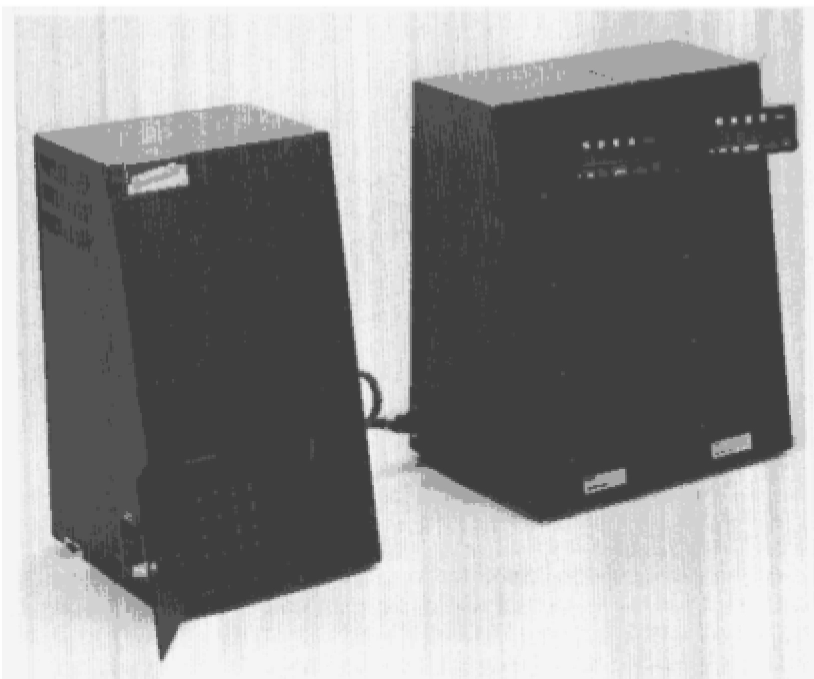

Figure 7 Norand Smart Base.

A variety of problems were hardware related. For instance, the SuperTracker hardware design was often ahead of the chip manufacturers' development schedules. As an example, the SuperTracker manufacturer was waiting for the chip manufacturer to certify a surface mount version of their low-power consumption static RAM for vapor-phase soldering so production could begin. In general, most hardware delays put a strain on internal resources at Federal Express as the engineers worked as hard as vendors and suppliers to overcome the problems.

Other hardware problems also arose with the SuperTracker clone. After a promising start with several innovative ideas, the development effort was soon mired down in difficulties meeting Federal Express's specifications.

The hardware configuration for the clone was essentially the same as in a piece of existing equipment, and the changes required were a bit too excessive to make the unit operate up to snuff. Federal Express was finally compelled to drop the development of the SuperTracker clone.

The changing business conditions had compelled the project team to request frequent changes in the requirements of Smart Base. Norand was having problems fitting these changes into the original design, and the modifications were causing significant delays to the Smart Base implementation date. These delays caused a group in Colorado Springs to initiate a separate effort to develop the Smart Base functions on a personal computer.

SuperTracker Implementation

Finally, in January 1987, 3 1/2 years after the issuance of the RFP, contracts had been modified and approvals obtained to purchase and implement 30,000 SuperTrackers to cover nightly volumes of 750,000 packages.

Training for the couriers and ground operations managers became a major task. The couriers had been hearing for 2 years that COSMOS IIB was coming, what it looked like, and how it would work. But, teaching 20,000 people how to operate a device—even one with software so refined and structured that it would not allow them to make a mistake—was a monumental task. The final plan was to train professional trainers who would travel around the country training operations management and volunteers from stations called information coordinators. These information coordinators would then return to the station and teach the other couriers. The information coordinators were given a full week of training including practice on teaching skills. When they returned to the station they gave 4-hour classes to groups of from 12 to 15 couriers at a time.

Although the rollout of COSMOS IIB had begun, the development was far from over. Before the SuperTrackers could be installed in non-DADS locations, the Smart Base had to work so the couriers could transmit their scans. Smart Bases with charging capability, but no communications ability, were developed. However, fully functional units for non-DADS locations appeared to be a long way off. Acoustic transmission by the couriers was seen as an alternative way to implement non-DADS locations, but the modems were still not operating properly. Soon the rollout plan was modified to put all the non-DADS stations at the end of the implementation.

The SuperTrackers were implemented with delivery scan software residing in EPROM for several reasons: (1) it had already been written, (2) the RAM software had not been completed, and (3) the Smart Bases required to load the RAM software into the SuperTrackers were not available.

Lack of a working Smart Base would have prevented RAM-based software from being implemented in January, but the fact was that the software had been caught up in some changes of its own. Shortly before the first version of RAM-based SuperTracker software was to be released, the risks associated with total loss of software were reassessed. It was felt that having a totally useless SuperTracker was unacceptable; a backup version of the scans had to be kept in EPROM so that even if the RAM software were lost, the courier could continue to carry out the basic elements of the scanning process. Several meetings ensued and a pared-down version of the scans was agreed to as was a new schedule. A major rewrite of the SuperTracker software was required to accommodate the new plan.

A Smart Base Breakthrough

The Smart Base looked as though it would become the nemesis of COSMOS IIB. Fitting the broader requirements into what had begun as a piece of equipment with limited functions seemed nearly impossible. But, a working Smart Base had to be available for the project to succeed.

The word was out that the group working on the PC-based "bootleg" version of the Smart Base had achieved a very functional unit using all off-the-shelf hardware and some already developed network protocol software. This "PC Base," as it came to be called, had not been tested on the quality assurance system, but it seemed to hold a lot of promise. Work on the PC Base intensified to ensure that a solution would be available to support the implementation schedule.

In July 1987, as the PC Base was completing its quality assurance process, Norand successfully delivered the working Smart Bases. Federal Express now had two working solutions, but Norand would require several months to supply its bases in quantity and the unit would still have to go through Federal Express's quality assurance process. Consequently, Federal Express decided to use both solutions.

The PC Base used the same computer that the Customer Automation Group was purchasing to use as customer meters. (A customer meter helps the customer automate his or her shipping process by generating airbills, generating reports, and providing paperless billing information to Federal Express.) A deal was made with the Customer Automation Group to "borrow" several hundred of their computers to use as PC Bases. When the Norand Bases became available, the computers would be returned to the Customer Automation Group.

Future

Table 1 presents a summary of the development of COSMOS IIB from 1977 through 1987. In early 1988 COSMOS IIB is still rolling out, is still evolving, and is still being developed. PC Bases are in the stations and some

TABLE 1 COSMOS IIB Timeline: 1977–1983

1977	COSMOS implemented.	First pass at computerizing customer and dispatch information,
1978	Bar code is investigated as a way to collect package tracking numbers.	Concept of recording package numbers at each change of possession first envisioned.
1979	COSMOS rewritten for IBM equipment.	Larger mainframe capacity required.
	Dalton moves to Colorado Springs and begins to test bar-code scanners.	First test of concept shows information can be collected, but better scanners are needed.
	Concept of terminals in vans for dispatching is conceived.	Dispatch volumes growing and logistics of voice becoming more difficult.
1980	COSMOS II is split into COSMOS IIA and COSMOS IIB.	COSMOS IIA only scans packages in the station; IIB will scan packages on road with tougher hardware. Norand equipment is selected and then modified to meet Federal Express's needs.
1981	COSMOS IIA is implemented.	Packages are scanned as they arrive at the station and as they leave the station. Scans are sent to COSMOS to provide customer information.
1982	First DADS systems implemented.	Radio channels too congested for voice dispatching; digital messages are sent to terminals in the vans. Near real-time link of Federal Express computers to courier van is established, the more recent the information the higher the value.
1983	Proposals for COSMOS IIB scanner are solicited.	Established companies are not listening to Federal Express's needs. Hand Held Products agrees to build exactly what is requested.
1984	Funds are approved to develop and test COSMOS IIB.	IIB scanning device called the Tracker. Specifications evolve from a simple scanner, to a scanner with an LCD display, to a unit with a display and a simple keyboard to a final design with a 2-line display and a full keyboard.
1985	Trackers are field tested.	Many hardware problems discovered and solved. Software changes make the unit much easier to use,
	Decision to use ZapMail network.	Station computer and developed software are dropped. Using the established ZapMail system provides business synergies.
	Smart Base planning initiated.	With no station computer, a separate charging unit with hardware controlling functions had to be developed.

1985	Decision to wait for large memory Trackers.	More memory allows electronic routing that enables a more efficient distribution system.
	Plans to develop a second source for Trackers are made.	Federal Express's executive management was not comfortable with the success of the project so heavily dependent on a single vendor.
1986	1,100 Trackers are implemented as a major test.	Using small memory Trackers allowed the system to be exercised and provide data to develop a better ultimate system.
	Decision to stop ZapMail.	Without the ZapMail network the programs had to be rewritten to run on IBM.
	Development of the SuperTracker 11 is discontinued.	A second source of SuperTrackers could not meet the specifications.
1987	SuperTrackers are implemented for delivery scans.	25,000 SuperTrackers are implemented nationwide to scan packages as they are delivered.
	PC Bases are installed and Norand provides working Smart Bases.	PC Bases, started as a bootleg project, are used to keep the rollout moving while Norand completes last-minute changes to their production unit.

have already been replaced with Norand Bases. Equipment challenges still exist with the acoustic transmission not working as reliably as desired and with future plans to have customer meters and SuperTrackers exchange information on a daily basis. Special scan types are being developed for Federal Express sort locations and international-customs operations. These special scans and the accompanying reports will soon bring the benefits of COSMOS IIB to every area of the company.

COSMOS IIB is actually a window to the future. Even before the project is fully implemented, new products are being developed that would not be practical without the information provided by COSMOS IIB. As the field personnel become experienced with COSMOS IIB, they are quick to identify new applications to help them in their day-to-day operations. Ideas never conceived when COSMOS IIB was designed will continue to enhance the value of the project for many years.

CONCLUSION

COSMOS IIB was driven by the concept that the information about the package is as important as the package itself. In fact, the objective of providing real-time tracking and tracing information about the package even tually became a requirement for long-term viability in the marketplace. Because the vendors of bar-code scanning equipment would not provide the proper equipment required to capture the information, Federal Express took it on itself to have the scanners developed.

Federal Express soon discovered that developing specialized hardware with a vendor was a skill in itself. It took 3 1/2 years of working with a vendor to finally define, develop, and take delivery of a scanner that met all the needs of the corporation. However, after acquiring the internal expertise for defining, testing, and monitoring development, a similar hardware project, in the form of the Smart Base, took only about a year to complete.

Even when the equipment was defined and being manufactured, it was a major challenge to the project team to keep the process moving and on target. Continually changing circumstances required multiple iterations in virtually every aspect of the project, particularly in the writing of software. Optimism frequently gave way to frustration as the rules changed or were reinterpreted. The ability to react, adapt, and keep on moving became the hallmark of the project team.

COSMOS IIA experience provided important guidelines in the confusing maze of a constantly changing COSMOS IIB program. Complex decisions about what to implement and how quickly to implement were always arising. COSMOS IIA was a constant reminder that pieces of the system could be added, but the sum of the pieces had to represent a self-sufficient system. For example, delivery scans with only DADS to communicate the information could be implemented, but those scans had to have management reports before they could be integrated into the corporate package tracking and tracing system.

Implementing pieces of the project was considered part of the "don't let best get in the way of better" philosophy of Federal Express. These pieces often required some reworking as later parts were added, but this was not necessarily negative. The experience gained from each new piece helped direct the future phases of the project to an overall better solution. Even entire report programs that were never implemented because of a change in the system hardware design provided new ideas for better database designs and report layouts. This is not to say that the project team preferred to move in bits-and-pieces instead of implementing a total project. However, given the scope of the project and an environment that could not be controlled, there was no choice but to keep what was good and implement what could be developed in the time available.

Despite its trials and tribulations, COSMOS IIB is considered one of the best managed technical projects ever developed at a company that has always pushed to use technology. Federal Express has adopted a learn-as-you go methodology that allows the company to benefit from the mistakes as well as the advances that it contributes to the technology it is attempting to use.

Harry Dalton sums up this general approach in his oft-heard comment—"The only thing we know for sure is that we won't get it quite right the first time, but we'll come back and fix it."

ACKNOWLEDGEMENTS

I would like to express my appreciation to David Dietzel, Jim Turpin, Pat Mulvey, Chris Demos, Arun Kumtha, and Chuck Theobald for their direction, input, and support in the preparation of this chapter. My special thanks to Harry Dalton for sharing his project management insights, his unwavering intent to provide a factual, instructional case history, and his overall participation in making this endeavor a success.

Electronic Automation at the New York Stock Exchange

Christopher Keith and Allan Grody

In mid-October 1987 the New York Stock Exchange (NYSE) was deluged by a percentage increase in volume of orders greater than anything that had occurred in the great 1960s volume surges that created the major "back-office crises" of that period. Happily, with some creaks and groans, the NYSE's systems withstood the unexpected strain, which was far larger than the NYSE's peak load forecast. This case discusses both the evolution of those systems and some of the major issues in their development.

There are two crucial aspects of almost any system effort: (1) the system as a pure technology (getting the technology to perform on its own as a system); and (2) the system as a component of a broader environment (interfacing it correctly with its users). These might be likened to the two problems facing the would-be inventor of an artificial heart machine. Problem one—get the little valves to work properly (that is, oxygenate and pump the blood). Problem two—related but very different—get the body to accept the device in the first place (whatever its supposed solution might be). At the NYSE, problem one was formidable, exacting, and full of the unexpected. Even so, it was in the area of problem two, the system in relation to its context, that the NYSE encountered its most demanding and distinctive challenges. Accordingly, in what follows, the system as a component of its environment will receive special emphasis.

The Trading Process

Ignoring for the moment the Exchange's automation programs, the flow of floor trading takes place more or less as follows:

First, orders arrive at the Exchange at "booths" situated on the floor periphery. The booths are maintained and operated by brokers, either "upstairs" brokerage houses or independent "floor" brokerage firms that work on a commission basis for a number of upstairs brokerage houses. Orders may be either transmitted from upstairs systems to printers on the floor or phoned in. In either event, floor brokers take the orders received at their booths to the appropriate "post," where the stock in question is traded. The NYSE has its own central switch, Common Message Switch (CMS), which routes orders to the floor from a variety of users and systems and routes reports back to these same systems. The bulk of this traffic is routed to specialists and the posts, but some is routed to the booths. This is referred to as the *Order/Report Delivery* subsystem.

Second, at the post, the specialist for the stock in question and the "crowd," other brokers interested in transacting that stock, are waiting. The specialist operates in one or more of the following roles according to circumstances: (a) as a broker's broker, taking the order from the floor broker and executing it, when appropriate (for instance, in the case of limit orders); (b) as a dealer, obligated to make a two-sided continuous market, and to "bid" if no reasonable contra-side is available (although having to better any contra-side in order to "deal"); (c) as a trading arbiter/manager (auctioneer), responsible for a fair and orderly market; and (d) as a matchmaker/ consultant, providing a variety of informal services that help a broker to find the other side.

The order, once brought to the trading post, is executed or, if it is not at current market, is left with the specialist for representation. If it is executed, the execution may be with another broker in the crowd, or with the specialist representing either his or her own quote (as dealer) or with another order left with him or her previously (as broker's broker).

The functions that support this activity directly are called the *Trade Support* subsystem. The crucial environmental factor here is that trading must take place more or less instantaneously in the trading environment, and requires some effort from those trading. Yet the systems must also allow accurate subsequent processing of the trades made.

Third, once a trade is executed, a member of the NYSE staff, a reporter positioned to "overhear" the trade, reports its elements by stroking an optical sense card and feeding it into a reader. This report goes to the Market Data System, which generates the ticker output, the High Speed Line. The reporter has a similar function in reporting quotes. Subsystems associated with these functions are called the *Market Data Reporting* subsystem.

Fourth, after execution (and reporting by the NYSE staff reporter), the broker returns to his or her booth and ensures that the execution is reported back (sometimes by telephone, often by terminal) to the entering party. If the specialist represents one side of the execution, he or she reports similarly.

The NYSE facilities (particularly CMS) are generally used for specialists' reporting and sometimes for brokers' reporting purposes. This, too, is a function of the *Order/Report Delivery* subsystem.

Fifth, the various parties to the trade then report their portions of the trade to the clearing corporation, which compares these reports. Discrepancies, if any, are usually resolved there. Once the trades are matched, they are netted and settled. The parties to the trade deliver either the requisite securities, perhaps using the Depository Trust Company, or the requisite money.

These comparison, clearing, and settlement processes represent the major components of the *After-Trade Processing* subsystem—and in times past, the major paperwork bottleneck. In addition to enabling comparison, clearing, and settlement, this system must accommodate the interfacing needs of depositories, transfer agents, registrars, varying state laws, and Employee Retirement Income Security Act regulations.

A Cooperative Activity

Even the rather cursory overview of trade processing above suggests that the consummation of a trade, and hence its processing, involves the cooperative efforts of multiple independent parties. Of the four major subsystems introduced above, the NYSE, from a systems operation perspective, controls only Market Data Reporting from the input process through final service generation (generation of the ticker and the high-speed lines). In all the others, an independent entity either is responsible for input submission or, at least in part, "operates" the service offered. One prime environmental difference, then, between the NYSE's automation system and most others is the complexity of the partnerships needed to produce a result.

The NYSE does not enjoy the position of a hierarchial organization with an implied line of authority to discipline all parties. Nor is there a single, distinct provider/user line controlling interfaces at all points. The NYSE's systems must serve a number of partners, each of whom has a different mix of functions and priorities. Even institutions within the same category can have very different perspectives; for example, exchange staff, specialists, $2 brokers, floor brokers, upstairs brokers, or independent brokers with widely differing policies and mixes of business. However, there is another difference that can hardly be passed over; that is, the nature and characteristics of the typical "automatee." Elsewhere, automation may involve automating production workers, shipping departments, or payroll departments—operations with little individual power. On the NYSE floor, however, the "automatee" is likely to be an independent businessperson or person of some wealth and prominence in the community, and quite possibly a member of the NYSE's Board of Directors.

Over this pluralistic universe the NYSE staff, the "automater" in this case, can perhaps impose some level of automation standards, but the ability to impose is more the ability to persuade. For a new system to work reasonably well, a substantive consensus is needed among a small but mixed community, few of whom report to a common source of authority.

In a sense the Exchange floor might be compared not to a typical department or business function but to a coral reef whose members share a common agreement over basic rules, but who probably have very different ideas about the future, its needs and imperatives, or the priorities and procedures of the present day. Like the coral reef, any new arrangements of the Exchange's system will depend in large part on the persuasiveness and persistence of the members themselves. At the NYSE, members had to contribute significant time to efforts ranging from problem identification and prioritization, through persuading other members of the need for change, to the choice among possible solutions, and the implementation of critical elements in the system.

The Marketplace

Marketplaces are notoriously difficult, it seems, for nonpractitioners to understand, especially automation experts schooled in traditional applications. At first, the Exchange floor and the trade-processing function might appear to be a "plant" producing transactions. But that is only part of the trading arena's service. In addition to consummating trades for those who enter orders, the Exchange provides vital services for those who have no direct connection with it, and who do not use its services explicitly. Like any marketplace, it not only "transacts," it also "sets value." The state of the Dow Jones Industrial Average, for instance, is taken by much of the public as a measure of the financial and commercial state of mind of the country. Businesses who neither buy nor sell on the NYSE significantly guide their business plans by the values set on their own stocks, their industry's average price movements, or a potential raider's relative stock value.

The establishment of value—in both how it is determined and the fact that it is taken as valid throughout the financial world—has profound consequences. The Exchange produces not only the trade and its associated information trail but also an accepted mechanism for establishing value or price. In Exchange argot the "price discovery" mechanism is itself an important product. What price means is not what someone ought to be willing to pay for a stock, or might be willing to pay, but what someone is willing to pay for it—now. To be meaningful, however, a price must first be actionable. To be actionable, there must be buyers and sellers, each willing to transact in a given volume at a certain price. This may seem self-evident, but the result is that an exchange floor must be looked on as a kind of date bureau, where highly dispersed and diverse customers with similar interests can find each other at the moment they desire.

What this translates into, in technical terms, is respect for a process. It is not just what the market support system produces as a result that is consequential; it is, to a much greater extent than elsewhere, the process by which the result is produced. The question frequently arises, "The current processes by which the Exchange floor operates often appear cumbersome, even anachronistic; why is it they are not replaced with something really efficient?" Or, expressed another way, "Why isn't everyone upstairs sitting at efficient computer terminals?"

What such questions seem to imply is that efficiency (particularly paper processing efficiency) ought to be the priority. What else is data processing for, if not to be efficient'? One should also ask, "Should the 'price-discovery' mechanism be compromised for more 'efficiency'?" Even in the interest of supposed efficiency, the environment of the Exchange moves only with great caution in tampering with the evaluation process. Perhaps the reader will see why in the next section.

The Valuation Process

Of the valuation methods so far devised, there is little question in the minds of most professionals that the fairest process for setting prices is the *Call market*. There are numerous versions of the Call market, but in all versions the various participants—each of whom has a disposition to buy or sell, in differing quantities, depending on price—meet to reach a consensus price. In some markets this is called (and unfortunately so) the "fix" (see Figure 1). The NYSE uses a variant of the Call market at market opening. The trouble with the Call market is that although it serves admirably in establishing a price, it does not handle the almost equally important variable of time. The pace of the modem commercial world has become such that continuous pricing is an absolute essential for many types of instruments. Both of the two principal types of order, the market and the limit, involve at least implicit settings of both price and time. The market order specifies a fixed time (now), with price as available. The limit order fixes price (as indicated), but allows time to vary.

A reasonable successor to the Call market might be one that offers both price and time advantages—a *Competitive Dealer market* (see Figure 2), in which the combined judgments of dealers theoretically simulate the Call market and back up those judgments with actionable prices. There is much about such a system that appeals to the intuition. For instance, if one were selling one's car, it would be reassuring to know that there would be relatively continuous competition on the issue of price. One imagines oneself shopping at various dealers over many days and then taking the best timing and price offer.

BRIEF HISTORY OF TRADING

The emergence of a Competitive Dealer market was the first major evolution that took place at the NYSE. Just after the Civil War, a broker who had not yet fully recovered from a broken leg settled down on the floor and decided to "specialize" (the stock was Western Union). He was not a specialist in the current sense of the word; he was a competitive market maker—no one awarded him an exclusive franchise. By the 1930s, there were such competing "specialists" in more than 300 of the NYSE's stocks; and as late as the 1960s, there were competing specialists in more than 50 stocks.

The Unitary Specialist Market

Gradually, the competing market system was replaced, de facto, on the NYSE by its current *Unitary Specialist market*. The Competing Dealer market seems plausible, but has a flaw. In such a market buyers and sellers do not transact directly; they transact only with market makers, and they pay an interposition cost which is equal—considering first a buy and then a sell—to the dealer spread. Returning to the analogy of selling a car, it would be clear there is a potentially better alternative than finding the best dealer—and that is finding the ultimate buyer directly. Providing that the cost of finding the customer is low and that one can be sure of a fair price, the buyer and seller can then share the dealer's spread.

This is the central rationale behind the evolution of the NYSE to its current structure (see Figure 3). When activity rates rose, particularly for the larger, better known companies, the chance of finding the other side directly was no longer a theoretical possibility, but a practical and achievable objective. Now at the NYSE, for example, the specialist operates as a dealer in less than 20 percent of the shares traded. Instead, the specialist helps the broker find the other side, sometimes charging a commission for this service, sometimes not, but always keeping any commission below the spread. A simple computation, using easily verifiable statistics, places the value of this reduction in interposition costs at more than $5 billion in 1986 alone—a much greater savings than the cost of the entire NYSE, with specialists' profits included.

Using its current market structure, the NYSE became the world's dominant equities exchange, and one whose value-setting mechanism had unequaled credibility in global financial circles. The NYSE tape print became an assurance of an equitable deal, or at least a best effort to that end, throughout the world. This being the case—or at least the NYSE's impression of the

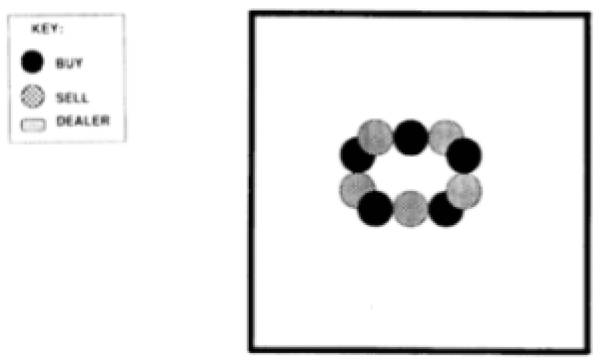

Figure 1 Call market. Buyers and sellers transact directly at call price.

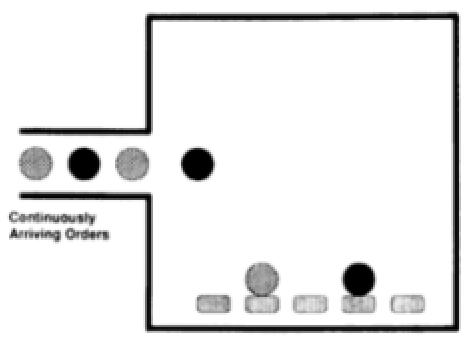

Figure 2 Competitive Dealer market. Buyers and sellers transact through dealer.

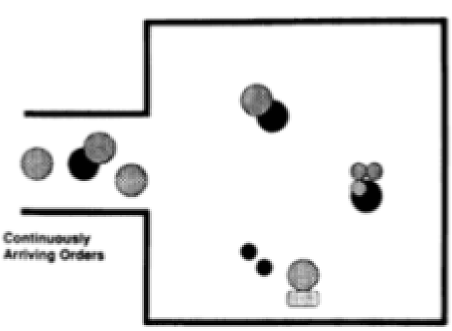

Figure 3 Unitary Specialist market. Specialists act either as brokers (bringing together buyers and sellers) or, if buyers and sellers cannot be matched, as dealers.

case—it is no wonder that the NYSE is reluctant to make fundamental changes in the price discovery process, no matter how tantalizing the seeming efficiencies. An understanding of this priority is essential to an understanding of the NYSE's automation. First, the institution's value determination methodology must be preserved, with efficiencies achieved only within that overriding constraint.

How automation developed within these constraints is instructive. We will introduce various subsystems in the chronological order that they were first put into practical use. The subsystems were determined essentially not by their final purpose in a grand system, but by the environments in which they operated and the human-related constraints imposed on them by those environments. That is what led to the chronology.

Market Data Systems

The NYSE's earliest use of electronic technology dates back to 1867, when the Morse Code Ticker was created. The technology was the telegraph. The pace of trading was such that there was no particular problem in keeping up with transactions. The system's constraint was the capacity to read symbols at the receiving device—and little more, other than maintainability. This simple, successful system imposed no demands on its environment. In 1881 the first electromechanical fstmechanical board was installed on the Exchange, displaying last sale.

The basic theme of ticker upgrades, starting as early as 1890 and continuing to the present, was speed. The 1867 ticker ran at a stately 50 to 60 characters per minute. In 1930 the ultimate seemed to have arrived, the first "high speed" teletype system with a capacity of 500 characters per minute. By 1953 developments in magnetic drum recording made it possible to add automated voice announcements. Quotations, recorded directly from the floor, were automatically relayed to subscribers who dialed the appropriate number for any of 200 active stocks (eventually expanded to 300). This was many times faster than the manual quotation system, which dated back to 1928 and depended on Exchange operators to read the quotations from boards and relay them to member firm subscribers by telephone.

In the late 1950s many new approaches—including the use of television, teletypewriters, and electronic rather than telegraphic printers—were studied by the Exchange. In 1961 the Exchange inaugurated the Special Bid Ask Ticker Network, making possible the transition of quotations on 800 stocks to electronic interrogation devices used by member firms throughout the United States. That same year, the Exchange's work with Teletype Corporation resulted in a technological breakthrough—a new design for a high-speed printer that could relay between 500 and 900 characters a minute.

To this point, what had paced technology was essentially only technology.

As the pace of trading grew, technicians applied available technology to increase speed and improve accessibility to the market. But the joint NYSE-Teletype Corporation system in 1961 reached a new kind of barrier. Nine hundred characters per minute was about as fast as anyone could read; faster tickers would only be illegible.

This led to the NYSE's first major real-time computer system—the IBM-based Market Data System (MDS), a development begun in 1962. What was most remarkable was the attention paid, for its day, to fallback recovery and to the first technical attack on the 900 character limitation. When the pace of market trading approached the 900 character barrier, the system automatically began abbreviating, using a kind of shorthand, a technique that has been in the arsenal of ticker system designers ever since. In addition, a much more readable price display unit was developed—the 900 character moving ticker display. This new device utilized luminescent plastic disks that were flipped by air jets as they moved along at speeds matching those of the 900 character ticker.

Early in 1965 a new system, the Quotation Service, became operational, constituting the first major application of the new MDS. Subscribing member organizations could dial a four-digit number for any of the more than 1,600 listed stocks. Less than a second later, a spoken message was heard giving the latest bid-ask and last sale information. A message containing the open, high, low, last sale, and volume was also available. Each message was automatically composed for prerecorded words by a special voice assembly unit in the Exchange's Computer Center. Designed especially for the Exchange, the MDS Quotation Service could handle as many as 400,000 telephone inquiries each day—an increase of 60 percent over the previous peak load then available.

The MDS was designed to handle a trading volume (then thought quite impossible) of as many as 60 million shares each day. Inauguration of the Exchange's comprehensive market price index, calculated by the MDS's computers, followed in July 1966, the same year in which optical card readers' mark-sense cards began driving the ticker. Referring to the MDS, *The New Yorker* in its June 1965 "The Talk of the Town" said:

> The New York Stock Exchange has recently automated its stock quotation services to brokerage offices, the purpose of which is to give by voice over private telephone wires the pertinent current statistics on any listed stock that the inquiring broker wants to know about. . . . One should not confuse the Stock Quotation Service with the more familiar ticker which records completed trades by teleprint or on paper tape rather than by voice. . . . The voices used to be those of a covey of girls who sat at a switchboard on the fifteenth floor of the Exchange's building at 11 Wall Street; now the voice in all cases is that of a young IBM engineer named Robert Rea, whose voice is stored in a drum inside a computer on the third floor of the building.

Order/Report Delivery

In 1966 the Odd-Lot Automation (OLA) program began with the ostensible intention of increasing efficiency, improving service, and providing for anticipated high-volume levels in odd-lot (less than 100 share lots) trading. Odd lots were not (nor are they now) traded under the NYSE's agency auction principle, which fosters direct meeting of buyer and seller. Instead, odd-lot orders were executed by special dealers (odd-lot houses) who executed trades for odd-lot customers at the "next" execution price, plus or minus the odd-lot differential of one-eighth. The NYSE had other objectives in mind as well. By switching odd-lot orders by computer, it reduced its dependency on a manual pneumatic tube system still in place at that time. The NYSE intended to switch the odd-lot orders to an area below the floor, so it could save precious floor space.

The project was initially limited in scope. What was little understood at the time was that the project had effectively ushered in the era of automated order and report delivery. For the first time, a "receptive enabler" was available with sophisticated computer capability that could be integrated with the Exchange's own systems. By 1969 the OLA was completed, providing manual underfloor pricing of odd lots for all posts and, in the process, the most extensive standardization undertaken in the securities industry until that time. At the end of the project, the securities industry and the NYSE had something that would prove more valuable than the elimination of the tube system—a set of uniform formats for teletypewriter messages in a user environment still very suspicious of any talk of computerized order delivery. Experiments and changes in the odd-lots arena did not really threaten the main, high-volume, profitable core of most key players' businesses.

Not surprisingly, the successor to this project was again an odd-lot effort, now switching odd-lot traffic to the one remaining odd-lot house, Carlyle and Decoppet, off-site. The system called Odd-Lot Switch (OLS) was no sooner cutover, in 1972, than the third odd-lot project began.

At approximately this time, the NYSE and the American Stock Exchange (AMEX) combined their data processing departments into a new corporation, Securities Industry Automation Corporation (SIAC)—owned two-thirds by the NYSE and one-third by the AMEX. What AMEX brought to the new corporation, among other things, was an ambitious on-line project based on a Collins front-end, and a Univac back-end. The project, Amex Computerized Order Delivery and Execution (AMCODE) was one of those far-seeing, vastly ambitious efforts that contemplated supporting a broad range of data processing services for a decade. The system, already under development, was taken over by SIAC and became a component of an SIAC plan then under development, Centralized Exchange Network Trading and Unified Reporting

(CENTAUR), even more ambitious and sweeping than AMCODE. The AMCODE project was renamed CENTAUR Message Switch, and again was geared initially as an odd-lot system.

The CENTAUR plan ultimately collapsed but the system would survive with a somewhat more modest name using the same acronym, CMS, for Common Message Switch. (Remember all that stationery with the acronym CMS?) CMS added to the standardization of order and report communications by accomplishing a remarkable technical feat. It interfaced with more than 40 firms on their terms. That is, it appeared to all these firms as a terminal on their networks. In a technical tour de force, CMS was brought on line to all 40 firms with which it interfaced in a single night, after having run in parallel with OLS for more than six months to ensure that it duplicated OLS results in all cases.

What the NYSE now had was ostensibly an odd-lot system, interfacing with all of its principal customers and using standards that would apply equally to round lots. But it was an order-gathering tool of such competitive significance—although still seen as an odd-lot system—that the Securities and Exchange Commission, during its active pursuit of a National Market System in the mid-1970s, talked of nationalizing it. What CMS turned out to be—although significantly upgraded in capacity and since migrated to Tandem computers—was the major component of the NYSE's Order/Report Delivery subsystem through which all electronic equity orders to the posts (and much to the booths) now pass.

Trade Support

Most of these early attempts at automation had been, on their own, at least modestly successful. If these systems did not achieve everything they set out to do, they at least worked and provided some benefits. However, beginning in the early 1970s, pricked by the national attention caused by the back-office crush of the late 1960s, and spurred on by aggressive automation projects among the member firms themselves, the NYSE (and its subsidiary SIAC) made various attempts to "automate the floor," to support floor trading using automation systems.

These were uniformly unsuccessful. Typical among them was Automated Trading System (ATS), which purported to provide electronic support for the specialist—in effect, becoming the first electronic book. Unfortunately, an insufficient portion of the order flow arrived on the floor in electronic form to make any such system practical. Added to this were an overly ambitious objective, too little understanding of the environment, and the lack of a strong "champion" who would defend and overcome the new experiment's awkwardness.

The result was that ATS remained only a pilot project, and slipped from pilot to limbo status almost without notice. Much the same thing happened to such other large-scale projects as Locked-In Trade (LIT) and Floor Derived Clearance (FDC). Fifteen years later there is still no general LIT system. Fifteen years later a practical FDC system (which presumably captures substantially all order and report traffic flow) is only on the threshold of feasibility.

This era culminated in the grandiose CENTAUR project. CENTAUR was a plan involving two years of detailed effort in the making, including substantiation by outside consulting firms of proposed savings. What CENTAUR envisioned, among other things, was floor brokers with batteries strapped to their belts carrying around floor terminals. These could not be nearly as miniaturized as today's models, and would be plugged in by hand at the various trading posts. This introduction was attempted in an environment where the CMS was still represented as an adjunct to odd-lot service—and where the head broker of one of the major wire houses, on learning that fields of his order forms had been crosshatched to expedite keypunching, stood in the middle of the floor, tearing the offending tickets into little bits and refusing to trade until the old forms were returned to him.

The problem with all of these systems (CENTAUR, ATS, LIT, and FDC) was that they required too broad an acceptance before they could even theoretically deliver any benefits. There was no way to start them small; the nonacceptance of CENTAUR was so emphatic that it ended the entire utopian frame of mind that had characterized so much of the NYSE's (not to say member firms') automation programs.

It was perhaps coincidental that a new figure at the NYSE began to take an active interest in automation, one who understood the diffuse (coral reef) interests superbly, and had an extremely pragmatic outlook. This was John Phelan, the current NYSE chairman. His basic question was in effect, "Can't we do something simple for a change?"

The answer was the Designated Order Turnaround (DOT) system, a system as elementary as CENTAUR was sophisticated. DOT was conceived, designed, and planned essentially in the course of an afternoon. The core of the solution was a set of cards—mark sense cards—that looked rather like order forms, could be printed at the post just like other order tickets, but had the added value of being turnaround documents (Figure 4). The specialist who used a card to report on a trade no longer needed to repeat all the fixed information on the order (information that the trade in a way affected), but merely had to stroke a turnaround number that summoned that information from a computer bank. All specialists did not have to use the system before it became useful. The system's merits were obvious to those persons using it, and the cards did not look like a big change from the order tickets they were accustomed to.

Coincidentally, the vehicle necessary for handling the electronic traffic

Figure 4 Designated Order Turnaround (DOT). The first successful automation trading support system at the NYSE. The breakthrough step was nothing sophisticated but a simple step that looked like procedures already in place.

for DOT (CMS) was already in place—and already linked to the NYSE's principal customers. With DOT, the era of meaningful support to the trading process had begun.

THE CAPACITY ISSUE

We must now introduce another motif central to the NYSE's automation priorities in the decade succeeding DOT—the absolute capacity to meet peak loads. To understand the importance of sufficient capacity at the NYSE, one must recognize another difference between the Exchange and most of the world. The NYSE lacks a degree of freedom that most corporations take for granted. Telephone companies, whatever the public spiritedness of their systems, have the option when facilities get overloaded to inform would-be callers that "all circuits to X city are busy" and to make the call later. Manufacturing firms can plan inventory levels so that unexpected surges of orders result in minimal customer delays. Airlines have the pleasant option of being "fully loaded" (on occasion), at the risk of turning away customers and creating some mild ill will.

It takes little reflection to realize that the NYSE simply does not have a similar option. For instance, a market order placed at 10:30 A.M. but not executed until 10:45 A.M. could cost a major enterprise millions of dollars; the legal consequences of a mishandled order could be enormous. The product of the NYSE is execution of a market order *now*—or immediately when a contingent event occurs. As opposed to most factories, even moderate delays at the NYSE (never mind breakdowns) are sufficiently important to be reported in the newspapers. Witness, for example, the mystique of the "tape running late," a circumstance still dutifully reported by the media although no longer very germane. The High Speed Line that feeds the vendor and major firm systems is not limited to the eyeball-reading speed of the ticker. It is never late.

In short, NYSE's action systems, as well as its market data reporting systems, have a special character. They are not real-time systems in the ordinary commercial sense of that term, where delays create what are essentially annoyances, such as making a customer wait at an airline terminal. They are more akin to the process-control systems found in an aircraft, where failure to deal with a problem leads to calamity.

The importance given the capacity issue at the NYSE is reflected in its planning assumptions for 1986 where, for planning purposes, the premise was 125 million shares per day. However, the targeted plant capacity was 500 million shares per day—equivalent to a corporate directive to build a very expensive plant under the assumption it would run at an average of only 25 percent capacity.

Variability and Unpredictability

What makes this task even more formidable is that the variability and unpredictability of demand is so great. This is perhaps the most uncommon and noteworthy aspect of the challenge facing the NYSE's capacity planners. Variability appears most dramatically on an interday basis, where capacity requirements can exceed 600 million shares one day and be a "mere" 140 million the next. But the same principle applies to intermoment demand as well (see Figures 5 and 6). To those of mathematical inclination, normal (Poisson) distribution assumptions are ordinarily used in queuing calculations. These assume that probability of an "arrival" at any moment is independent of an "arrival" at any other moment. But at the NYSE, the more pertinent model of order arrival is an Erlang distribution with a high k. In other words, the probability of an order arriving at any moment is not independent of whether an order arrived at the preceding moment; rather, an order arriving the previous moment increases the likelihood of an order arriving the next moment.

The high variability of potential transaction rates translates itself into one of the most demanding requirements for trade support systems. At times the floor is relatively quiet; at others as one member described it, the floor resembles "sharks in feeding frenzy." To would-be automaters, this is an awesome circumstance. These are known as "fast markets." The action around a trading post begins to resemble the tumult of a commodity pit, but the sequentiality of trades must be maintained (not a requirement in commodities, but an important one for the NYSE with its many rules based on "plus tick, zero plus tick," and so forth). The inability to answer the ultimate question, "But how would it operate in fast markets?" led to the demise of many otherwise promising automation schemes.

In the mid-1970s, the NYSE was deeply aware of this high potential volatility of demand as well as the hundreds of millions of dollars the brokerage community was spending on automation and communications gear with the aim of being able to deliver order flow from its end customers to the market more quickly. The NYSE was also aware that, in an era of rapidly moving communications, this could mean demands on the marketplace that could become ever more sharply peaked.

One of the more underplayed characteristics of the NYSE's floor technologies and procedures (in addition to their peculiar value in price discovery) is their extraordinary suppleness as transacting mechanisms. Volume surges in the past have taken down virtually every other version of an electronic market; they have occasionally even caused substantial queuing in the NYSE's electronic systems; and they have resulted in snafus and backlogs in after-trade processing (partially because traders have less time to execute the reporting function in hectic circumstances). Through all these surges, how

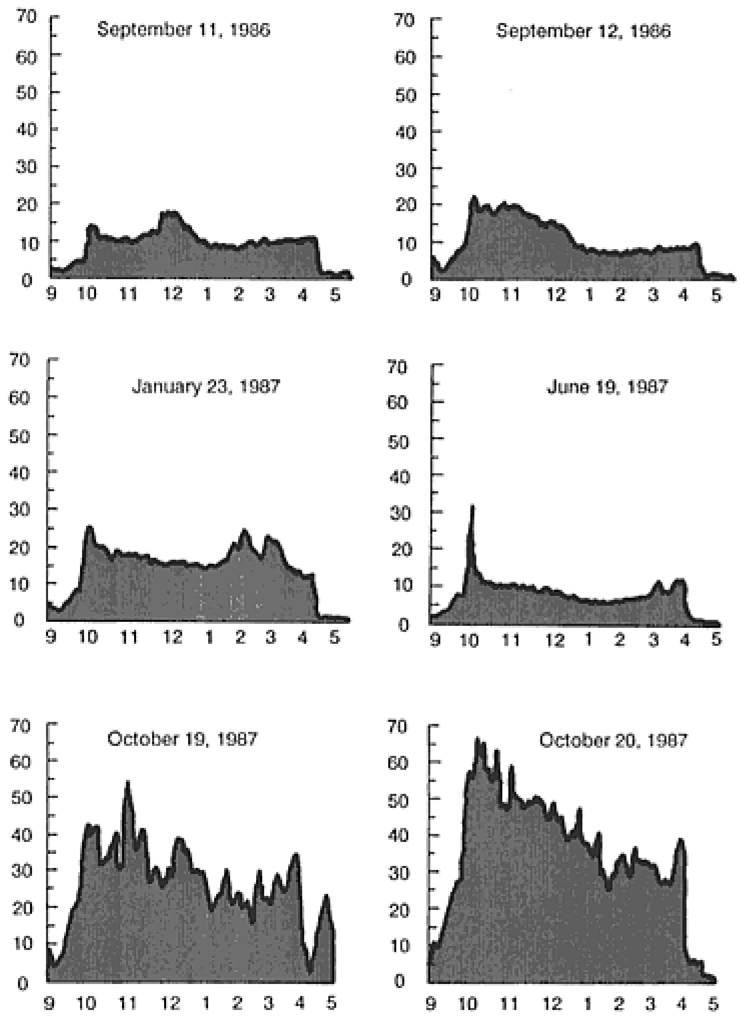

Figure 5 System nonprogram traffic in messages per second during selected trading days, 9 a.m. to 5 p.m.

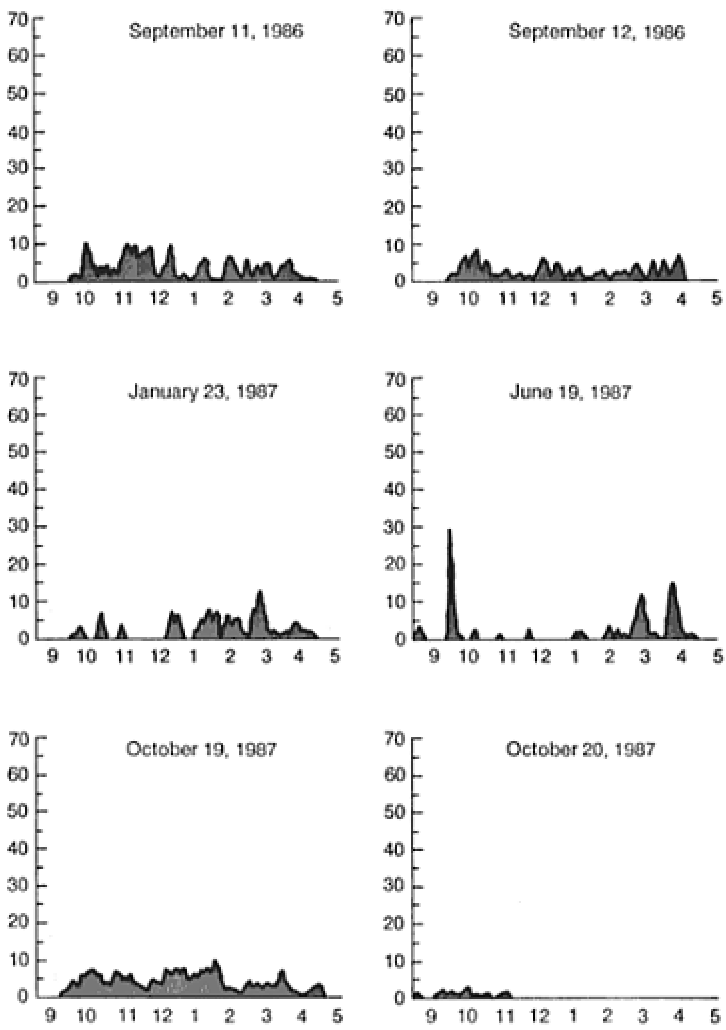

Figure 6 System program traffic in messages per second during selected trading days, 9 a.m. to 5 p.m.

ever, floor trading itself has never been the bottleneck. Although heavily automated, the system allows people to intervene instantly (and intuitively if they like) to radically change circumstances without having these circumstances specifically allowed for in advance. The mechanics and systems that allow the floor to "rev up" this way are fascinating studies in themselves, but beyond our scope here.

Constant Upgrades in the System

With the installation of DOT, an acceptable base had finally been established in the most demanding of the four environments—the trading environment. In retrospect, it seems clear that no radically different mechanism developed externally or imposed from above or the outside (like the grandiose CENTAUR plan) would have been feasible. What was needed was an engine, simple at the outset, that could be improved in place and that was susceptible to being upgraded in small, careful steps. To be successful the purported improvement had to (1) be of clear benefit to users, (2) support the essential methodology of price discovery, and (3) bring the floor community (and the NYSE ethos) along in congruence with the evolving technology.

The DOT system turned out to be very successful, and its genius was its very simplicity. It was developed in a 4-month period from authorization to cutover, used a UNIVAC-Collins message switching system originally intended only to route odd-lot orders, and used for turnaround a UNIVAC disk recorder number already generated in that system for disk logging. But the system was scarcely more than a beachhead in the face of capacity requirements that would appear essential during the coming decade.

The NYSE's entire physical plant would require an overhaul. First, there was the trading floor itself with its wood paneling and its look of history, somewhere between a cathedral and a nineteenth century-style railroad station. The Exchange was undoubtedly a national landmark, but hardly what any industrial engineer would have considered an ideal venue for efficient plant operations. The "posts" were not designed for automation-assisted operations, there was no raised floor of the type necessary for complex electronic interconnections, and the power and air-conditioning facilities were grossly inadequate. The price display units on the floor were electromechanical, and the floor was already crowded with things such as the old IBM 1937 card readers dedicated to market data reporting.

Nevertheless, a line of advance acceptable to the floor and customers alike had to be found, which led to a number of projects that might be likened to shoring up the foundations of an existing building, while erecting a far higher structure incorporating the original in its plans. A grudging consensus on a redesigned floor took hundreds, perhaps thousands, of member meetings. Once agreed on, changes had to be implemented while the floor continued to operate as a trading arena. On evenings and weekends, when the traders left, the workmen appeared; but by morning everything had to be shipshape when the market reopened. Trading often took place under scaffolding.

To handle the anticipated volume in the future, there could no longer be one set of card readers for MDS, another for DOT, and so forth. The various devices on the floor had to be linked through an internal switch so that a single card reader could access multiple systems, and multiple systems could address common printers in an already constricted space. This support system, the Universal Floor Device Controller, like the floor upgrade itself, had to be installed without interrupting operations. When the cutover date came, on a Friday evening, there were card readers for market data and card readers for DOT. By Monday morning, there was a unified set of card readers and new cards with new formats, allowing floor personnel to access both systems via common devices.

There was a similar step-by-step phasing at SIAC. The major effort in the development of CMS involved providing message assurance and adequate fallback. After this system was in place, it was obvious that a more efficient means of developing reliable applications had to be found. A new line of fault tolerant computers, Tandems, was used for the Intermarket Trading System (ITS), an application of the DOT idea to National Market System initiatives. In ITS someone on the floor could use a DOT-like card and reach across to regional markets for "better prices" if available. A conservative incremental approach was followed to lower risks throughout the new systems' introductions. What had happened in introducing the DOT system was that familiar computing machinery (UNIVAC-Collins equipment) was used for an untried application on the floor. ITS reversed this order, using new computing equipment for an application of the type that DOT had already pioneered. With the success of the Tandems in ITS, the DOT system was converted to use them to.

A Step-by-Step Process

DOT's features began to evolve in parallel, a careful step at a time. The original cumbersome turnaround number was simplified. The reader was upgraded several times. DOT eligibility (the size order eligible for DOT treatment) increased gradually from a very conservative 199 shares, and each step evoked anxiety that the "auction character" of the market might be compromised. Even as simple an item as routing order traffic directly to the specialist—rather than to brokers for representation at the specialist post—involved serious price discovery issues, and a rethinking of agency responsibility.

To add flexibility, a DOT "bunching" feature was added. During volume surges, this feature allows individual orders to be bunched according to parameters that can be set by floor officials on-line. Buy and sell orders for a stock can be aggregated, either up to the bunching factor or according to a maximum time delay parameter, whichever is triggered first. Thus, if the bunching factor is 5,000 and the time parameter is 1 minute, buy orders for a stock are bunched in pseudo orders up to 5,000, but in no case do they wait longer than 1 minute, effectively reducing the number of transactions, but not altering the manner in which the transactions are handled. To ensure proper customer service even during volume surges, a guaranteed execution feature was installed for specially designated stocks. This feature provided the customer with either the bid or asked price extant at print time (in theory at print time, in actuality at printer queue time) in the event the specialist had not responded within a specified time to the order (a typical parameter is 2 minutes).

Throughout all of this, the attempt was not just to improve throughput and efficiency, but to do so in balance with the essential tenets of the auction market. In particular, the Exchange wanted to maintain the ability of any order to receive price improvement—that is, a price better than the bid or asked price. The NYSE's product was to be produced more efficiently, but the character and quality of the product could not change.

Electronic Books

By the early 1980s both a modernized computing facility and a floor capable of supporting electronic equipment were in place. DOT had come to handle a majority of the transactions on the floor (although a minority of shares traded). Nevertheless, it was obvious that not even these improvements would enable the NYSE to achieve the capacity likely to be required of it. Consequently, the NYSE resumed experimentation with what had not proved successful the decade before—electronic books. Traditionally, when a specialist received a limit order, either the specialist or a clerk took down his or her book for that stock, and wrote in the order. An electronic book, on the other hand, could receive and post these orders automatically. This may sound straightforward enough to the reader who has some vision of traditional order files in mind, but there were some ferocious complexities in creating an electronic book for specialists.

To suggest the kind of complexity it involved, remember that the NYSE has approximately 22 order types, some of which are not, by their very definition, susceptible to being placed in single price dimension. (For example: restricted quantity, cash orders cannot be treated the same as "regular way" orders.) Consequently, some orders simply cannot be integrated effectively in an "automatic" book that is relatively easy to operate. Another problem is that the electronic book is not a market unto itself; it must be interfaced with a human crowd. It should come as no surprise that during surges, Cancels, Cancels with Replacement, and Inquiries arrive in the same surges as orders themselves—usually when the user has the least time to deal with them. In an electronic book environment the specialist must keep one eye on an electronic screen that is changing at typical electronic speeds, and the other eye on a human crowd that is operating at merely human speeds. In addition, he or she must mediate accurately and effectively between the two. During the time required to say "Done, 1,000 shares at 16 1/2," literally dozens of orders could be entered or canceled or both in the electronic book, which looked one way when the specialist began saying "done," and another when at "16 1/2."

None of these problems was insoluble. However, introduction of an electronic book into a human crowd represented a much more significant step in the trading context than providing an "optical sense" card as stand-in for an ordinary order ticket a decade earlier.

Competing Approaches

The NYSE's approach, especially considering its earlier failures, was to develop not one, but two, competing pilot projects for electronic books, hoping that one or the other might be feasible. They were based on very different approaches and technologies. One, Touch Trade, was externally designed (by IBM) and implemented by using PC technology and touch-sensitive screens (as the name suggests). The other, Display Book, was internally designed (by SIAC) using a standard keyboard, with SIAC serving as system integrator for Intel boards. As a commentary on both the Exchange's sociology and the importance assigned these efforts, one of the Exchange's two vice-chairmen served as the personal leader, and, to a significant extent, the detail specifier, for Touch Trade. For several years the two pilot projects coexisted side-by-side on the floor, being tried out and assessed by various specialists. Each was periodically updated with new features and improvements.

Ironically, whereas no previous electronic book attempts had worked adequately, both of these projects, perhaps owing to more experience or to the competition, appeared to be feasible. Touch Trade was an imaginative, easy-to-use system with a checkerboard-type, touch-sensitive display incorporating a valuable innovation—color used not for design or visual contrast but for information, that is, color indicated the order's state. It was geared to do a wide variety of things—handling limits, markets, and quotes—and it did them well. The early Display Book, on the other hand, had a more traditional keyboard and was far narrower in ambition. It was targeted to handle only limit orders, but it was fashioned, above all, for efficient use in heavy traffic situations. Its objective was to do only one thing but do it extremely well.

It will remain forever arguable what actually tipped the scales in favor of

Display Book. One legend has it that the watershed was a particularly hectic day's trading in PanAm stocks, when the specialist user (not entirely convinced until then, and ostensibly a disinterested judge) concluded that he could not have survived that day without Display Book. It came to be rumored, even if not entirely believed, that this new gadget not only would not have to be abandoned during "fast markets," but also would actually be of assistance during them. Perhaps the Display Book success was an echo of the earlier DOT theme—one slice at a time rather than a single broadscale solution. In any event, Display Book was the initial survivor, but its subsequent versions slowly incorporated most of the worthwhile ideas introduced in the competing Touch Trade. Current Display Book is deployed in 280 of the 300 most active stocks and is proving to be exceedingly effective (see Figure 7 for a comparison of display and nondisplay operating efficiencies on October 19, 1987).

Display Book could not have been introduced without the predecessor DOT system it eventually will replace. Display Book itself existed for two years on the floor as a pilot project and went through at least a dozen upgrades before it was thought solid and accepted enough for deployment across the floor generally. Here again, its deployment was as much a result of jawboning

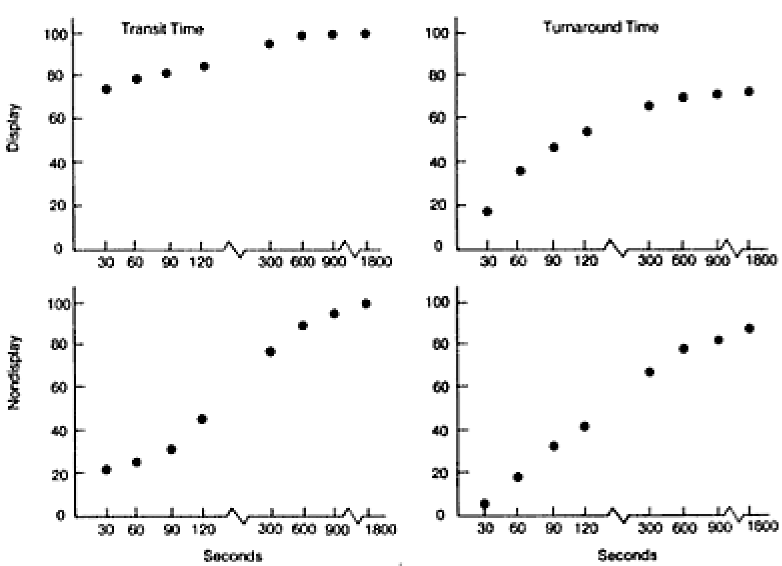

Figure 7 Order-handling performance; cumulative percentage of orders handled. Transit and turnaround time for display and nondisplay systems, October 19, 1987.

by members who had used it successfully as it was of any persuasiveness on the part of the staff. In 1988 priorities call for further assimilation of the technology into the specific needs of the floor; for instance, enabling one terminal to handle multiple stocks or for multiple terminals to handle one stock, according to instantaneously reassignable priorities.

CAPACITY PLANNING AND PROGRAM TRADING

The formal approach at the NYSE has been to define the performance characteristics, constraints, and dependencies of the 20 significant systems (or system clusters) on which effective NYSE operations depend (not all discussed in this document). For instance, the trading floor was expanded. On a Friday in January 1988 a number of brokers and several specialists left their traditional facilities, and on Monday resumed their operations, with little obvious disruption, in a new space. The cathode ray tubes (CRTs) that replaced the electromechanical floor display devices in the 1970s are now in the process of being replaced by flat screen technology, which takes less power and space and provides more display information. The underfloor wiring installed at the time of the floor upgrade was to be upgraded by local area network (LAN) technology. The computing base for all these applications had to be upgraded without changing the visible product. Systems developed over a period of 10 years had to be reintegrated, so that computing resources could be applied more flexibly, and not be tied down to supporting a single function or module.

Along with such planning, external sources, with a prudent eye toward the NYSE's capacity policy, forecast anticipated volume. Historic trends and data patterns enable the NYSE's modelers to convert these inputs first into a set of overall peak-day requirements for each of the major systems. Peak-day requirements are then converted into peak 5-minute-load requirements for the various on-line components and into total item requirements for batch systems. Input assumptions for this process are again based on historic patterns, such as the ratio of peak 5-minute loads to peak 1-minute loads, skewing characteristics, peak-to-average characteristics, and so on.

The outputs from this process provide a target capacity for each of the 20 system components. The measurement of the capacity of each of these components is—as with the forecasts and estimates—a bit of an art. Scatter figures are taken at various levels of channel utilization, CPU utilization, and so forth, and projected into assumed effective maximum rates (somewhat less than 100 percent). Separately, during product tests of new versions of software, the systems are driven at very high rates, and observations are taken as to when queuing begins. Obviously, any particular traffic distribution will impose different loads on different facilities, so that a given system might not begin to queue at 40 messages per second in one test, but, with a slightly altered input, might begin queuing at 35 messages per second.

The process is at least serviceable, and has reliably forecast (within the accuracy of overall volume forecasts) system demand over the last 3 years. Although the volume surges of October 19 and 20, 1987, were in no way anticipated, their impact on the system was very much as the modeling process would have predicted—so much so that further increasing capacity to meet what appeared to be much higher and more urgent capacity goals did not involve starting any single project, merely accelerating ones already well under way.

However, during the last 18 months many of the old truisms (if not all of them), on which so much apparently sophisticated analysis depends, have begun to break down. For instance, for years traffic loads on peak days averaged about 2.5 times traffic loads on an average day. This is no longer true. In the last year this ratio has increased to approximately 4. In the past it was always safe to estimate peak 5-minute demand and 1-minute demand together, since one exceeded the other by only approximately 10 percent. This is no longer true. The peak traffic loads also almost always occurred right after the morning's opening. Now higher peaks occur irregularly, and so forth.

One culprit, of course, is program trading—particularly arbitrage plays between the futures markets and the NYSE (Figure 8 suggests how this impacts the NYSE). But it is doubtful whether this one strategy is the only culprit. Perhaps the real villain is speed itself and the rapidity with which news can spread, resulting in the growing importance of the "turkey farm syndrome"—that is, when one turkey begins to gobble, all suddenly follow suit. With increased communications and computing capacity throughout the industry, the possibility of suddenly encountering immense loads will exist; with the breakdown of all the traditional ratios, one will be able to predict peak requirements with less and less confidence. Therefore, plans will have to provide for greater and greater margins for error—and higher flexibility throughout the system.

CONCLUSIONS

The pattern of the NYSE's automation over the years seems to be cyclic. In Stage 1 of this cycle some very cautious experiment is tried on the floor. With some modifications and adjustments, this ultimately works at least well enough to continue its line of development. In Stage 2 this development, perhaps begun for a specialized application, is deployed across many stocks and trading positions. In the process of introduction the automation support base on which the edifice is built undergoes serious strains. In Stage 3 a significant parallel effort becomes essential to deal with these strains. During

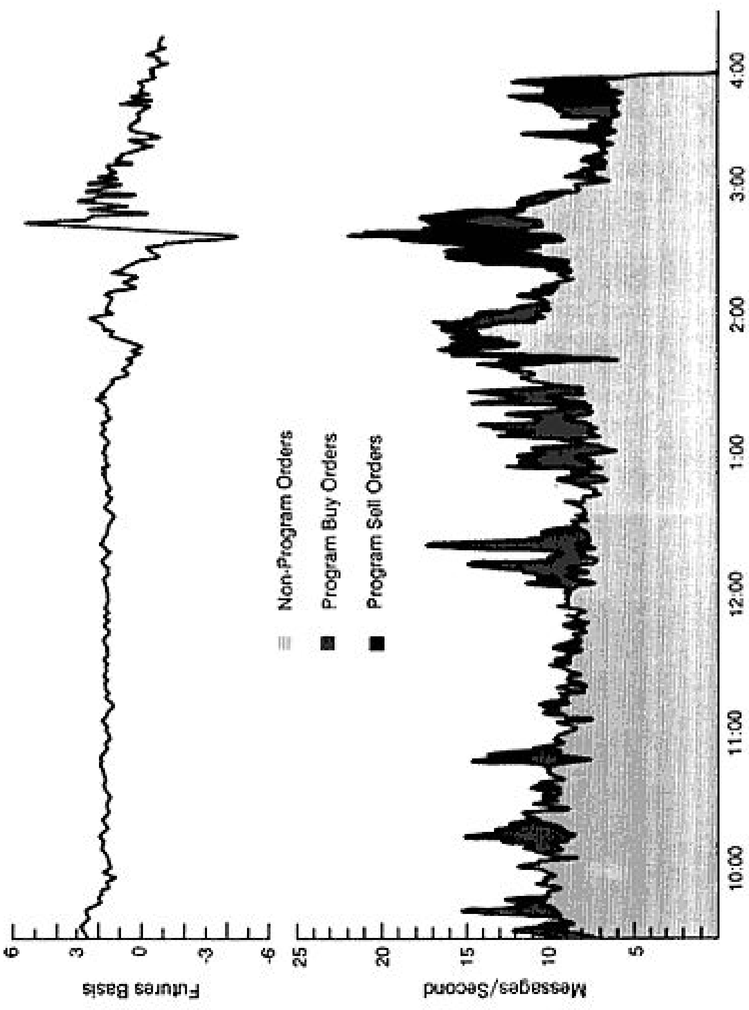

Figure 8 Time distribution of March S&P 500 futures basis and total orders, January 23, 1987.

the process of Stage 3, either unexpected responses to the program instituted or exogenous forces in the industry make it clear that future requirements could well exceed anything the current process could deliver. In Stage 4, therefore, the last efficiency is wrung out of the development that was begun as a cautious experiment in Stage 1; and planners begin to define and evaluate the pros and cons of the next experiment.

By the end of 1988 the floor of the NYSE should be essentially electronically automated, with orders (bunched as necessary) arriving at electronic speeds at specialist positions. The capabilities will be in place to call in or handle multiple added participants (through helper terminals) for particularly hectic stocks. In 1989 SIAC should be well on its way toward creating a more supple system, so that computing capacity may be assigned throughout the system as required. The troubling thing is, of course, that all this attention to capacity is no doubt being matched by the NYSE's customer base. As a result, despite its enormous progress, the NYSE's automation program may not be much further ahead of its requirements than it was at the time of the installation of DOT.

ACKNOWLEDGEMENT

Research and graphics production by B. R. Graphics, Inc.

Using Technology for Competitive Advantage: The ATM Experience at Citicorp

Paul F. Glaser

In the late 1960s external and internal studies performed by Citicorp concluded that advances in communications and computer-based technology would support effective on-line interactive transaction processing services and would become instrumental in changing the manner in which financial services would be delivered to customers. The external studies were done in 1966 by the General Electric Tempo consulting organization headed by Thomas Paine, who later became administrator of the National Aeronautics and Space Administration. Senior management within the corporation conceptually agreed with the findings and initiated efforts to determine the types of products and services that the new technology could either support or bring to the marketplace.

During the period of the late 1960s and early 1970s, contacts were made with the major computer and communications suppliers to generate interest on their part to work with Citicorp to develop the technology necessary to deliver on-line interactive transaction products and services. Qualified suppliers were interested in supporting Citicorp, but felt that it was too early to apply resources to such an effort and that they would be willing to participate only when other financial institutions and industries saw the need for the technology.

Citicorp management recognized an opportunity to provide innovative products and services based on this new technology early in its development. Delivery of the new on-line product concepts being generated would, without significant vendor support, require internal technical resources to initiate the development of a technical systems' infrastructure to support a reliable and secure product delivery capability.

Attempts to initiate internal efforts within the existing management and operating structure of the bank, which at that time was deeply engaged in getting the back office under control because of an increasing volume of transactions, met with little success. Moving ahead of the financial community required a radical management solution because organizations responsible for day-to-day operations were not responsive to such approaches. It was difficult to convince a business manager responsible for day-to-day management and earnings to invest in, or initiate, a project whose payback was questionable and could possibly take more than 5 years when the typical job life cycle of such managers was no more than 2 years. It became obvious that achieving the goal of maximizing the competitive benefits of the new technology as early as possible required resources not tied to day-to-day operations.

In 1968 Citicorp formed a technical product development and marketing company to define and market on-line transaction products and services. The company, Citicorp Systems, was formed in Cambridge, Massachusetts, for proximity to the university environment, personnel, and resources and the emerging computer-vendor community. The attempt failed to produce significant results owing to both the inability to define a corporate direction and a desire to operate independently of existing Citicorp resources, outlets, and infrastructure.

During 1972, after several unsuccessful attempts to achieve significant technical or marketing results, an agreement was made with Quotron Corporation in Los Angeles allowing Citicorp to use its hardware and software patents as well as processors of its on-line quotation system. In addition, a group of 30 people were transferred from Quotron to Citicorp as the nucleus of a staff to apply the technology to financial service applications. These people became part of the new company formed for this purpose named Transaction Technology, Incorporated, a wholly owned subsidiary of Citicorp in Los Angeles. The original Cambridge organization was also merged into this new entity. This team emerged as the driving force for identifying the technically based infrastructure and product requirements.

The aim of the new organization was to demonstrate success as soon as possible. A plan was developed in 1972, and work was immediately begun to deploy an on-line transaction processing network and some limited applications. The goal was to provide customer convenience and to have broad external visibility with minimal need for support from the existing back-office operating organizations of the banks. This eliminated the need to develop immediate, extensive cooperative relationships prior to deployment of any useful services.

The applications selected were to support check cashing in New York City branches of Citibank along with MasterCard authorization and private-label credit card authorization at the point-of-sale. The check-cashing service al lowed customers to cash checks at any of Citibank's 260 branches and in New York City through an on-line determination of available funds and placement of a real-time hold on the account for the amount of the cashed check. This differentiated product gave Citibank customers access to 260 branches for no-hassle check cashing, a true market differentiator at that time.

The MasterCard and private-label credit card point-of-sale authorization system was designed to reduce credit card transaction risk and provide rapid and convenient service to the customer at the time of purchase. A system to provide these services would generically consist of a terminal for input, a means to identify a customer, a network to transmit data from the terminals to the central applications, a data-management application system, and an interface process to exchange customer data between the on-line system and the banks' existing batch-processing accounting systems.

The technical staff that was assembled believed that to achieve reliable system performance with a minimum of risk, at that time, would require three major developments. These were a reliable card encoding and reading technology with low fraud susceptibility; a network and efficient processing environment to deliver applications; and a terminal that could interface with the network, the card technology, and the user.

In the early 1970s the reliability and security of credit card encoding and reading, using magnetic stripe technology, was seriously questioned by the technical staff that had been assembled. As a result, an electro-optical technique was developed consisting of a permanently embedded code in a plastic card—a code that was unalterable without destroying the card—that could be read by a low-cost sensor system whose security and reliability was, and today still is, more effective than magnetic techniques. The electro-optical technique that was developed operated only in an on-line environment, a primary basis for its security.

During the early 1970s, most organizations were concerned with the reliability of on-line technology and the pervasiveness of its deployment. Primarily because magnetic stripe technology supported off-line as well as on-line systems, it became the industry standard. Magnetic stripe data can be altered with little difficulty which leads to fraudulent use and subsequent loss to financial institutions and other users of the technology.

The network was designed to deliver services through a communication system to the terminals. The processing environment was required to both maintain on-line customer data and process high transaction-rate applications. The developers postulated that achieving the reliability and integrity of service that customers, branch personnel, and retail clerks required involved developing a system that would continue to operate in spite of any single point of failure. The solution was to develop a system consisting of two data centers, at separate sites, each capable of operating together in a duplexed mode, or independently, without human intervention. Hardware and software meeting these system requirements were not available and required a major internal development effort for implementation. The telecommunications network also required a major internal development effort to manage redundant paths and to switch between data centers in the event of any component or communication on-line failure.

The computer selected for use in the system was the one Citicorp purchased the rights to assemble for its own use from Quotron. The system also relied on other network components and software from Quotron's network to deliver stock quotation services to brokerage houses. Software for the system, file management, and applications were essentially developed from scratch. There were no commercially available packages that could provide the integrity and transaction rates required by the system, and the hardware that was selected did not have a system's software base that supported the more common software languages of the time. In fact, there was no other solution than developing the applications in machine language owing to the performance requirements and the limited processing power and memory of the minicomputer.

To provide universal access to the system, a simple terminal was devised consisting of an electro-optical card reader, a telephone-type keyboard, and a light-emitting diode (LED) single line display in a package the size of a standard telephone. The terminal was designed specifically for the system under development and its anticipated applications.

By the end of 1973 the initial rudimentary network was installed in New York City with terminals deployed at teller and point-of-sale locations throughout the city. The customers of the bank were issued the debit card, which by this time was categorized by its technology as the "Magic Middle Card." Private-label customers of the system and Citibank MasterCard customers were also issued cards containing the magic middle technology. This provided Citibank customers a "no hassle" check-cashing service at all bank branches in New York City and rapid reliable MasterCard and private-label credit card authorization at those points-of-sale that had terminals installed.

The goal of the system, which eventually would be developed, was to give customers direct access to their accounts 24 hours a day, 7 days a week, without interposing a teller or retail clerk. This goal was partially achieved in 1974 by placing the terminals in branch lobbies and allowing customers to use them without staff support. Customers could, in complete privacy, check real-time balances and obtain other basic information about their accounts. This helped to develop a customer base that was comfortable with technology and its benefits of accurate information and high quality service.

Market research studies performed during the early 1970s consistently indicated that bank customers desired two basic service improvements—time and place utility. This translated into the consumers' desire to have access to their funds 24 hours a day, 7 days a week, and not be confined to their home branch. Attempts to achieve this functionality by installing terminals in retail establishments for access to cash and account information were unsuccessful for many reasons, none of which was technical. Who benefited from guaranteed check cashing or direct account debit at the point-of-sale and how incremental costs would be covered could not be settled. This argument, although new at the time, continues to this date.

It was, therefore, not possible to provide the transactional capability of the branch in a retail establishment. The benefits of increased traffic and services for the retailer, as perceived by the system developers, were not salable. Alternative approaches needed to be considered. In 1975 the technical staff concluded that cash dispensing and depository mechanisms had developed sufficiently that reliability was not a significant problem and automated teller machines were viable with state-of-the-art technology.

It became apparent that the automated teller machine was the most likely candidate to provide the time and place utility and services that the bank customer desired. This was a major conceptual change to the approach originally conceived which was to furnish, in concert with the retailer, increased utility at the point of sale. The functionality on the other hand was no different, only the means of dispensing—from a retail clerk to an automated machine. Fortunately the system was originally designed for the function and was not restricted to a specific delivery mechanism.

Tests performed with commercially available systems, at that time, found serious shortcomings in the areas of the human interface and technical systems' operating characteristics. To furnish a product that could meet the operational requirements of Citicorp, and interface with the consumer in a friendly manner, a design effort for an automatic teller machine was begun in 1975, using as many commercially available components as possible. This effort led to a product design that, after market tests, could be deployed in branches by 1977. These market tests were conducted in a simulated branch environment without consumer knowledge of the test sponsor.

In 1976 Citicorp made a commitment to deploy the automated teller system throughout the New York City branch system. This would be accomplished by taking advantage of the proven network that had operated reliably for more than 4 years and that could be incrementally enhanced to support the broad product set provided by the automatic teller system. To ensure uninterrupted service under all conditions and to retain the system goal of not allowing a single failure to inconvenience the customer, it was decided to place a minimum of two machines in all locations.

On the business side, the aim of the program was to provide benefits to both the customer and the bank. These benefits had to be based on increasing revenue because it could not be proved, at that time, that an automatic teller machine was justified on the basis of operating cost reduction or displace ment. In fact, several limited test programs by major U.S. banks all determined that with projected transaction volumes, cost per transaction could not compete with a teller and, therefore, proved that the machines were not cost competitive.

Market research, however, indicated that the system would give a clear competitive advantage to Citibank by meeting several of the stated customer desires. The system would give customers 24 hour, 7 day per week access to the bank and allow them to use any of the branch locations throughout the city. Customers could get cash, make deposits, and manage account relationships. This functionality would increase the benefit of multiple account relationships with Citibank, allowing customers to manage account balances for maximum return through the on-line information and transfer capability.

The primary motivating factor for Citibank was that these differentiated services would attract additional customers to its underutilized branch system. Additional benefits would result by off-loading time-consuming teller transactions and allowing existing staff to perform additional sales and service functions.

On installation, the system was an immediate success. Customers who had been using their Magic Middle Cards in the branch lobby terminals were completely comfortable using the next so-called generation of machines, and new customers also found the system relatively easy to use. The results were dramatically seen in changing market share. Until 1977 Citibank's share of the New York City market was approximately 4 percent. Immediately after 1977 the market share began to increase to its present rate of 13.4 percent in a marketplace with little or no population growth.

Attributing all this growth entirely to the system is not justified. The system did have a catalytic effect on market share as seen by its dramatic growth that began concurrently with its deployment. This effect could be seen through the actions of both customers and staff. The staff became convinced that their bank had a clearly defined, differentiated set of products and services that could be sold based on measurable customer benefits. Customers felt they were getting something that no other bank could offer, symbolized by the advertising motto "The Citi Never Sleeps," which captured the feeling of the time.

It has been estimated that the total costs of developing and installing the network and the automatic teller machines were returned within the first year of operation as a result of increased revenue within the existing branch system. In 1988 that branch system served three times as many customers as it did before 1977. The branch staff during the same period increased from 7,100 in 1977 to 8,400—a significant productivity improvement in itself.

Citibank retained this differentiated product position with essentially no competitive response until the early 1980s. It took about 4 years for the major banks in New York City to decide how to respond and then develop both the network infrastructure and teller machine base that could compete with Citibank.

Although the development risks were high and required organizational approaches that, at the time, were extremely unconventional, the returns more than justified both. Turning an unprofitable branch system into a significant corporate earning center could have been achieved only with dramatic approaches to increase both revenue from existing customers and new customers. A "me, too" attitude, based on a utility type of service with only price as a competitive advantage, would not have led to the resulting profitability or the development of a fundamentally new approach to consumer banking in the United States and many other parts of the world.

Operations Research and the Services Industries

Richard C. Larson

A dispatcher is using a color-graphics computer workstation to design efficient routes for 14 trucks that tomorrow will carry liquid nitrogen to customers in the Northeast.

After consulting a computer printout, an operations engineer has just directed crews at 6 dams in a 12-dam hydroelectric system to increase outflows by 10 percent for the next 3 days.

The chief executive officer of a major forestry products company is using a video-game-like system to understand better how mill operators in the field can increase profitability by improving their log-cutting activities.

A social service volunteer in Atlanta, Georgia is using two card files to assign drivers to vans and vans to routes to deliver "meals on wheels" to elderly and handicapped individuals.

A vice president of a large railroad is scrutinizing a consultant's report that recommends an ambitious $1.5 billion capital investment program over the next 5 years.

What do all these activities have in common? In each case the individual mentioned is a consumer of a product of operations research. Representing a quantitative knowledge-based service industry, operations research has established a foothold in corporate America, both in the goods-producing and non-goods-producing (service) sectors.

Despite the growing importance of the field, as demonstrated by numerous documented case studies, relatively little is known about it outside its own "inner circles." As illustrated above, the material product of operations research can assume numerous forms, making the field ill-defined to outsiders. Operations research can be highly mathematical and is often embedded in less-than-transparent computer software or mathematical models. As a result, reactions of fright, suspicion, distrust, insecurity, and irrelevance are not uncommon.

In many ways, operations research is a perfect example of an emerging class of technologies (often provided in the marketplace as a service) that might be characterized as decision-aiding "software technologies." Decision-aiding software technologies range from spread sheet programs to complex optimization algorithms specially tailored to a specific application. The professional fields that provide such software and services include computer science (especially with recent advances in "fourth generation" languages, relational data base systems, and "expert systems"), various branches of engineering and operations research/management science (or, "ORMS").

Despite an impressive array of successful implementations, the market penetration of OR/MS in services industries in the United States is low, raising several questions related to operations research. How does a firm choose to invest in OR/MS and what determines success or failure in implementation? Does low market penetration reflect difficulties in evaluating the likely returns to investment? Does it reflect difficulties in translating the analytically rigorous academic discipline of OR/MS into application?

This chapter, therefore, has two purposes. First, to provide an overview of successful applications of OR/MS in services industries. Second, by example and with some admitted speculation the paper addresses the ways in which investment decisions are made about OR/MS applications and the ways in which those investments are evaluated. To get to that point we must first backtrack, spending a little time on a description of the field and a brief review of its history and some of its major accomplishments.

OPERATIONS RESEARCH: BACKGROUND

Decision-aiding Technology

Operations research focuses on developing improved procedures for planning and operating complex systems. To distinguish it from other fields having similar objectives, operations research tends to utilize the scientific method to discover the "laws of physics" of the system under scrutiny. By a process of trial and error not dissimilar from that of a physicist who is both an experimentalist and a theoretician, the operations researcher attempts to develop an accurate mathematical abstraction (i.e., mathematical model) of the system. By manipulating the model, the operations researcher tries to discover improved ways for operating the system. Operations research does not exclude inputs from social scientists and organizational theorists, and a large number of the founding members of the Operations Research Society of America (ORSA) were from these fields (in the year 1952); however, the mathematical model seems to remain a central feature of operations research studies and products.

Most operations research models contain decision variables whose values are to be "optimized" subject to certain constraints. If one views a decision as an irrevocable allocation of resources, the values of the decision variables represent a particular allocation of resources. The optimization objective may be to maximize profits or to minimize costs or to maximize customer satisfaction, or it may be multidimensional, including two or more such objectives. In any event the desired goal of an operations research enterprise is the identification and implementation of improved decisions (i.e., allocation of resources).

Operations research is a state-of-the-art technology. It uses the latest scientific knowledge from such diverse fields as mathematical programming (i.e., computer-based optimization of mathematical functions subject to often complex constraints), stochastic processes, graph theory, and computer science. Since its focus is on improved decisions, we may regard operations research as a decision-aiding technology and thus admissible to a discussion of technologies in services industries.

Institutionally, operations research is carried out by consultants, in-house technical groups, university professors, and—ever more frequently—software firms. Operations research is itself a services industry. And, as will be argued below, its services are most often sought by other services industries, including transportation, finance, government, health care, education, and consulting.

Brief History

Operations research was identified as a field of scientific inquiry and named during World War II. The initial important work, focusing on radar utilization, antisubmarine warfare, and other military operations, was done by two groups of diverse scientists, one (in the United Kingdom) under the direction of the physicist P. M. S. Blackett and the other (in the United States) under Philip M. Morse, also a physicist. The group in the United Kingdom christened the new field "operational research" [see Morse (1977) and McCloskey (1987) for more details]. Among the numerous accomplishments of these groups was the creation of "search theory," a new integrated mathematical formalism combining ideas of probability, geometry, and mathematical optimization and used initially to deploy planes and ships to find enemy submarines. Search theory has subsequently found wide application elsewhere, including design of search strategies to find lost items over vast areas; for instance, it was instrumental in helping search parties to locate the wreckage of the Shuttle *Challenger* crew module.

After the war there was considerable interest in developing and applying methods of operations research to problems of the private sector and the nonmilitary public sector. Considerable momentum was given to this effort by the simultaneous developments in computers and algorithmically oriented mathematical optimization, arising initially as "linear programming" with the celebrated "Simplex method" due to George Dantzig. Subsequent advances in algorithmic optimization, many developed at the Rand Corporation, included dynamic programming (Bellman, 1957), various special forms of linear programming (see Dantzig, 1963), and network flows (Ford and Fulkerson, 1962); the methodological developments fit nicely with the concurrent technological advances in digital computation.

The ORSA was founded in Cleveland in 1952. The first academic programs in operations research were established at Case Institute of Technology and (under the direction of Philip M. Morse) at Massachusetts Institute of Technology. The first Ph.D.s were awarded in the late 1950s.

Although the field coalesced as a result of the war effort and subsequent developments, important components of operations research reach back prior to 1940. Thomas A. Edison, serving during World War I as head of the U.S. Naval Consulting Board, used statistical and gaming ideas to develop important early results in antisubmarine warfare (Whitmore, 1953). Queuing theory, which focuses on the development of mathematical models of waiting lines, had its roots in Denmark during the period 1910 to 1915 when the Danish telephone engineer Erlang used probabilistic reasoning to develop the first queuing models to help engineers determine the capacity of telephone switching systems. Graph theory, which has been used extensively by operations researchers to model transportation networks, is rooted in the efforts of the Swiss mathematician and physicist Leonhard Euler who in 1736 attempted to route a parade over the seven bridges of Konigsberg (now Kaliningrad) in such a way that each bridge was crossed exactly once; in developing the initial important results of graph theory Euler proved that such a route did not exist (and he showed how to design a minimal length parade route that crossed each bridge at least once) (Larson and Odoni, 1981, p. 385). With regard to linear programming and the Simplex method, the now famous vertex-to-vertex descent method was at least suggested by the mathematician Fourier in 1826 and additional important early work was done by other mathematicians: Farkas in 1902, von Neumann in 1937, and, especially, Kantorovich in 1939 [see Dantzig (1963, Chapter 2) for details].

As this brief history shows, operations research uses techniques and approaches from many disciplines. A persistent problem for the field has been a labeling one, in which it has often proved difficult to determine how operations research is distinguished from various branches of applied math ematics, physics, or engineering, or (more recently) computer science and artificial intelligence.

Prize-winning Works

The Lanchester Prize of the ORSA is given annually to the book or paper (in the English language) judged to be the most outstanding contribution to operations research during the year of publication. Since the award's inception in 1954, 39 publications have been honored as Lanchester Prize winners. Using a rough categorization scheme, 15 of these publications have been general in nature and 24 have focused on applications in a particular industry or service category. Of the 24 applications-driven winners, 14 (or 58%) are clearly directed at services industries. These include applications in transportation (Leslie Edie, in 1955, was awarded the first Lanchester Prize of ORSA for his 1954 work on traffic management over the bridges and through the tunnels of the Port Authority of New York), banking transactions, university operations management, urban systems, library management, communications, criminal justice, logistics, postal operations, and health care provision. Of the remaining ten applications-focused winners, only two were prompted by problems from the manufacturing floor; the other eight were directed at less easily categorized problem areas: search theory, inventory management, military operations, mining exploration, and purchasing policies (data provided to the author by ORSA).

Each year The Institute of Management Sciences (TIMS) selects up to six finalists for the Edelman Award for Management Science Achievement. Each entry is judged on its use of operations research and management science techniques within an implementation context in which real dollar savings or service level improvements or both are reported and verified by external referees. Of the most recent 28 finalists, 19 (or 68%) are clearly in services: transportation and logistics, 8; financial planning and asset management, 3; marketing and sales, 3; urban services, 2; work force planning, 1; postal services, 1; general corporate planning services, 1. Only 3 are motivated directly by concerns in manufacturing. The remaining 6 focus on water systems, inventory, and energy systems.

This review of OR/MS prize winners was not meant to imply that operations research is unimportant in the goods-producing sector. Many services such as logistics are often "services in support of manufacturing." In specific manufacturing processes, OR/MS can provide design and analysis tools to assist in the engineering design and operations control of those processes.

A Roadmap

In the following three sections I have selected cases from three general areas of OR/MS application to illustrate the variety of contextual settings for

OR/MS, the driving forces behind the decision to invest in OR/MS, the types of products that emerge, some broader organizational impacts of OR/MS, and estimates of the return on investment in OR/MS. The first and third topic areas are directly focused on services industries: logistics and work force planning. The second is "production-related services," selected to indicate how OR/MS provides technical engineering service in the goods-producing sector. Most of the cases are drawn from the open literature, particularly the flagship OR/MS applications journal *Interfaces*. Several were nominated for the prestigious Edelman Award, thus no claim is made that the sample is "random" in any sense. On occasion the discussion is augmented with information provided to me by the author(s).

The space-constrained limited descriptions of these cases do not give a full picture of the institutional and organizational factors that come into play in OR/MS work. Thus I have included, as an appendix, a more detailed case in the area of work force planning. It focuses on the scheduling of emergency telephone (911) operators in New York City.

Following the cases I offer a suggested set of conclusions and discuss the problem of estimating the value added by OR/MS, particularly focusing on how managers make the decision to invest in OR/MS.

ILLUSTRATIVE CASES IN DISTRIBUTION AND LOGISTICS

One of the most successful areas of application of operations research has been in improving operations of spatially dispersed systems. Usually the problems focus around issues of transportation, deployment of vehicles, location of facilities, design of service territories, and inventory management. "Distribution/logistics" is the label we assign to these types of problems.

Tactical Planning

A recent well-publicized example involves the efficient routing and scheduling of trucks delivering industrial gases (nitrogen, oxygen, and argon) to spatially dispersed customers. In the production and distribution of industrial gases the major costs are due to electricity (to separate the gases from the air) and to distribution. Typically distribution costs amount to 30 to 40 percent of total direct costs. In day-to-day operations a dispatcher matches customer orders with available trucks and drivers, tries to identify other customers who may benefit from a delivery, and attempts to devise cost-efficient routes for the trucks while not violating any one of myriad constraints (e.g., Department of Transportation-imposed constraints on maximum allowable driver time per trip). The potential number of combinations is often enormous, and operating in manual mode the dispatcher must rely on experience and intuition to devise the trip assignments.

In 1983 Marshall Fisher (University of Pennsylvania) and his colleagues devised an operations research procedure based on mathematical programming techniques that allowed the testing and sorting of thousands (perhaps millions) of trip combinations; the new procedure produced solutions typically 10 to 12 percent less costly than those produced manually. In implementation at Air Products, Inc., the reported savings were reduced to approximately 6 percent, still representing substantial dollar volume when projected over the entire corporation. This work was honored by TIMS, which awarded the authors the Edelman Award for excellence in the practice of management science (Bell et al., 1983). The work has been favorably reviewed in the *Wall Street Journal* , the *New York Times*, and elsewhere. [Also see Bodin et al. (1983) for additional references in vehicle routing and scheduling.]

The "product" of the Fisher et al. work is a complex computer program that must be executed at least once daily to devise the next day's trip assignments. The success of the implementation in Air Products, Inc., is due in part to corporate commitment from top management to quantitatively based, computer-implemented tools for improving operating efficiencies. The record of success and failure of similar attempts indicates clearly that a necessary condition for implementation success is the existence of a broadly based constituency within the organization (including top management) supporting the effort and able to maintain the delivered product after the operations researchers have left the scene.

Not all operations research products in the logistics area are sophisticated computer programs. In 1983 John Bartholdi and Loren Platzman (Georgia Institute of Technology) studied the problem of "meals on wheels." In this application a charitable organization in the inner city delivers meals daily to elderly and infirm individuals in their homes. The organization is staffed with a combination of volunteers and near-minimum-wage employees; resources are scant, and a state-of-the-art computer for solving complex mathematical optimization problems is out of the question. Yet distribution costs represent a large fraction of direct costs of operation, and even casual observation of operations revealed that then current methods of distribution were far from optimal. Ingeniously, applying some ideas from the mathematical field of "space-filling curves," the authors devised a scheme for assigning drivers to vehicles and vehicles to routes that (1) only required two card files (no computer); (2) produced solutions vastly superior to previous solutions; and (3) naturally included certain operating realities, such as, the fact that the number of drivers showing up for work on a given day is a random quantity. As in the Air Products case, the procedure must be performed daily, but it only requires 15 minutes; before the procedure was introduced, 3 hours had been consumed daily to construct routes. The devised operations research procedures are now used widely throughout the United States for delivering "meals on wheels" and in a variety of commercial endeavors as well (package delivery, supplying fresh pastries to restaurants, servicing banks' automated teller machines) (Bartholdi et al., 1983; Bartholdi, private communication, 1987).

Strategic Planning

Whereas routing of industrial gas trucks and vans for meals on wheels required vastly differing computational power, each represents an example of tactical or operational planning in a logistics setting. Many of the successful operations research implementations in logistics have been of this type, namely involving near-term decisions in an operational setting. Yet perhaps a more powerful area for application in logistics is in strategic or long-term planning.

Rails Consider the case of the Boston and Maine Railroad. Between 1977 and 1982 the Boston and Maine Railroad made extensive efforts to improve its operating performance, especially in the areas of freight service, terminal control, and freight car utilization. In the arena of long-haul transportation, railroads achieve their competitive advantage by using a single locomotive to pull a great many freight cars. The "operating plan," the most fundamental control at the disposal of the railroad, governs the movements of cars and trains (blocking policy, train schedules, and dispatching policy). Procedures for developing and modifying operating plans are extremely important aspects of the railroad control system. Specific problems arise in establishing consistent standards for yard, train, and system performance, due in part to availability of only aggregate information and to the limitations resulting from analyzing each train and yard somewhat independently, despite their clear interdependence.

As reported in 1986, Carl D. Martland and his colleagues (MIT) developed the "Service Planning Model" (SPM) that estimates the service and cost impacts of alternative railroad operating plans. Using available data on traffic flows over the network, costs, operating constraints, and parameters describing the proposed plan, SPM estimates yard performance, trip times, aggregate performance per user-defined traffic categories, and numerous types of costs. As a consequence of analyses conducted with SPM, major changes were made in the organizational structure and decision-making processes of the company, as well as in physical facilities and information systems. Savings attributable to this effort amounted to more than $3 million annually, or roughly 3 percent of total operating expenses (Martland et al., 1986). According to Martland, "The Service Planning Model in and of itself did not cause the benefits, but provided an impetus to create an effective interdepartmental planning process" (Martland, private communication, November 1987).

Operations research has produced other major strategic planning impacts in the rail industry. As another example we consider the Canadian National (CN) Railway, Canada's largest railway. In 1984 CN Railway's total traffic volume was 174.3 billion gross-ton-miles, which generated C$3.8 billion in revenues, resulting in a net income of C$304 million. In the late 1970s traffic volumes through 1990 were projected to double on CN Railway's already congested single track main line. Faced with complex constraints that were physical, financial, and operational, CN Railway embarked on a "Plant Expansion Program" (PEP) whose implementation would handle the increased traffic of 1990 while maintaining existing levels of service. Proposals called for capital expenditures during the 1980s of C$2.2 billion, of which C$1.3 billion would provide double track in some congested links of the system.

After analyzing state-of-the-art line-capacity methods, CN Railway decided to develop detailed simulation models to estimate the capacity of transport on segments of the line. An important component of the analysis was the Signal Wake Model that determines for a given configuration of signals the minimum train headway that can be maintained as a given fleet of trains follow each other in the same direction over a specified track layout. Another component was the Route Capacity Model that estimates train delays by simulating operations of trains over a rail line under specified track maintenance activities. The resulting analyses produced a package of cost-effective improvements for capacity expansion, which included a combination of control technology (closely spaced signals) and strategically located sections of double tracks. The major cost savings of the analysis was the identification of 128 miles of track, originally slated for expansion to double track, that with extra signaling could remain single track until after 1990. This allowed CN Railway to defer C$350 million in capital expenditure beyond 1990 (Welch and Gussow, 1986).

Banking Banking is a major (financial) service industry that is not usually associated with logistics. However, as the case of BancOhio demonstrates, logistical concerns can play a key role in efficiency of banking operations. Partially due to the relaxation of branch banking restraints, U.S. banks are increasing their branch networks. The wider geographical dispersion of bank branches can complicate the check-processing function resulting in the need to determine (1) how many operations centers should be used and where should they be located; (2) which branches should be served by each center; and (3) what costs and performance measures should be included in evaluating alternatives?

From an operations research point of view the flow of checks through a bank can be viewed as a "pipeline inventory model." Items are input at various entry points (banking offices), flow through the processing pipelines and exit in the form of outgoing cash letters dispatched to clearing banks.

The problem becomes complex due to external time constraints imposed on outputs of the system (clearing deadlines). Transportation is a significant component of the system as checks are moved from receiving branches to encoding sites to capture sites.

In January 1984 the BancOhio network consisted of 266 branches representing 42 individual banks. Checks were encoded in 31 of these locations and eventually transported to one of two capture sites (Columbus or Cleveland) for computer processing and clearing. Management felt that centralizing processing facilities would achieve economies of scale, and initiated an operations research analysis to determine the validity of their views.

The analysis used a check-processing simulation model (CHECKSIM) to generate efficient transportation routes for the messengers who pick up checks and deliver them to the processing center. The simulation was run for each processing center configuration under consideration. The analysis showed the expected result that consolidation would produce economies of scale, but perhaps even more importantly, that even greater savings could be accrued by moving certain ancillary support functions to consolidation centers. In fact, the savings in transportation and encoding efficiencies ($287,000 per year) were dwarfed by savings associated with transferring certain retail and operations functions ($1,381,000 per year). The total identified savings represented approximately 9 percent of then current operating expenses. In implementation the most difficult problems were associated with reassignment (and displacement) of personnel (Davis et al., 1986). According to Davis, the OR/MS implementation effort lasted 270 days and cost BancOhio $80,000 (Davis, private communication, November 1987).

Urban Services The use of operations research in logistics is not confined to the private sector. Let me briefly cite two projects that I recently directed involving agencies of New York City.

The first was with the New York City Department of Sanitation (DOS). In 1981 DOS was confronted with imminent closings of major in-city landfills, resulting in a projected doubling of daily refuse tonnage transported by barge to the world's largest landfill on Staten Island (Fresh Kills Landfill). The strategic planning rule then "in good currency" was to "size" the fleet of barges in direct proportion to daily tonnage carried. If tonnage doubles, barge fleet size should also double, according to this tradition-based rule-of-thumb. If the fleet size were to double the city would have had to purchase an additional 40 barges, estimated then at $1 million per barge, representing a potential commitment in capital expense of $40 million.

Not willing to trust the "linear rule-of-thumb" for such an important decision, DOS commissioned an operations research study to determine the required fleet size to handle the projected new loads. The study resulted in the creation and implementation of a simulation model Barge Operation

System Simulator (BOSS). DOS personnel, while performing numerous production runs with BOSS, identified a savings through 1990 of at least 10 barges and perhaps 20 barges without impeding service levels. The savings were not unexpected by the operations researchers who saw the barge and tug refuse transportation system as a closed "multiserver queuing system;" such systems almost always display economies of scale, in which workloads may be increased and servers (i.e., barges) need not be increased in direct proportion, while still maintaining similar (acceptable) performance characteristics.

At the time of contracting for new barges (1983) the ship-building industry was severely depressed, resulting in new barge purchase cost of only $600,000 per barge; hence, as a result of using BOSS for fleet sizing, the city has saved $6 million and may save an additional $6 million (if an additional 10 barges are not ordered in 1990). BOSS cost New York City $100,000, yielding an immediate return on investment of 60:1. (Larson et al., 1988).

New York City's Department of Environmental Protection (DEP) commissioned an operations research analysis in September 1985 to provide a computer-based tool to help DEP planners design a new logistics system to transport sewage sludge to a new ocean dumping site. Sewage sludge is the final product of primary and secondary sewage treatment; it is 97 percent water, 3 percent solid and has a specific gravity of 1. For decades New York City had been dumping its sewage sludge a few miles outside New York's harbor entrance. In 1983 the federal Environmental Protection Agency (EPA) placed New York City under court order to commence a scheduled process whereby eventually all of the city's sludge would be transported to a new EPA-designated site approximately 106 miles south southeast of the harbor entrance. It is at this "106-mile site" that sludge is to be dumped in the future.

After requesting bids to transport the sludge from "private haulers," DEP decided that the new sludge transport system should be primarily under DEP's (not a private hauler's) control. The commissioned operations research model was to be able to depict alternative ways of operating the new sludge transport system, including computation of costs and performance characteristics of alternative fleet sizes and fleet mixes, use of transshipment points, impact of dredging and other capital improvements, and increased land-based sludge storage capacity at one or more sites.

The model that was ultimately developed, Strategic Logistical Unified Design GEnerator (SLUDGE), accomplished all the desired tasks; it operated on an IBM PC AT desktop computer and required only approximately 2 to 3 seconds for each production run. DEP planners have executed the model in production runs well over 1,000 times in determining the appropriate type and size of vessel to assign to the oceangoing link and to the inner harbor. SLUDGE has also been used to determine the best locations for a primary and a backup transshipment site. It is currently being used to "fine tune" the design of the inner harbor system.

Although it is difficult to identify a precise "savings" accrued by New York City as a result of using the model, the dollar magnitude of the decisions being made is significant. For instance, the result of the first set of production runs was to design the oceangoing fleet as four 15,000-ton oceangoing barges (towed by special tugs under contract to DEP); the size of the barge construction contract was $21 million. Other decisions of almost equivalent dollar consequence are currently being made. The cost to New York City for the model (and accompanying analyses) was $330,000 (Larson, 1988).

CASES IN PRODUCTION-RELATED SERVICES

The efficient production of things often requires careful coordination and timing of flows of materials, subassemblies, and the like through a complex sequential (and at times, parallel) process. Due to the logistical nature of many production processes, one is not surprised to learn that operations research has played a significant role in the design and operation of production processes.

For years the "production paradigm" has been successfully used to model certain services industries [see, for example, Heskett (1986)]. A well-known Harvard Business School case cites the design and operation of Benihana Restaurants as following a production process, where in effect the customer is the item being processed ("Benihana of Tokyo," HBS Case No. 9673057). ORSA awarded its 1985 award in the practice of operations research to Burger King restaurants who used production processing ideas to develop mathematical models of alternative Burger King operations; among the innovations adopted was the decision in many high-volume restaurants to have customers "make their own drinks," a very labor-intensive activity.

Production of Hydroelectric Power

But operations research has played major roles in more standard aspects of production. Consider, for instance, the production of hydroelectric power by Pacific Gas and Electric Company (PG&E), the world's largest privately held utility. Supplying gas and electricity to more than 10 million people over 94,000 square miles in northern California, PG&E had operating revenues in 1984 of $7.8 billion. PG&E generates electricity using a mix of hydropower, fossil fuels, nuclear energy, more wind energy than any other utility, solar power, the world's largest geothermal steam engine, solid waste, and biomass.

The Sierra Nevada hydrosystem provides more than 20 percent of the electricity the company sells, representing 16 billion kilowatt-hours, the demand of 2.5 million homes. If PG&E had to burn fossil fuel instead of using hydropower from the Sierra Nevada system, an additional 20 million. barrels of oil would be required in an average year. The efficient management of the system is necessary to provide energy at low cost. The magnitude of the management problem becomes clear when one considers that the system includes 86 hydropower plants in 23 river basins, covering an extensive region with marked variability in water supply due to seasonality, randomness of weather, and other uses of the water (e.g., drinking and irrigation). Other complications include the cumulative effect of flow in the river basin and the fact that energy produced depends on the pressure of the water flowing through the turbines, producing a nonlinear relationship with total rainfall.

Over a 2-year period PG&E developed an operations research software package called HYSS, whose purpose is to compute optimal water release policies for each of the 86 plants in the basin. The objective of the mathematical model is to maximize the total megawatt hours of electricity generated over a 1-year planning horizon. HYSS has been used to compute release schedules for PG&E for more than 3 years. According to company estimates, the value of the increased energy attributable to the use of HYSS schedules is $10 to $45 million per year. In addition to improved operations through release scheduling, HYSS has contributed to better resource planning, particularly with regard to scheduling of construction projects. As an illustration, HYSS was used to show that PG&E could generate $13 million worth of additional hydropower if construction of the Kerckoff 2 plant could be completed 2 months early, due mainly to California receiving record quantities of rain and snow that year (Ikura et al., 1986)

A somewhat different but equally complex energy planning model was adopted by 18 utilities in Brazil to determine optimal allocation of hydro and thermal power-generating resources in the system. The model stochastic dynamic programming (SDP) was extensively validated by the 18 utilities before adoption. Comparisons with a previously adopted model projected savings of $87 million in 5 years, a 28 percent reduction in generating costs. Actual (measured) savings from 1979 to 1984 are reported to be approximately $260 million. The OR model required 540 days to develop and implement, at a cost of $50,000. Like the PG&E case, the model originally developed for near-and mid-term operational planning is now the focal point for longer term generation expansion-planning activities (Terry et al., 1986; Terry, private communication, January 1988).

For those further interested in the use of operations research models for hydroelectric power management and, more generally, in water resources management, see Goeller et al. (1985) who report on a 125-person-year effort in the Netherlands for nationwide water resource planning. This path-breaking effort, initiated in 1977, has already resulted in reported savings in excess of $50 million per year.

Production Scheduling

More traditional use of operations research methods in production is in the coordination of a wide range of manufacturing-related activities. Production planning can be seen as a hierarchy of managerial decision-making activities. The hierarchy ranges from strategic planning through tactical planning to operations control. Hierarchical integration of production planning, scheduling, and inventory control is required to coordinate organizational levels responsible for developing and executing plans.

Owens-Coming Fiberglas (OCF) produces in a large facility in Anderson, S.C., a variety of mat products, sold in rolls in various widths and weights, treated with one of three process binders, and perhaps trimmed on one or both edges. There are two parallel production lines with distinct characteristics, producing 200 distinct mat items, with 28 of these products (the "standards") representing more than 80 percent of total demand and the remaining low-volume products called special orders.

The scheduling system developed for the mat line addresses the interaction between aggregate planning (relevant costs for work force, overtime, and inventory), lot size determination (production line quantities, line assignments, and inventory levels for each product), and ultimate job sequencing (for standard and special order items). The methods, all computer-implemented, range from simple "back-of-the-envelope" approximations to mathematical optimization modules involving thousands of variables. The key optimization module uses derived aggregate monthly inventory levels and individual standard product demands to generate inventory levels, lot sizes, and line assignments for each of the standard items. The resulting scheduling covers a 3-to 12-month planning horizon. OCF used the module to schedule over 20 million pounds of mat production during 1981–1983. As a result the average number of monthly production changeovers decreased from 70 in 1981 to less than 40 in 1982/1983, resulting in an estimated savings of $100,000 or more. Operating efficiencies improved dramatically during the same period of time, although net effects are difficult to quantify (Oliff and Burch, 1985). The OR/MS development cycle lasted about 120 days, costing OCF approximately $30,000 (Oliff, private communication, December 1987).

In 1983 Monsanto opened a second production plant (this one in Pensacola, Florida) to produce maleic anhydride, a chemical used in polyester resins, oil additives, agricultural chemicals, and fumaric acid. The sudden additional production capacity mandated close coordination between and within the plants to minimize costs. Three operations research models were developed, one for each plant and a global model combining both plants. The effort was similar in spirit to that of Owens-Coming, but with one more level of hierarchy due to multiple plants. The total time spent on all three models was approximately 1,000 person-hours, with a direct cost not exceeding $50,000.

The direct cost savings were estimated to fall between $1 and $3 million per year. Additional cost savings were derived from using the model in the following areas:

- the modeling system is used to evaluate many operating policies that would cost thousands of dollars per policy to evaluate in engineering time
- the system was used to determine whether or not certain compressors should be repaired, the decision yielding a net savings of more than $250,000
- the possibility of using vent steam in summer months to run the compressors was analyzed, resulting in a savings of more than $100,000 (Boykin, 1985)

For another example in production planning see Liberatore and Miller (1985) who describe a system implemented at American Olean Tile Company. The system reportedly required less than $100,000 to develop and is saving $400,000 to $750,000 in distribution costs annually, representing approximately 5 percent of the variable costs of production and distribution (Liberatore, private communication, December 1987). It has also resulted in improved coordination and communication between manufacturing and marketing and the development of new sales forecasting procedures.

Programming Sophisticated Machines

Weyerhaeuser is one of the world's largest forest products companies. In 1984 company revenues were more than $5 billion, predominantly through domestic and foreign sales of logs and timber, lumber, plywood, and paper products. Forest products is primarily a commodity industry, which means there is little control over the prices realized from the sale of products. Together with a very competitive environment (in 1984 profits averaged 2.5% of sales for large firms), it is imperative to use raw materials efficiently. Achieving improved use of raw materials meant seeking the best use of each individual tree (i.e., how a tree is cut at the mill). Depending on the decisions that the cutting machine operator makes, the value of the tree can vary by 50 percent or more. Weyerhaeuser cuts approximately 15 million trees annually, approximately 100 trees per minute, and there are hundreds of cutting machine operators at dispersed locations.

In response to such a challenging problem, M. Lembersky while working at Weyerhaeuser developed over several years a sophisticated "product" called VISION. The first component of VISION is a dynamic programming optimization procedure that determines the best economic use for any tree. The second component is a video-game-like computer system that allows mill personnel as well as company managers to grasp easily the best use of any given tree. Operated together, the result is better decisions in practice. VISION allows its user to make his or tier own decisions, see what revenues they yield, and compare them with mathematically derived optimal decisions in a very user friendly environment. Implementation of VISION was not without its difficulties, particularly convincing top management to devote sufficient resources to its development and to train and motivate field workers in its use.

VISION has been used at Weyerhaeuser since 1977, yielding dramatic results. The value added through additional profits through 1985 is estimated to exceed $100 million. In addition, VISION has had important ancillary impacts up and down the line. At the operational level it has helped workers to adjust quickly to changing operating conditions in the field. At the top corporate level, VISION has been used by George Weyerhaeuser (chief executive officer) and others who have committed millions of capital dollars to in-field real-time implementation of VISION and related (subsequent) operations research products in such different areas as facility design and truck routing (Lembersky and Chi, 1986).

The mere use of operations research ideas does not guarantee success. The model may be inappropriate, the data too costly to collect and update, or aspects of organizational structure may inhibit proper implementation. One of my favorite examples from the last category involves a Boston-area computer manufacturer.

One particularly complex large printed circuit board required more than 10,000 holes to be drilled in it at precise but irregular locations. The time to drill each board using a state-of-the-art programmable electric drill exceeded 1 hour. This total "service time" was roughly 75 percent drill time and 25 percent drill movement time, the latter required to position the drill for the next hole. Anyone who watched the machine operate could see that the drill was routed "all over the place," crisscrossing previous paths, thus significantly increasing total service time per board. Yet when one spoke with technical staff associated with programming of the drill, they insisted that a well-known operations research "heuristic" solution to the "traveling salesman problem" was used and that the machine should be operating near optimally. This heuristic is nothing fancy, being called the "nearest neighbor heuristic," meaning that the drill, when finished with one hole, would be routed next to the nearest undrilled hole location. A drill properly programmed with the nearest neighbor rule would not meander all over the board several times as could be observed in practice.

The solution to this puzzle was found when it was discovered that the four different device types to be placed were the responsibilities of four different engineering groups within the firm. Each group, with the help of an in-house optimization group, had programmed the drill for its "own holes," oblivious to the need to coordinate their activities with the other groups. In particular, the in-house optimization group never realized that the same drill bit was being used for all four hole drilling processes. As a result, the drill implemented each group's instructions sequentially: the machine would first drill "device type A's" holes, then "device type B's," then "C's," and finally "D's," resulting in the observed meandering (A. Port, private communication, 1987). As a consequence of such a "suboptimal" programming of the machine, one could reasonably expect the drill traveling distance to be approximately 70 percent greater than necessary (Larson and Odoni, 1981, p. 408). The percentage reduction in drill positioning time was attenuated somewhat by the "fixed" effects of acceleration and deceleration in the positioning process. Still, attainable improvements in the total processing time per board ranged from a few percent to more than 10 percent, not insignificant considering the cost ($150,000) per drilling machine.

The failure of operations research here was in fact a failure for like groups to communicate and coordinate their activities at a focal point in the manufacturing process where their interests merged. Although this example is small in scale, it illustrates a general condition that when projected over the entire organization implies an urgent need for new forms of internal communication in order to stay competitive.

At Metelco S. A., a medium-sized manufacturer of printed circuit boards near Athens, Greece, the drilling machines prior to March 1983 were often sequenced in the order of coordinates of holes specified by the company's customers. There, too, drilling of a single board often consumed more than 1 hour. Magirou (1986) reports that he developed a "nearest neighbor heuristic" and implemented it on Metelco's programmable drills, thereby increasing average throughput by more than 10 percent. Metelco has reportedly saved at least $10,000 per year because of fewer machine operators' hirings than planned. A. J. Nicolitas, managing director of Metelco, reports additional side benefits, especially by saving "money by avoiding common human errors, i.e., multiple drillings of the same hole," and in making "our management aware of the potential benefits of the interface between management science and electronic hardware" (Magirou, 1986). This "software" technology has recently been "hard coded" on a microchip and the chips are being sold throughout Europe as device controllers for programmable drilling machines (A. R. Odoni, private communication, 1988). This example illustrates as well as any the interchangeability between software and hardware, and between services and goods.

ILLUSTRATIVE CASES IN WORK FORCE PLANNING

One of the most important areas of application of operations research to the services industries has been in deploying and scheduling of personnel. Various work force planning "packages" have been developed in such widely divergent services as retail sales, air transport, urban services, telemarketing and telephone sales, clerical services, restaurants, and banking. Like logistics and production-related services, work force planning relates to the scheduling (and sometimes movement or placement) of "allocable resources," in this case people. The potential returns in this area are often sizable due to the labor intensiveness of many services; for instance, work force scheduling procedures have been developed for services whose costs are 95 percent attributable to salaries and related fringe benefits.

Scheduling

In 1984 United Airlines recorded earnings of $259 million on a revenue base of $6.2 billion. The profitability resulted in part from an ambitious expansion plan implemented in the previous year which brought about a 6 percent growth in revenue with only a 2 percent growth in costs.

A major factor in cost containment is the airline's newly created computer-based work force planning system for scheduling shift work at its reservation offices and airports. United's 11 reservation offices employ more than 4,000 reservation sales representatives (RSRs) and support personnel, with requirements for work force determined by a forecasting of call volumes based on historical trends and a queuing model to determine (at a given demand level) the number of employees to provide the desired level of service. Also covered in the developed Station Manpower Planning System (SMPS) are the 1,000 customer service agents (CSAs) at its 10 largest airports, the CSAs divided between counter and gate employees. Although work force requirements vary widely by time of day and day of week, work rules require employees to have the same starting time every day and to work the same shift length every day. SMPS uses developed requirements for 30-minute intervals over a 7-day period to produce monthly shift schedules. The system uses state-of-the-art mathematical programming techniques and encompasses the entire scheduling process from forecasting of requirements to printing employee schedules.

Because of a company-perceived urgent need for an effective scheduling tool, not enough time was initially devoted to involving employees in developing the new procedures. Although economically "optimal" in some (infeasible) sense, the model's exclusion of factors important to employees delayed implementation until 1983. However, since 1983 SMPS has been used to develop work schedules for 4,000 employees on a regular basis and is eventually expected to schedule 10,000 employees or 20 percent of United's total work force. The system has produced savings of more than $6 million annually while earning strongly positive reviews from United's upper management, operating managers, and affected employees. Hard-to-quantify capital benefits include the following:

- additional revenue generated by improved service levels
- benefits from the use of SMPS in contract negotiations
- savings from reduced support staff requirements
- savings from reduced manual scheduling efforts
- reduced training requirements (Holloran and Byrn, 1986).

According to Holloran, the development cost of SMPS was approximately $500,000, allocated over a 260-day development period (Holloran, private communication, November 1987).

Services associated with "income tax" time are highly seasonal, requiring careful scheduling of personnel during short peak work load periods. For the Financial Services Group (FSG) of Canada Systems Group, Inc., the "season for giving" to one's tax-deductible retirement fund apparently lasts approximately 6 weeks (late January to early March). In 1984 the incremental cost to FSG for managing this intense period was approximately $500,000. Prior to the following season, the FSG developed a linear programming work force planning tool that, based on projected work loads, developed hiring needs and shift assignments for the 6-week period (well in advance of that period). The incremental cost of managing the 1985 season, with the new tool, was reduced to $170,000, a 64 percent reduction, despite somewhat higher wages and a 25 percent increase in volume. As has been shown to be common with other operations research installations, other intangible benefits were also reported, particularly an enhanced reputation for reliability of service that has resulted in successful acquisition of new clients and all but one major client renewing contracts (Haehling von Lanzenauer et al., 1987).

Spatial Deployment

Work force planning may also relate to the allocation of personnel over service territories. In fact the subfields of "optimal location" and "optimal districting" are two of the most active fields in operations research.

In the 1970s the literature of both operations research and marketing began to offer detailed consideration to the use of mathematical programming models to assist in sales territory design decisions. Models were developed to allocate work load among a fixed number of salespersons, to calculate the best number of salespersons, and to determine territory boundaries. Later refinements dealt with constraints on time limitations of the salespersons, supervision, salesperson experience and competition, as well as adjustments to boundaries taking into account natural obstacles.

The Houston-based Variable Annuity Life Insurance Company (VALIC) markets annuity contracts to not-for-profit organizations and governments. In 1982 VALIC decided that it needed quantitative guidance in the design of its service territories and in the structure of its field offices. At that time there were 336 salespersons nationwide, allocated over 16 regions, each with a manager and an office. Among the management issues were the number and design of sales territories and regions, while considering equity of "market potential" in each and morale problems associated with redesigns. An operations research analysis commenced whose purposes were threefold: (1) to determine the cost associated with the current 16-region configuration; (2) to determine the lowest cost solution in both number of regions and their geographic configuration; and (3) to estimate expected cost savings if the change in configuration were to be adopted. The analysis presented an interesting trade-off between fixed and variable costs, the fixed costs associated with regional offices and the variable costs associated with intraregion travel times.

The first use of the resulting program was focused on the then present regional configuration, showing a model-derived cost of $18,826,000. Also, it was determined that by closing one regional office and moving a few regional boundaries, VALIC could reduce total costs by 4 percent. More interesting was the cost difference when the number of regions was allowed to vary. The total cost of the solution resulting from 25 regions was $9,933,000, a savings of $8,833,000. Not surprisingly there were obstacles along the way, such as initial results violating constraints on disproportional market potential among regions and the apparent uncaring attitude of the company toward changing the locations of current regional offices. As of the time the case was reported (1984), VALIC had launched a 5-year phase in of the resulting "fine-tuned" recommendations. Management appeared confident in the projected cost savings but had decided to "go slow" in the sensitive area of personnel relations (Gelb and Khumawala, 1984).

The deployment of ambulances throughout a city represents a totally different type of spatial deployment problem. The reader is referred to Brandeau and Larson (1985) and Eaton et al. (1985). In redeploying ambulances in Austin, Texas, Eaton reports that his $30,000 study saved the city $10.8 million over the following 7 years; ambulance first response times decreased 7 percent in the face of a 52 percent increase in demand. This example, coupled with the earlier "urban services" cases, demonstrates that significant returns to investment in operations research are available in the public as well as private sectors.

CONCLUSIONS

What have we learned from our tour of OR/MS applications in services industries and in production-related services? I would like to offer the following:

- An OR/MS product can assume many forms, from a computer program implemented in color graphics to a consultant's report, to card files, to a "smart" machine tool, to an educational "video game." The numerous embodiments contribute to the field's fuzzy image.

- In implementation, often unanticipated "side benefits" of an OR/MS effort will dominate the benefits accrued in the original target area of the work. Such side benefits can be limited to additional unanticipated cost savings (or service enhancements), or they can extend more fundamentally into managerial structure and flow of information. In the latter case the potential magnitude of the effect of OR/MS is often accompanied by a comparably large organizational resistance to change.
- The costs of implementing OR/MS can vary enormously, from $50 per implementation for "meals on wheels" to millions of dollars for large-scale, multiple-site decision support systems.
- The reported benefits of OR/MS work are often one or two orders of magnitude greater than the costs.
- In-the-field knowledge of even rudimentary properties of operations research models (and thus of operating systems) is often lacking.
- Managers do not like to state explicitly target service levels that implicitly admit to failure a certain fraction of the time (i.e., "probabilistically stated objectives").
- Markedly successful OR/MS efforts seemed to be accompanied by (1) top level corporate enthusiasm and long-range commitment and (2) involvement of operations personnel during implementation.
- Due in part to the amorphous nature of its products and the highly technical nature of its process, as a profession OR/MS runs the risk of being absorbed by related and more easily identifiable fields such as computer science.
- Although there are numerous OR/MS "success stories," several of them reported here, the overall market penetration of OR/MS in services remains shallow. The field's limited impact to date may be due to excessive academicism in the field, fear of technical approaches by operating managers, need until recently to use mainframe computers, and exclusion by many operations researchers of broader nonmathematical aspects of the problem.
- Operations research offers the potential for great productivity improvement in languishing services sectors, improvements often greater in percentage terms than those typically associated with the manufacturing sector.

VALUE ADDED FROM OPERATIONS RESEARCH

Is it possible to estimate a priori the "value" of any proposed OR/MS effort? Investment in capital equipment is a familiar activity of U.S. corporations. By now standard techniques exist to estimate costs and benefits of many "hardware" investment alternatives, including discounting cost/ benefit streams into the future, estimating time until the investment is recouped, and computing total (discounted) returns on investment. But considerably less is known about investing in various kinds of "software" technologies. Occasionally, as in "desktop publishing," the savings and productivity improvements are so demonstrable that the investment is clearly a good one. More problematic are investments in software and services that are aimed to improve "decision making," either at high corporate levels or at operational levels. Examples would include software/services for investment planning, assembly line balancing, airline scheduling, allocation of marketing dollars, analyzing potential new markets, deploying work forces, analyzing customer service satisfaction levels, and designing a new logistics system. Many of these types of decisions are based on intuition and methods derived from "years of experience."

When I contacted them in relation to this paper, OR/MS practitioners and researchers were doubtful that any formal mechanism could be devised. According to Y. Sheffi [a well-known logistics specialist and coauthor of a case reported for Marshalls, Inc. (Carlisle et al., 1987)],

> Estimation of cost/benefit: the burden is on me and a project champion in the organization. . . . Mostly, no formal analysis is undertaken as decision maker in the organization gets finally convinced by hand waving. No specific time is used to recoup costs. The potential has to be enormous—otherwise the project is not done. In other words, the benefit/cost ratio has to be (an implied value of) 20–200 or people to feel comfortable (Y. Sheffi, private communication, October 1987).

The emphasis on large benefit/cost ratios was repeated by others. According to Amedeo Odoni (a recognized expert in OR/MS as applied to airport planning),

> In my experience I have not really encountered any formal mechanisms for evaluating the costs and benefits of an OR study. The reason may be that the benefits are usually of a different order of magnitude than the costs (e.g., in a $100K study of an airport's layout, one may "save" $25 + million, a real example). Airport benefits are also often difficult to quantify in dollar terms (Amedeo Odoni, private communication, November 1987).

According to John Bartholdi (coauthor of the "meals-on-wheels" project),

> In my consulting experience costs/benefits must be clear and large before a client undertakes action. [He or she] expects immediate payback (or at least within I year) and wants insignificant risk. Future is not discounted, since action not taken unless improvement will be enormous (John Bartholdi, private communication, November 1987).

There appear to be some settings in which "scientific" a posteriori evaluation of an OR/MS product is possible. For instance, in logistics, if the concern is solely transportation cost reduction, one can analyze the decisions of truck dispatchers with and without the OR/MS technology to estimate savings. This was apparently the approach taken in the celebrated Air Products case. According to the principal author of the work, Marshall Fisher, with regard to evaluating the benefits,

> Air Products was quite thorough and methodical in this regard and developed a model of how distribution costs related to various parameters of the distribution operation, such as location and volumes of customer demands during a particular time period. This model was used to predict what costs would have been in the future if the vehicle schedule system had not been introduced, and therefore to provide a benchmark against which to assess improvements (Marshall Fisher, private communication, December 1987).

But after-the-fact evaluation still does not answer a manager's question regarding investment in OR/MS technology prior to demonstrated beneficial results. I particularly like what Warren Powell, an expert in logistics, has to say about this problem area:

> For some companies it (the decision to invest in OR/MS) is a pure cost/benefit decision, using very conservative estimates of cost savings as benefits. For other companies the decision is driven by one or two individuals "with a vision" that a model is critical to success. Operations people generally fall in the first group, marketing/finance types in the second. The person with a vision is critical to implementation.
>
> In reality, success usually depends on making someone's life easier. Rigorously documenting savings is rare (e.g., numbers prepared for the Edelman Award are generally not reliable). Service and profit benefits are virtually impossible to quantify, because side-by-side analyses with and without a model are never available.
>
> It is most tempting to evaluate the value of a model in terms of how much money it saves each year. To be sure, it is a useful and often important exercise to at least try to estimate the economic impacts of a model. . . . The current emphasis on cost numbers is having the result that (a) only implementations at big companies which may yield substantial cost savings are important; (b) traditionally conservative people in operations, who often will acknowledge only savings they can rigorously verify, are to be avoided; and (c) traditional applications to operations, which yield direct cost savings are preferred over richer applications to improve pricing, marketing, customer service, or financial planning with notoriously intangible benefits (Warren Powell, private communication, November 1987).

It may be that no satisfactory formal mechanism will ever be devised for deciding before-the-fact whether or not to invest in OR/MS. For certain narrowly defined applications areas, one can simulate proposed new procedures and compare them with status quo procedures to assess potential benefits. But the cases in this paper illustrate that the greatest potential benefits of an information/knowledge technology such as OR/MS are organizational, affecting fundamentally the ways firms manage and operate. Sometimes the OR/MS model serves as the catalyst for managers from dis parate departments within a firm to communicate; perhaps they should have communicated before creation of the model, but the at once bald and scientifically neutral assumptions of the model can focus a group's discussions on difficult decisions. OR/MS products implemented on a day-to-day basis, by affecting information flows and providing immediate evaluative results of decisions, can markedly change managerial behavior.

In increasingly competitive environments one can argue that the effective processing of data to develop decision consequential information remains for many services industries a viable mechanism for achieving competitive advantage. For the impacts of OR/MS and related information technologies to grow, many managers may need a new point of view regarding investment. According to George Kozmetsky, "Managers need to understand that information, science, and technology are not free economic goods but are assets to be used, planned, earned on, and replenished" (Kozmetsky, 1984, p.4).

ACKNOWLEDGMENTS

I would like to thank the MIT School of Engineering for providing support for developing OR/MS course material (of which this paper is a part) in a new engineering schoolwide undergraduate elective on operations research in engineering. I would also like to thank the National Academy of Engineering for supporting my very productive research assistant, Luiz F. M. Vieira, who is a doctoral candidate in operations research at MIT. Finally, particular thanks are due to Bruce Guile at NAE who carefully read earlier versions of this paper and greatly contributed to its final form.

APPENDIX

A Detailed Case: Scheduling 911 Operators

In 1968 the mayor of New York, John Lindsay, opened the first-in-the-nation big-city "911 system" for responding to calls for emergency service (police, fire, ambulance) from the public. To call the police one formerly had to memorize seven emergency telephone numbers, one for each of seven dispatching zones throughout the city. (And, not insignificantly, one had to know from which dispatching zone one was calling.) The new system allowed a caller simply to dial "911" from anywhere in the city.

Approximately 15,000 calls per day were processed by "911 operators" located at a central dispatch and communications room of the New York City Police Department (NYPD) in lower Manhattan. Within weeks after opening the new facility, complaints started pouring in (by telephone, letters, radio talk shows, and letters to the editor) that the new multimillion dollar system that was supposed to speed processing of calls was plagued with delays. One letter to the editor of the *New York Times* complained of calling 911 on a Saturday evening and waiting more than 25 minutes for someone to answer the phone; eventually the caller thought he might have called the "wrong number," so he hung up and tried again, this time getting an answer after "only" approximately 20 minutes of waiting. These types of complaints forced the police commissioner to put together a study team to analyze the problem.

I was one of three members of the team, the other two being police lieutenants trained in police planning and operations. Quickly we jointly discovered that the hourly average volume of 911 calls varied predictably by a factor of eight or more on a daily basis (and more when measured over a week), while the hourly number of 911 telephone operators varied only by a factor of two (with maximum deployment averaging 25 operators during all hours of the day except the early morning period, 3:00 to 7:00 A.M., when the number of operators dropped to 12 or 13). In other words, hourly deployment of operators virtually ignored predictable changes in call volumes, except during the very quiet early morning hours.

We desired to develop (in 1 month) an easy-to-use scheduling procedure that took advantage of economies of scale. Although the 911 system incorporated several complications not found in more standard telephone answering systems, we found it acceptable as an approximation to apply Erlang's original formulas (circa 1915) describing the operating properties of multiserver queues to schedule the operators.

Two interesting encounters during the implementation process, both with a senior managing police officer, deserve mention. First, when in a formal briefing I displayed graphically the data showing the true (deplorable) state of affairs with regard to queue delays (with 40 percent of callers on Saturday evenings experiencing delays greater than 30 seconds), the senior officer declared that the data I was using were inaccurate; after all, his officer in charge of the Communications Division had informed him that there were few problems and that the loud public outcries were not representative of the service levels being provided. Luckily, the two lieutenants and I had worked together side-by-side within the Communications Division for I month; when questioned by the senior officer, the lieutenants verified the accuracy of the data.

Second, after the presentation of the data, I requested from the senior officer his department's "performance objectives" with regard to 911 operator scheduling. In particular, I requested from him two numbers, T and P, such that operators would be scheduled so that during no hour would more than P percent of the callers incur delays greater than T seconds. For instance, if T and P were set at 15 seconds and 5 percent, respectively, I would use Erlang's formulas to schedule 911 operators each hour so that no more than 5 percent of the callers would experience delays exceeding 15 seconds. At first the senior officer refused to give any such numbers. Then, when pressed, he relented, announcing his department's values: $T = P = 0.00$ (!). Queues with uncertainty can virtually never achieve such perfect operation with a finite amount of resources. Ultimately, we backtracked and rescheduled the department's existing number of operators on a weekly basis, achieving major reductions in delays and "time equity" in level of service (i.e., with all hours of the week having nearly the same delay characteristics) [see Larson (1972)]. Unlike many other operations research studies, this one was implemented in its entirety within I month after completion.

Prior to the study, 17 percent of all 911 calls had been delayed 15 seconds or more (when averaged over an entire month), with terrible congestion at predictable times (e.g., 40 percent of calls delayed more than 30 seconds from 8:00 P.M. to midnight, Saturday evenings). After implementation of the recommendations of the study, no hour of the week experienced more than 5 percent of the calls having 15-second delays. The "product" of the study consisted of seven charts or tables, each containing for a particular day of the week the recommended number of 911 operators to assign each hour of the day. The "cost" to the NYPD was 3 person-months of professional effort. No additional 911 operators were hired; rather the hours of working operators were simply reassigned. If additional 911 operators had been hired under the "old" scheduling scheme to obtain the same new performance levels at all hours of the week, the operator pool would have increased by approximately 50 percent.

Before leaving this case, two other reflections are in order. We found early on in the study that nearly all the data we needed had been recorded meticulously by a full-time officer whose only job was to place the operating statistics into a large loose-leaf book. By 5:00 P.M. each day, he had completed the previous day's entries, inserted the final completed sheet of numbers, and went home. Not one decision had been influenced by the entries in the book! As far as we could tell, virtually no one other than this recording officer had ever looked at the numbers. This book was so complete (and accurate) that for our work we needed only approximately 10 percent of its entries.

To ensure implementation, the two lieutenants decided to modify the incentive and reward system for the captains on duty in the 911 center. The "data recorder's" job was modified so that the first thing he or she did each morning was to draw a large graph displaying in an hour-by-hour fashion the previous day's performance. The name of the captain on duty for each 8-hour tour of duty was prominently displayed, as well. Each time during a tour that one or more callers experienced a call-answering delay exceeding 30 seconds, bells would sound and lights would go on and revolve, not unlike that in modern discotheques. Each such event was labeled a "bell." On the large graph, next to each captain's name, the data recorder displayed in large font the number of "bells" incurred on that captain's 8-hour tour of duty. This display was one of the first things seen by the many New York citizens, school children, and tourists who were given tours of the 911 center each day. Needless to say, the number of bells was kept to a reasonable minimum.

REFERENCES AND BIBLIOGRAPHY

Bartholdi, John J. III, and Loren K. Platzman, R. Lee Collins, and William H. Warden III. 1983. A minimum technology routing system for meals on wheels. Interfaces 13(June):1–8.

Bell, W, J., L. M. Dalberto, M. L. Fisher, A. J. Greenfield, R. Jaikumar, P. Kedia, R. G. Mack, and P. J. Prutzman. 1983. Improving the distribution of industrial gases with an online computerized routing and scheduling optimizer. Interfaces 13(June):4–23.

Bellman, R. E. 1957. Dynamic Programming. Princeton, N.J.: Princeton University Press.

Bodin, L., B. Golden, A. Assad, and M. Ball. 1983. Routing and scheduling of vehicles and crews: the state of the art. Computers and Operations Research 10(1):63–211.

Boykin, Raymond F. 1985. Optimizing chemical production at Monsanto. Interfaces 15(January/February):88–95.

Brandeau, M., and R. C. Larson. 1985. Extending and applying the hypercube queueing model to deploy ambulances in Boston. In Urban Service Systems, A. Swersey and E. Ignall, eds. New York: North Holland, pp. 121–153.

Browne, Jim. 1984. Management and Analysis of Service Operations. New York: North-Holland.

Carlisle, David P., Kenneth S. Nickerson, Stephen B. Probst, Denise Rudolph, Yosef Sheffi, and Warren Powell. 1987. A turnkey micro-computer based logistics planning system. Interfaces 17(July/August):16–26.

Charnes, A., and W. W. Cooper. 1984. Creative and Innovative Management. Essays in Honor of George Kozmetsky. Cambridge, Mass.: Ballinger.

Cochard, Douglas D., and Kirk A. Yost. 1985. Improving utilization of Air Force cargo aircraft. Interfaces 15(January/February):53–68.

Dantzig, George B. 1963. Linear Programming and Extensions. Princeton, N.J.: Princeton University Press.

Davis, Samuel G., George B. Kleindorfer, Gary A. Kochenberger, Edward T. Reutzel, and Emmit W. Brown. 1986. Strategic planning for bank operations with multiple checkfstprocessing locations. Interfaces 16(November/December):1–12.

Eaton, David J., Mark S. Daskin, Dennis Simmons, Bill Bulloch, and Glen Jansma. 1985. Determining emergency medical service vehicle deployment in Austin, Texas. Interfaces 15(January/February):96–108.

Edwards, Jerry R., Harvey M. Wagner, and William P. Wood. 1985. Blue bell trims its inventory. Interfaces 15(January/February):34–52.

Finke, Gary F. 1984. Determining target inventories of wood chips using risk analysis. Interfaces 14 (September/October):53–58.

Fitzsimmons, James A., and Robert S. Sullivan. 1982. Service Operations Management. New York: McGraw-Hill.

Ford, L. R., Jr. and D. R. Fulkerson. 1962. Flows in Networks. Princeton, N.J.: Princeton University Press.

Gardner, Everette S., Jr. 1987. A top down approach to modeling U.S. Navy inventories. Interfaces 17(July/August):1–7.

Gelb, Betsy, D., and Basheer M. Khumawala. 1984. Reconfiguration of an insurance company's sales regions. Interfaces 14(November/December):87–94.
Goeller, Bruce F., and the PAWN Team. 1985. Planning the Netherland's water resources. Interfaces 15(January/February):3–33.
Guile, Bruce R., and Harvey Brooks, eds. 1987. Technology and Global Economy. Washington, D.C.: National Academy Press.
Haehling von Lanzenauer, Christoph, Erwin Harbauer, Brian Johnson, and David H. Shuttleworth. 1987. RRSP Flood: LP to the Rescue. Interfaces 17(July/August):27–40.
Harrison, H. 1986. Management science and productivity improvement in Irish Milk Cooperatives. Interfaces 16(July/August):31–40.
Hax, Arnoldo C., ed. 1987. Planning Strategies that Work. New York: Oxford University Press.
Hehnen, M. T., S. C. Chou, H. L. Scheurman, G. J. Robinson, T. P. Luken , and D. W. Baker. 1984. An integrated decision support and manufacturing control system. Interfaces 14(September/October):44–52.
Heskett, James L. 1986. Managing in the Service Economy. Boston, Mass.: Harvard Business School Press.
Holloran, Thomas J., and Judson E. Byrn. 1986. United Airlines station manpower planning system. Interfaces 16(January/February):39–50.
Huysmans, Jan H. B. M. 1970. The Implementation of Operations Research. New York: John Wiley & Sons.
Ikura, Yoshiro I., George Gross, and Gene Sand Hall. 1986. PGandE's state-of-the-art scheduling tool for hydro systems. Interfaces 16(January/February):65–82.
Kantorovich, R. E. 1939. Mathematical methods in the organization and planning of production. Publication House of the Leningrade State University, 68 pp. Translated in Management Science, 1960. 6:366–422.
Kleutghen, P. P., and J. C. McGee. 1985. Development and implementation of an integrated inventory management program at Pfizer Pharmaceuticals. Interfaces 15(January/February) 69–87.
Kozmetsky, George. 1984. Creative and innovative management, a new academic frontier. Pp. 3–25 in Creative and Innovative Management. Essays in Honor of George Kozmetsky, A. Charnes and W. W. Cooper, eds. Cambridge, Mass.: Ballinger.
Larson, Richard C. 1972. Improving the effectiveness of New York City's 911. Pp. 151–180 in Analysis of Public Systems, A. W. Drake, R. L. Keeney, and P. M. Morse, eds. Cambridge, Mass.: MIT Press.
Larson, Richard C. 1988. Transporting sludge to the 106 mile site: an inventory/routing model for fleet sizing and logistics system design. Transportation Science.
Larson, Richard C. 1972. Urban Police Patrol Analysis. Cambridge, Mass.: MIT Press.
Larson, Richard C., Alan Minkoff, and Paul Gregory. 1988. Fleet sizing and dispatching for the marine division of the New York City Department of Sanitation. In Vehicle Routing: Methods and Studies, Bruce Golden and Arjang Assad, eds. New York: John Wiley & Sons.
Larson, Richard C. and Amedeo R. Odoni. 1981. Urban Operations Research. Englewood Cliffs, N.J.: Prentice-Hall.
Lembersky, Mark R., and Uli H. Chi. 1986. Weyerhaeuser decision simulator improved timber profits. Interfaces 16(January/February):6–15.
Liberatore, Mathew J., and Tan Miller. 1985. A hierarchical production planning system. Interfaces 15(July/August):1–13.
Magirou, Vangelis F. 1986. The efficient drilling of printed circuit boards. Interfaces 16(July/August):13–23.
Maltz, Michael D. 1984. Recidivism. Orlando: Academic Press.

Martin, Clarence H., and Sanford L. Lubin. 1985. Optimization modeling for business planning at Trumbull Asphalt. Interfaces 15(November/December):66–72.

Martland, Carl, Henry S. Marcus, and George B. Raymond, Jr. 1986. Boston and Maine achieves control over railroad performance. Interfaces 16(September/October): 1–16.

McCloskey, Joseph F. 1987. The beginnings of operations research: 1934–1941. Operations Research 35(1):143–152.

Morse, P. M. 1977. In at the Beginnings. Cambridge, Mass.: MIT Press. Oliff, Michael D., and E. Earl Burch. 1985. Multiproduct production scheduling at Owens-Corning Fiberglass. Interfaces 15(September/October):25–34.

Quinn, James Brian. 1987a. Managing strategic change. Pp. 1–18 in Planning Strategies that Work, Arnoldo C. Hax, ed. New York: Oxford University Press.

Quinn, James Brian. 1987b. The impacts of technology in the services sector. Pp. 119–159 in Technology and Global Economy, Bruce Guile and Harvey Brooks, eds. Washington, D.C.: National Academy Press.

Parsons, Gregory L. 1987. Information technology, a new competitive weapon. Pp. 196–211 in Planning Strategies that Work, Arnoldo C. Hax, ed. New York: Oxford University Press.

Riccio, Lucius J., Joseph Miller, and Ann Litke. 1986. Polishing the big apple: How management science has helped make New York cleaner. Interfaces 16(January/February):83–88.

Shapiro, Roy D., and James L. Heskett. 1985. Logistics Strategy, Cases and Concepts. St. Paul, Minn.: West Publishing Co.

Stankard, Martin F. 1981. Successful Management of Large Clerical Operations. New York: McGraw-Hill.

Sullivan, Robert S., and Stephen C. Secrest. 1985. A simple optimization DSS for production planning at Dairyman's Cooperative Creamery Association. Interfaces 15(September/October):6–53.

Swersey, Arthur J., and Edward J. Ignall. 1986. Delivery of Urban Services. Amsterdam: North-Holland.

Terry, L. A., M. V. F. Pereira, T. A. Araripe Neto, L. F. C. A. Silva, and P. R. H. Sales. 1986. Coordinating the energy generation of the Brazilian National Hydrothermal Electrical Generating System. Interfaces 16(January/February): 16–38.

Walker, W. E., J. M. Chaiken, and E. J. Ignall, eds. 1979. Fire Department Deployment Analysis. Amsterdam: North-Holland.

Welch, Norma, and James Gussow. 1986. Expansion of Canadian National Railway's line capacity. Interfaces 16(January/February):51–64.

Whitmore, William F. 1953. Edison and operations research. Operations Research 1:83–85.

Cellular Mobile Telephone Services

John H. Davis

The cellular mobile telephone system is an American technology, driven by American need and created by American innovation. At a time when conventional mobile telephone service was available only to an elite few, the cellular concept promised a thousandfold increase in the availability of the service in the United States, but its birth encountered significant delays before it became a reality. Ultimately, more than a decade passed from initial proposals to implementation. With strong and continuing pressures from legal, regulatory, and competitive forces, cellular development has been characterized by delay, changing rules, and shifting market structures.

This case will examine the genesis, evolution, and externalities of the cellular industry—from proposals in the 1960s to the service that exists today. Although it does not provide answers, hopefully the case will stimulate discussions on a number of difficult questions: Could the long delay have been shortened? What roles did the participants have in causing this delay? Was delay inevitable given the legal, regulatory, and market forces of the time? How have these forces shaped and reshaped industry structure? Were the basic issues involved regulatory or were they technological? How were the public benefits of service weighed against the participants' issues? What was the effect of these forces on foreign companies selling in the United States? On U.S. companies selling in other countries? Is this environment conducive to the investment of resources by U.S. industry?

BACKGROUND: CHRONOLOGY OF EVENTS TO 1968

After years of development and testing, mobile telephone service, that is, a service that allows people in cars and trucks to talk to land-based telephone customers who are using ordinary telephones, was first introduced in this country by AT&T in 1946. A mobile user in those days had to select a channel manually, depress a "push-to-talk" button on the mobile unit, and place the call through a telephone operator. Thus, the early mobile service was much like the operator-based service provided by telephone companies before automatic dialing systems were available. As mobile service expanded, various technological advances combined to make it operationally comparable to the improving land-based telephone service.

Throughout the history of mobile telephony, the increasing popularity of the service outstripped the capacity of available channels. By its nature, mobile telephony uses a scarce public resource—the radio spectrum. Because this resource is limited, much as land is limited, the history of the service is dominated by technological developments that attempt to make better use of the resource. Despite such innovations as splitting of channels in existing bands and the opening of new bands,[1] channels remained overloaded, causing existing customers considerable difficulty in completing calls and causing extremely long waiting times for new customers to obtain service. In New York City customers could wait many years or even indefinitely to obtain the service.

Mobile telephone service was offered by two types of service providers: (1) the local telephone companies or Wireline Common Carriers (WLCCs) and (2) the Radio Common Carriers (RCCs). The RCCs were established by the Federal Communications Commission (FCC) as a competitive force in the mobile market and were assigned a number of channels equal to the number assigned to the WLCCs. This early action taken by the FCC in 1948 to foster competition weighed significantly on the later events in cellular development.

Long before the advent of mobile telephone service, privately owned and operated radio systems were developed for governmental, law enforcement, industrial, and similar organizations (e.g., police, taxi, ambulance). Over the years, the amount of spectrum made available for these private systems substantially outweighed the spectrum allocated for mobile telephone systems. By 1968, a frequency interval of 40 MHz was allocated for private mobile systems and an interval of less than 2 MHz was allocated for common carrier mobile telephone systems.[2] Despite this relatively large allocation of spectrum, the private radio users were forced by their large numbers to share channels extensively in urban areas. When a channel is shared, the users rely on cooperative "party-line" operation, which often results in interference and generally poor performance. As the private radio and mobile telephone industries emerged and grew, they put continuing pressure on the FCC for additional channels. Although their goals were in concert, they were actually in competition with each other for radio channels.

In 1958 the needs of the mobile radio industries for new channels were considered by the FCC. The parties representing the mobile radio and tele phone industries were pitted against the television broadcast industry. After deliberation, the FCC determined that the public interest would best be served by allocating the spectrum in question to broadcast TV, creating the UHF (Ultra High Frequency) TV channels numbered 14 to 83. Because part of this frequency band was the one the FCC later proposed to reallocate for cellular service, the broadcast industry (in its opposition) was also to play a major part in the cellular proceeding.

Another group to become a stakeholder in the cellular proceedings was the radio equipment manufacturers. Although uniformly supporting the allocation of additional channels for mobile services, this group would split into two camps: potential suppliers of equipment for cellular systems, who favored a large cellular allocation, and the suppliers of equipment for private systems, who favored an increased allocation for that use.

During this pre-1968 period, AT&T made a series of requests to the FCC for frequency allocations in which to provide a high-capacity mobile telephone service (including the above-mentioned determination in 1958). Each of these requests was denied, primarily owing to conflicting demands (for example, broadcast TV) for the available spectrum resource.

THE REGULATORY ARENA: 1968–1982

In 1968 the FCC initiated an inquiry to determine whether the upper part of the UHF TV band (channels 70 to 83 on the TV dial) should be reallocated from broadcast TV to mobile radio services. This 84-MHz interval, combined with other available frequencies, formed a 115-MHz block that could possibly be allocated for mobile telephone, private mobile, and air/ground radio services.

As might be expected, the broadcast interests vigorously opposed reallocation. All mobile interests, both common carrier and private mobile, strongly supported the proposed allocation (but with very different viewpoints). In 1969 the FCC held 2 days of oral arguments, and in May 1970, issued its First Report and Order.[3] As proposed in the 1968 inquiry, it reallocated a total frequency spread of 115 MHz for mobile radio systems. As illustrated in Figure 1, the order proposed the allocation of intervals 64 MHz for cellular telephone service, 40 MHz for private mobile service, and 11 MHz for air and ground service. In addition, the FCC asked for recommendations on how this large amount of spectrum could be used most efficiently and for the greatest public good.

In December 1971 AT&T responded to the inquiry with a technical description of a cellular telephone system. The cellular system was developed by AT&T's Bell Telephone Laboratories and was a major step forward in mobile system technology. It was made possible by the emerging power of microprocessor technology and large software controlled switching machines.

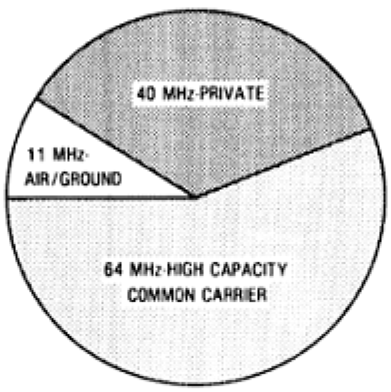

Figure 1 Mobile radio systems frequency band allocation, 1970. Total 115 MHz.

Although its distributed architecture initially required a large number of channels, it could multiply their usage to fuel continuing growth, through cell splitting.

The proposed cellular system differed completely from ordinary mobile systems. As illustrated by Figures 2 and 3, instead of using a single high-powered transmitter to reach users within a range of approximately 50 miles, cellular systems distribute their channels among small areas, or "cells," throughout the total coverage area of the system. Because the cells are relatively small, low-powered transmitters can be used, and cells only a few miles apart are able to reuse the same channels for different calls.

As a vehicle travels from one cell to another, the call in progress is automatically and imperceptibly "handed-off" to the next transmitter and a

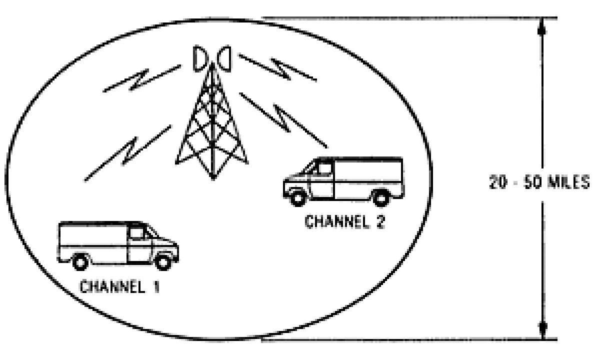

Figure 2 Conventional mobile telephone system. Mobiles 1 and 2 must operate on different channels to avoid interference.

different channel by software-controlled switching equipment. Whereas, in a conventional system a channel used in Philadelphia, for instance, cannot be reused nearer than New York, in a cellular system that channel can be reused many times in each city. Through a combination of channel reuse and the larger number of channels available, this system offered a thousandfold increase in available service.

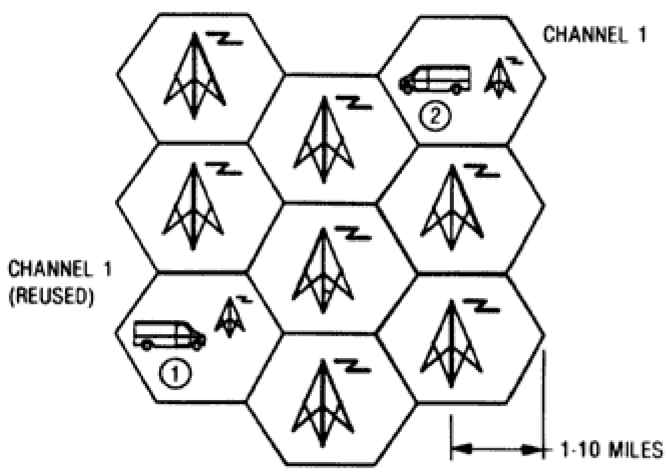

Figure 3 Cellular mobile telephone system. Mobiles 1 and 2 may operate on the same channels.

Because both the cost and the capacity of such systems increases with the use of smaller cells, it was proposed that larger cells (about 10 miles in radius) be used to limit cost at start-up. These would then be split to smaller sizes as the number of customers increased, with cells ultimately as small as 1 mile in radius.

After the May 1970 reallocation of spectrum by the FCC from TV broadcast to mobile radio, individual mobile radio interests began to strongly express their views on how the spectrum should be split within the mobile radio community. A number of private mobile users filed comments, requesting frequencies for their particular service and attacking the viability of cellular service. Interestingly, Motorola, which was to become a pioneer in cellular development, was also the dominant supplier of private systems and generally favored the latter in allocation proceedings. AT&T, on the other hand, had few business interests in the private systems, and it consistently favored allocation to the cellular systems.

Although this conflict was rooted in market strategy, it was often expressed as technological debate between AT&T and Motorola on the "right" cellular design.[4] These debates were difficult for either party to win and led to additional delay in the rulemaking. Ultimately, the U.S. standards allowed a good deal of flexibility to accommodate such differences, and designs became more complex to maintain service compatibility despite different system designs. In addition to debates among the manufacturing interests, the RCCs, traditional competitors of the WLCCs, filed strong protests contending that they were entirely neglected in the allocation and that with no frequencies dedicated to them they would be put out of business.

After more FCC orders and additional oral arguments before the FCC, a Second Report and Order was issued in 1974.[5] That order, as illustrated in Figure 4, changed the allocation scheme to:

- 40 MHz for WLCCs to provide a single cellular system per market area;
- 30 MHz for private mobile systems; and
- 45 MHz for reserve for future growth.

Most of the interested parties filed for reconsideration of the FCC order. As a result, in March 1975 the FCC issued its "final" order,[6] changing the allocation to:

- 40 MHz for *any* Common Carrier that qualifies to provide a single cellular system per market area;
- 30 MHz for private mobile systems; and
- 45 MHz held in reserve.

The RCC industry, still dissatisfied, took the FCC's order to court. They alleged that although the 40 MHz was available to *any* Common Carrier, the practical result would be a monopoly for the WLCCs since all the technology belonged to the Bell System. At about this time, in an apparent contradiction to that allegation, American Radio Telephone Systems (ARTS), an RCC, filed for a developmental system in Baltimore, using the Motorola cellular technology.

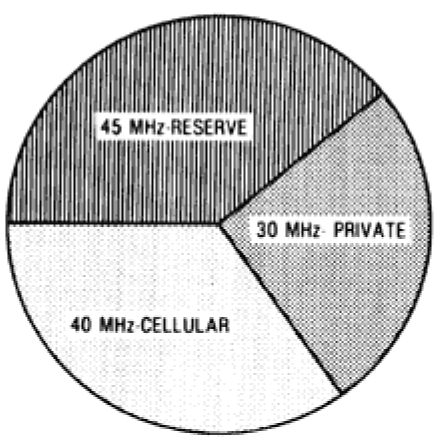

Figure 4 Mobile radio systems frequency band allocation, 1974 and 1975. Total 115 MHz.

In 1976 the U.S. Court of Appeals rejected the RCC argument. Although affirming the FCC's order, it reserved judgment on whether the plan would result in a breach of the antitrust laws—thus inviting further litigation "down the road."

Further, the court noted that because the FCC order provided for only one system to be authorized per market area, it was clear that applicants filing mutually exclusive applications for the same cities would cause lengthy hearing proceedings followed by appeals in virtually every large city in the country.

At about that time, Congress took an interest in the problem, and in 1979, in an appearance before a House Subcommittee, AT&T proposed a new division of the spectrum intended to answer the Court of Appeals concerns and bring the service to market more quickly. The proposal, illustrated in Figure 5, recommended two systems per market area (instead of one)—one for RCCs and one for WLCCs—maintaining the existing historical competitive market structure. In addition, all carriers would be required to sell to intermediaries for resale to ultimate consumers. As part of the proposal AT&T committed to first service within 2 years of the grant of a construction permit from the FCC, and to the construction of 35 systems covering 70 cities within 5 years.

Following the congressional inquiry, in January 1980 the FCC opened an Inquiry and Rulemaking proceeding to establish general policy and rules for commercial service.[7] In March 1982 the FCC issued an order that made cellular service to the public possible. That order, illustrated in Figure 6, allocated:

- 20 MHz to WLCCs;
  20 MHz to RCCs;
  30 MHz for private radio; and
  45 MHz held in reserve (20 MHz "earmarked" for cellular).

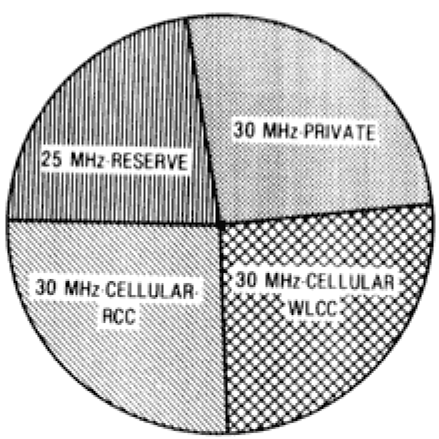

Figure 5 Proposed mobile radio systems band allocation, 1979.

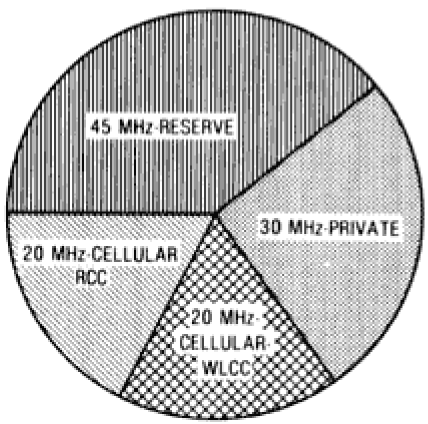

Figure 6 Mobile radio systems frequency band allocation, 1982.

The reduction of frequency allocations from 64 MHz per system in 1970 to 20 MHz per system in 1982 was estimated by AT&T to increase the per user cost, exclusive of the in-car unit, from approximately $650 to approximately $1,600[8] (see Figures 7 and 8). Likewise, the increase in systems per market area obviously reduced market potential per system. These changes, "hammered out" over a decade of debate, led to changing technical standards, new commitments, new cost assumptions, and, ultimately, to a new industry concept for cellular deployment throughout the United States.

System Trials and Adoption of Technical Standards

During the regulatory process previously discussed, the FCC attempted to confirm the validity and necessary standards for commercial cellular service by requiring developmental "trials" of the cellular concepts. In July 1975 AT&T [acting through Illinois Bell Telephone Company (IBT)] filed an application for a developmental cellular system in Chicago. Both the RCC industry and Motorola opposed the filing. This opposition, and a number of issues concerning the details of the trial configuration, caused a lengthy delay in approval. During this period ARTS filed for a developmental license in the Washington/Baltimore area, using equipment developed by Motorola. The FCC granted the IBT application in March 1977, and the ARTS application later in 1977.

AT&T's Chicago trial was configured as a fully operational start-up cellular system. In its first phase the trial used approximately 100 mobile units to verify proper operation of all equipment and gather operational data for evaluation. In its second phase the trial offered service to 2,000 real customers. This phase was conducted to predict the potential penetration of cellular service into various types of business segments and to estimate the elasticity of the market to price, different types of terminals, and so forth.

The Chicago trial system consisted of 10 cells, a switching center housing a large, specially programmed telephone switching machine (called a Mobile Telephone Switching Office or MTSO), and an installation and customer service center equipped to install and repair the 2,000 mobiles used in the trial. The performance of this system was generally considered excellent, and customer satisfaction with the quality of the service was extremely high.

To verify the ultimate cellular configurations, using cells as small as I mile in radius, AT&T's Bell Laboratories constructed an experimental system in Newark, New Jersey. The Newark system consisted of a single 1-mile radius cell surrounded by six interfering cells located at the appropriate distance away. This was intended to show that the ultimate small-cell configuration was attainable in a real-world urban environment, and to refine the processes for locating the vehicle and handing it off between cells.

Like the AT&T Chicago system, the ARTS system in Baltimore/Washington was configured as a start-up cellular system, although one cell was

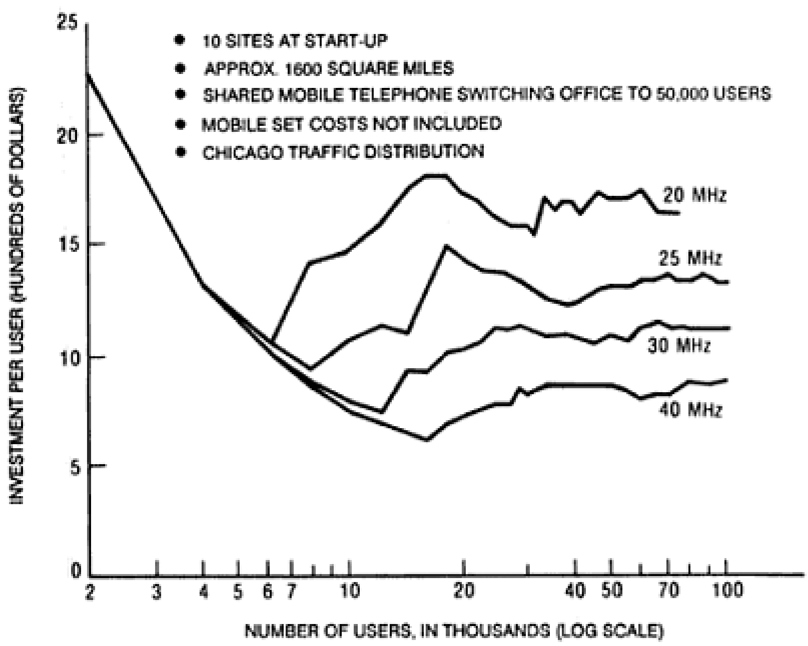

Figure 7 Cost sensitivity for spectrum allocation.

operated for a time in a configuration simulating a mature (small) cell. It served some 50 mobile units and 50 hand-held portables.

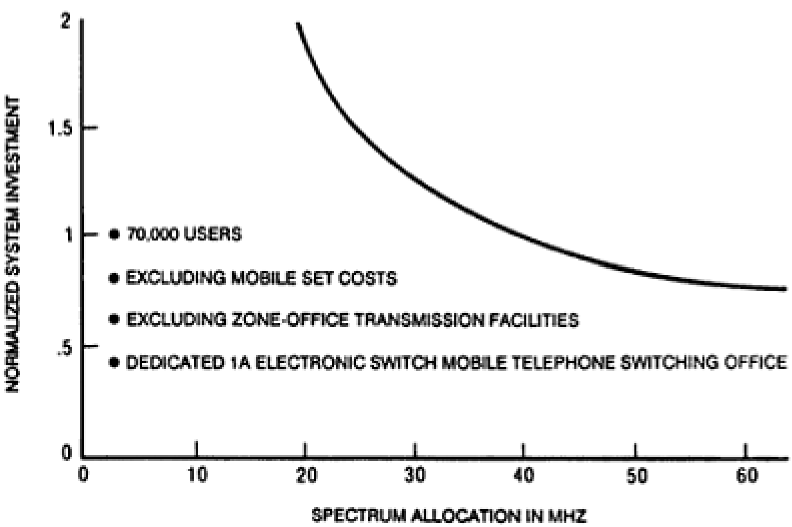

Figure 8
Effect of spectrum allocation on investment.

Although the AT&T and Motorola configurations differed, both supported the viability of the cellular concept and laid the groundwork for setting technical standards. These standards were developed under the aegis of an Electronic Industries Association (EIA) ad hoc committee and covered issues of interface compatibility between mobile units of competing systems. The standards were proposed to the FCC, which adopted them as part of its 1982 order. The Chicago market trial commenced in January 1979 and continued to be operated by IBT until October 1983, when it was replaced by a new system designed for commercial operation. The ARTS system commenced operation in late 1981 and was converted to commercial operation in December 1983.

Industry Evolution

Based on the evolving standards for cellular applications, the FCC permitted applications for the top 30 markets on June 7, 1982.[9] The first licenses were granted to the AMPS Corporation, a subsidiary of AT&T formed to plan, install, own, and operate cellular systems on a nationwide basis. As planning for the first group of systems moved forward, however, the dives titure of AT&T Operating Telephone Companies from the parent AT&T caused a major discontinuity. The AT&T-Justice Department Modified Final Judgement (MFJ) divested the Bell Operating Companies from AT&T as subsidiaries of seven Regional Holding Companies (RHC) on January 1, 1984. Cellular service was designated as a local exchange carrier service and, as a result, the AMPS Corporation was replaced by subsidiaries of the RHCs that ultimately specified and owned the new systems. As a result, the nationwide aspects of the WLCC service (equipment compatibility, nationwide automatic roaming, standardized nationwide features and functions, uniform coverage standards, nationwide billing, and so forth) could no longer be dictated by one provider. This development had a significant impact on the evolution of cellular systems in the United States.

The first commercial cellular system began operation in Chicago on October 13, 1983. It was operated by Ameritech Mobile Communications, Inc., which was a subsidiary of Ameritech, one of the RHCs. The Ameritech system was followed in December 1983 in the Baltimore/Washington, D.C. area by an RCC system that was owned by a consortium of the Washington Post Company, Metromedia, and other local groups. A few months later, in April 1984, Bell Atlantic Mobile Service, Inc., an RHC cellular subsidiary, also began service in the Washington/Baltimore area, making it the first competitive market area in the United States.

During the transition period between 1982 and 1984, the AT&T AMPS subsidiary, with encouragement from the FCC, negotiated settlements with other WLCCs for providing service through partnerships in most of the major markets. Table 1 lists examples of the settlements that were made between AMPS and the other WLCCs.

The RCCs, on the other hand, had a more difficult time organizing their business relationships. Unlike the WLCCs, the RCCs had no unique operating areas or history of cooperation in offering telephone service. As a result, long and arduous negotiations ensued for ownership of many of the RCC cellular markets. During the next 3 years, while these negotiations were taking place, it was very difficult for an RCC to deal with the details of planning the construction of its system or to have meaningful contractual discussions with vendors of cellular equipment.

Start-up of the industry was also affected by the difficult task the FCC faced in processing the large numbers of applications for cellular markets. Initially, the Commission required each applicant for a given market to submit an extremely detailed engineering plan for the design and construction of its system. The Commission reviewed these applications as a part of the procedure for awarding the market to a given applicant. As mentioned above, the WLCCs had largely resolved settlement issues in the early markets. However, with no such settlements in place for the RCCs, the Commission found itself inundated with applications for the early markets (194 applicants for markets 1 to 30, 396 for markets 31 to 60, 567 for markets 61 to 90).[10]

TABLE 1 Examples of Applicants and Settlements Between AMPS and Other WLCCs

City	Applications	Final Ownership
New York City	AMPS	Bell Atlantic
		NYNEX
Los Angeles	AMPS	Pacific Telesis
	GTE	GTE
		TDS
Seattle	AMPS	New Vector
	GTE	GET
		Whidbey Telco
Dallas	AMP	Southwestern Bell
	GTE	GTE
		Lake Dallas Teleco
Minneapolis	AMPS	New Vector
	Continental	Continental
	United	Scote Rice Telco
Detroit	AMPS	Ameritech
	GTE	GTE
		Mid Cont. Telco

As a result of this deluge, the Commission elected to use a lottery system to resolve mutually exclusive applications. Under this concept, if the applicants for a given market could not resolve their competing applications, the Commission ran a lottery to determine the licensee. Only a modest amount of information was required of each applicant prior to the lottery, so the Commission's processing load was substantially reduced.

Since the cost of submitting an application was also substantially reduced, the lottery developed attributes of a modern-day "gold rush." Entrepreneurs, small investors, and major corporations all rushed to "sign up." More than 90,000 applicants filed for markets 91 to 305. It became clear that many of these individuals were not interested in building and running a cellular system but were principally driven by the financial benefits of winning the lottery for an important market.

During this period of evolution in the industry, three major industry groups were also evolving in their roles within the cellular industry. The first, the Electronic Industries Association (EIA), had already played a key role in defining the standards for nationwide compatibility. During the early period of cellular evolution, the EIA continued to enter into issues of spectrum assignment and standards, often achieving a consensus among the manufacturers of cellular equipment.

The second group was the Telocator Network of America (TNA), a trade association that had represented RCCs for several decades. TNA had been vocal in demanding the split of spectrum into two bands and, in later years, became heavily involved in representing the RCC interests in having efficient and competitive interconnection arrangements with the WLCCs.[11] A major new issue faced by TNA during this period was whether to include WLCCs in the association and how to deal with the long-standing distrust between the WLCCs and the RCCs. This issue became extremely complicated as the consolidation of market ownership began to take place in 1985, when a WLCC, Pactel Mobile Access Company, purchased the ownership of an RCC owned by Metromedia in the San Francisco market. Once Pactel received approval for this acquisition a flurry of acquisitions took place, essentially blurring the distinction between WLCC and RCC in the spectrum that had been so laboriously segmented by the regulatory process. During this period TNA spent much of its energy on internal debates of organization and membership scope, defusing its ability to have a major shaping impact on the cellular industry.

Finally, a new trade association, the Cellular Telecommunications Industry Association (CTIA), was formed early in 1984. Although this group attempted to represent both WLCC and RCC interests in industry issues, it is viewed by a number of RCCs as still dominated by WLCCs. As a result there is a somewhat ambivalent attitude toward the organization from the RCC industry. Much of CTIA's effort has been spent on promoting the allocation of additional (reserve) spectrum and on pressing for legislation against eavesdropping on cellular calls.

Thus, no powerful central authority or representative body has emerged to guide the relationships among participants in this increasingly segmented industry, which has had an impact on the ability of the user to have a nationwide service. When a cellular subscriber "roams" (leaves his or her home serving area and enters another serving area) it is possible for that subscriber to place or receive calls. However, the details of how the subscriber goes about this vary from city to city, depending on the arrangements made among individual system operators.

This situation has made the service difficult to use, and has encouraged the fraudulent placing of calls by illegal roamers. It has also led to business opportunities, however. Several service firms have emerged to become "roaming clearinghouses." These firms attempt to handle the details of billing and customer authentication without requiring a detailed knowledge by the cellular subscriber of what transactions are required. Thus, even though cellular service was extremely fragmented as it emerged in the United States, job opportunities in the service industry did arise to help circumvent some of the problems.

PRESENT STATUS

Market Size

Despite all the aforementioned problems, cellular systems have generally achieved increases in customer demand that exceeded expectations. Table 2 presents the results of a survey of 27 systems that had been in operation at least 3 months. This survey was conducted by Compucon on behalf of TNA and CTIA, the two trade associations described earlier. The subscriber levels achieved after 6 months of operation were 14 percent higher than originally forecast. In the three largest markets the average was 20 percent above the original forecast. Based on this favorable early experience Compucon (and many other market analysts) revised their 5-year projections of the cellular market. Compucon's revised projections are shown in Table 3.

The story is much the same regarding the level of customer usage and resulting revenues from such usage. In general, average customer usage exceeds the early forecasts. Although price wars between competitors and special price packages have reduced rates in different markets at different times, they have also stimulated the markets, and the net effect has been positive.

Based on the results of the AT&T market trial in Chicago described earlier, the conventional estimate of market potential for cellular at the time of the first application filings was 2 percent of the population. Since that time, and

TABLE 2 Original Forecast and Actual Numbers of Subscribers the First Two Quarters of Operation, 1986

	Quarter 1		Quarter 2	
Markets	Forecast	Actual	Forecast	Actual
1–3	5,165	6,198	9,191	11,029
4–10	1,441	2,190	2,765	3,041
11–20	860	1,084	1,699	1,648
21–30	623	716	1,167	1,191

TABLE 3 Revised 5-Year Forecast of Subscribers by Years, 1986

Markets	Year 1	Year 2	Year 3	Year 4	Year 5
1–3	21,200	40,000	55,700	73,367	90,900
4–10	7,157	14,506	22,056	30,636	38,418
11–20	2,942	4,992	8,839	12,613	16,176
21–30	2,236	4,932	7,853	10,861	13,343

currently, there have been numerous forecasts that range from today's 1.5 percent penetration in most markets to 4 percent or greater within 5 years.

Table 4, a chronology of development of the technology and industry, reviews the full 40-year history of the industry. Although the simultaneous development of markets and technologies is frequently a slow process, regulatory decision making—necessary because of the fact that bandwidth is a scarce resource—played and continues to play an especially important role in bringing cellular telephony to the public.

The Cellular Manufacturing Industry

The industry of equipment manufacturers who design and build the systems and the mobile units grew in parallel with the development of the cellular service industry. In the United States AT&T and Motorola became the major suppliers of systems, each supplying more than 30 percent of the current installed systems (the AT&T systems tend to be in larger cities, giving AT&T the greater share of equipment sales). Northern Telecom, in partnership with General Electric, Nippon Electric Company, Ericcson, Astronet, NovAtel, and CTI/E.F. Johnson share the remaining third of the market.

The world situation is rather different. During the long period of delay in the United States the Japanese established standards, built operating cellular systems in Japan, and began to sell them successfully in the Middle East and Pacific regions. Later, as planning began for European and Canadian systems, U.S. companies were able to bid but have sold few systems to date. International sales are hampered by differing standards and the tendencies of countries to politicize this type of decision.

In the manufacture of mobile units, U.S. firms have fared even less well; only Motorola has achieved any significant success. The U.S. market, on the other hand, is shared by many manufacturers (see Table 5).

Ironically, the other significant U.S. manufacturer of cellular systems equipment, AT&T, was excluded from the mobile unit market by the FCC in 1974.[12] The FCC took that action in an effort to prevent AT&T from being both a service provider and a manufacturer. As a result, AT&T was forced to share its technology with other sources, including Motorola and the Japanese, to assure the availability of mobile units for its systems. By the time this prohibition was removed in 1981, [13] AT&T was far behind the competitors it had helped to create. After an attempt to catch up it was forced to abandon its attempt to compete in this large market.

FUTURE

The suppliers of cellular equipment find themselves facing a new dilemma. The larger systems will reach their ultimate capacity within a few years,

TABLE 4 Cellular Mobile Telephone Development: Synopsis of Key

Key Event	Background Technology Development	Regulatory Arena	System Trials and Adoption of Standards	Industry Evolution
1946 Bell System initiates first FM mobile telephone service at 35 MHz (manual operation)	✓			
1948 AT&T Bell Labs invents transistor, allowing miniaturization of electronic equipment	✓			
1948 FCC establishes RCCs to compete with local telephone companies		✓		
1958 FCC decides in favor of UHF-TV channels 14–83		✓		
1959 Bell System initiates first FM mobile telephone service at 150 MHz (manual operation)	✓			
1959 AT&T Bell Labs proposes high-capacity mobile system at 35 MHz	✓			
1964 Bell System initiates first improved mobile telephone service at 150 MHz (full duplex operation)	✓			
1968 FCC orders inquiry to reallocate UHF-TV channels 70–83 for mobile radio use		✓		
1969 Bell System initiates first improved mobile telephone service at 450 MHz (full duplex operation)	✓			
1970 FCC reallocates 115 MHz bandwidth for mobile radio services (64 MHz for common carriers)		✓		
1971 AT&T Bell Labs files technical description with FCC of proposed cellular telephone system at 850 MHz	✓			

continued

Key Event	Background Technology Development	Regulatory Arena	System Trials and Adoption of Standards	Industry Evolution
1974 FCC reduces its 1970 allocation to 40 MHz bandwith for common carrier use (one cellular system per market area)		✓		
1974 AT&T excluded from mobile unit market		✓		
1975 AT&T files for developmental cellular system in Chicago			✓	
1976 ARTS files for developmental cellular system in Washington, DC			✓	
1979 AT&T Bell Labs proves in cellular concepts at Newark, NJ, cellular test bed	✓			
1979 AT&T's Chicago developmental system operational			✓	
1980 FCC opens inquiry to establish policy and rules for commercial cellular service		✓		
1981 AT&T allowed to reenter mobile unit market		✓		
1981 ARTS' Washington developmental system operational			✓	
1982 Motorola introduces first hand-held cellular terminal	✓			
1982 FCC issues order to make commercial cellular service possible. Allows 2 cellular systems per market area with 20 MHz bandwidth allocated for each system		✓		
1982 EIA technical standards for cellular interconnection and nationwide service adopted by FCC			✓	

continued

Key Event	Background Technology Development	Regulatory Arena	System Trials and Adoption of Standards	Industry Evolution
1983 1st commercial cellular system operational at Ameritech Mobile Communications in Chicago				✔
1983 First RCC cellular system operational by Cellular One in Washington, DC				✔
1984 Justice Dept. modified final judgment designates cellular as a local exchange service. AT&T interests relegated to that of manufacturer		✔		
1984 Bell Atlantic Mobile offers first competing system in Washington, DC				✔
1985 Pactel Mobile Access purchases RCC				✔
1987 AT&T Bell Labs proposes compatible next-generation digital cellular system	✔			
1987 FCC opens inquiry on new cellular technologies		✔		
1987 One million subscribers using cellular service				✔

requiring more spectrum, narrower channels, and even smaller cells than were originally planned. Current systems were specified and designed during the 1970s and new opportunities such as narrow-band digital channels and distributed-control architectures have become possible. The FCC has recently taken an action that it hopes will encourage new technology.[14] It has proposed new rules that would allow cellular licensees to use advanced cellular technologies and provide additional services. These new rules, if adopted, would remove all existing restrictions on channeling schemes and types of emissions and modulations techniques. As a further part of the proceeding, the FCC requests comments on the desirability of requiring some portion of the cellular frequencies to be maintained for service that conforms to present compatibility and other technical requirements—providing assurance for continuation of roaming for customers using today's systems. Given the fragmented nature of the cellular service industry, however, it will be difficult to reach consensus on a path to the future, and the need to coexist with today's systems and mobile units makes every new path complex and thus expensive for the developer.

TABLE 5 Manufacturers of Mobile Equipment for U.S. Users

Audiotel	Motorola
E. F. Johnson	NovAtel
GE	OKI
Hitachi	Panasonic
NEC	Tandy
Mitsubishi	Toshiba
Mobira	Walker

Superimposed on the domestic market is the enticing yet hazardous vision of the burgeoning international market. The variables for participation in that arena are mind-boggling and any manufacturer willing to enter that marketplace will be faced with multiple technological policy issues. For example, a proposed Pan European cellular system contemplates using a state-of-the-art all-digital transmission scheme, which is totally incompatible with U.S. systems. Should U.S. manufacturers wish to enter that market, they face not only the traditional parochial obstructions but also the uncertainty about regulatory directions in the United States that would change standards to permit such systems domestically. Prompt domestic regulatory direction would remove one significant barrier. To the extent that such direction requires U.S. standards to be compatible with preliminary international standards, U.S. manufacturers will have the benefit of sharing the R&D costs over a much larger base. If, on the other hand, it were promptly determined that it was not in the public interest for U.S. standards to be compatible with the proposed international specifications, the opportunity would still exist to present the U.S. case before international decision-making bodies before their adoption of final standards.

Initially, U.S. technology developed and encouraged the cellular industry. Two U.S. companies had the resources and willingness to pursue development through more than a decade of uncertainty. As the industry has grown and expanded, the external domestic and international environment has dramatically changed. The U.S. cellular industry has reached another major discontinuity in its growth. The future is dependent on all the cellular stake-holders. The service providers will have to understand the importance to a healthy, competitive marketplace, of long-term as well as short-term technical advancements. Further, they will have to define and commit to the role they will play in advancing that technology—either individually or through their trade associations. U.S. equipment manufacturers will have to decide whether they are willing to persevere with the financial and technological resources to ensure a place in the domestic and international marketplace. Finally, the government must determine what role, if any, it should play to assure continuation of a U.S. presence in the cellular field of the future. These issues must be faced promptly or they will be resolved by default—to the detriment of U.S. industry.

ACKNOWLEDGEMENT

I am most indebted to Mr. Louis Weinberg, who served as the principal investigator and coeditor of this case. Lou was actively involved in many of the issues in AT&T's earlier cellular activities, and his insights have been most useful.

Thanks, also, to Mr. R.H. Frenkiel, who provided many of the early innovations in the development of AT&T's cellular technology and who helped Mr. Weinberg in pulling this case together.

NOTES

1. Channel bandwidth was twice "split" as the technology advanced, creating four times as many channels in the available spectrum. The first allocation at 35 MHz (35 million cycles per second) was supplemented by allocations at 150 and 450 MHz.

2. In this context, MHz is simply a measurement of frequency bandwidth. For example, each TV broadcast channel uses 6 MHz, although that same bandwidth can accommodate 100 or more narrower telephone or private radio channels.

3. An inquiry relative to the future use of frequency band 806–960 MHz 19 RR 2d 1663 (1970), *2nd Report and Order*, 46 F.C.C. 2d 752 (1974), *Recon. Granted inpart*, 51 F.C.C. 2d 945 (1975), *Aff'd. sub nom. NARUC* v. *CC*, 525 F. 2d 630 (D.C. cir 1976), *cert. denied* 425 v.s. 992 (1976).

4. The major issue in these design "differences" had to do with the propagation pattern of the cell-site antennas and the degree of channel reuse that could be tolerated. In later years it became clear that these extremely technical arguments were somewhat moot.

5. See Note 3.

6. See Note 3, 51 F.C.C. 2d 945 (1975).

7. An inquiry into the use of the bands 825–845 MHz and 870–890 MHz for cellular communications systems, 78 F.C.C. 2d 984 (1980); *Report and Order* 86 F.C.C. 2d 469 (1981); *modified* 89 F.C.C. 2d 58 (1982); *further modified* 90 F.C.C. 2d 571 (1982); *appeal dismissed sub. nom, United States* v. *FCC*, no. 82–1526 (D.C. cir. 1983).

8. AT&T estimated that the division of the available spectrum would significantly raise the cost per user, owing to duplication of facilities and less efficient use of channels.

9. By this time there were operating systems in both Japan and the Nordic countries.

10. It is an interesting phenomenon that the smaller the market size, the larger the number of cellular applicants. This fact flies in the face of earlier studies that indicated that the larger markets were much more profitable than the smaller ones.

11. The RCCs are dependent on their WLCC competitor's parent local exchange companies for the wireline facilities that connect their transmitting/receiving sites to the telephone network. TNA has continually complained that, despite a clear FCC mandate to do so, the local exchange companies have resisted providing the RCCs with the type of interconnection arrangements they desire.

12. See Note 3.
13. See Note 7.
14. FCC GEN Docket No. 87–390.

Bibliography

Bell System Technical Journal. 58-1:January 1979.
Business Week. Fewer busy signals for mobile phones. August 7, 1978, p. 60B.
Business Week. The boom ahead in cellular radio. April 13, 1981, p. 124c.
Business Week. Hello, anywhere. April 21, 1987, p. 84.
Cooper, Martin. The sibling relationship of paging and cellular. Cellular Business. August 1987, p. 20.
Frenkiel, Richard H. A high capacity mobile telephone system model using coordinated small-zone approach. IEEE Transactions on Vehicular Technology. May 1970.
Frenkiel, Richard H. A high capacity mobile telephone system overview. Bell Telephone Laboratories, September 16, 1969 (private document).
International Telecommunications Union. Mobile communications—a status report. Telecommunications Journal. June 1980, p. 386.
Mobile Times. Cellular: Up and running at Illinois Bell. November 1980.
Patrick, Dennis R. Cellular '86—New approaches to today's challenges. Speech by Commissioner Patrick before the CTIA. Federal Communications Commission, December 4, 1985.
Robinson, John 0. Spectrum management policy in the U.S.: An historical account (OPP working paper series #15). Federal Communications Commission, Office of Plans and Policy. April 1985.
Seneker, Harold. Beep! Forbes. September 3, 1979, p. 52.
Television Digest Inc. Cellular radio: Birth of an industry. A special report from Television Digest, Inc., Washington, DC. August 1987.
U.S. House of Representatives. Committee on Interstate and Foreign Commerce. Hearings before the Subcommittee on Communications. June 7, 1979.
U.S. Senate. Committee on Commerce, Science and Transportation. Hearings before the Subcommittee on Communications. May 1, 1979.
Vorhies, John. Comparison of Capacity and Demand for Determining Long-term Spectrum Requirements of Cellular Radio. Office of Science and Technology, Federal Communications Commission. December 1984.

Modern Bridge Construction and Engineering Services

Juan A. Murillo

In fiscal year 1988 the Federal Highway Administration (FHWA) apportionment for bridge replacement and rehabilitation amounted to $1.4 billion, which, when combined with local and state funds, makes up the $3 billion per year industry in bridge design and construction in the United States. This level of expenditure on bridge design and construction represents a surge in bridge activity in the United States brought on by the aging of existing structures and increased traffic demands on the U.S. highway system.

From a global perspective the United States has been relatively slow in applying innovation now available mostly from European sources to bridge design and construction. As discussed later in this paper, the demand for postwar reconstruction in Europe stimulated innovation and transformed European bridge design and construction practice. America, during the 1950s and 1960s, however, felt no comparable urgency. As a result, there has been a noticeable lag in innovation and the transfer of technology from Europe to the United States—and there are many opportunities for improvements in the efficiency of design and construction methods in the United States. It is crucial that suppliers and buyers of bridge design and construction services in the United States recognize that innovation is necessary and that the barriers to enriching daily practice with the new technology lie not in technology itself but in the limited roles and expectations that have become traditional to bridge designers, contractors, and owners.

In particular, in the United States a contractor building a bridge works under the supervision of the engineer responsible to the owner for compliance of the work with the contract documents and the design intent. This U.S. practice is not universal. In Europe, bridge work is usually handled using the design/build concept. Under this arrangement, the owner/agency specifies the project requirements and calls for bids for design and construction. This approach has the advantage of promoting innovation at all stages of design and construction by allowing the design engineer to consider both the type of structure most suited to the site as well as the cost and time of construction.

If U.S. bridge design and construction teams are going to innovate in analytical and construction techniques (or be more effective at adopting innovations that originate elsewhere), adjustments are needed in the ways that clients, legal counselors, insurance carriers, contractors, and engineers conduct their professional and business activities. Moreover, the engineering community, regulating agencies, government, and universities must reconsider their respective roles if the United States is to produce the required innovation. Finally, the local, state, and federal agencies that buy bridges have the responsibility of identifying and managing procurement procedures that promote the best bridge design and efficient construction.

It is worth examining the postwar experience in Europe—an experience that created substantial innovation in bridge design and construction—for lessons applicable to U.S. policy.

THE EUROPEAN EXPERIENCE

Europe's dominance in bridge engineering and construction emerged during reconstruction efforts after World War II, when hundreds of major structures had to be replaced rapidly under tough economic and industry conditions. Steel fabricators were not ready to roll structural steel for bridge applications, so as concrete substituted for steel as the dominant material for bridge building, the combined design/build concept was adopted to encourage competition and reduce design and construction time.

The postwar European experience can be divided into two distinct historical decades, each with its own special character. In the 1950s Ulrich Finsterwalder introduced cast-in-place, segmental balanced cantilever prestressed concrete construction, and the structures built between the introduction of this technology and 1962 have been generally labeled as the first generation of modem bridge development.

In the 1950s it was demonstrated that by using the "balanced cantilever" method of erection, precast or cast-in-place concrete segments could be joined together progressively on top of a pier to form half of a bridge span on each side of the pier. In similar fashion and simultaneously, concrete segments erected on another pier, opposite the first, could be used to close the center span. This method of erection, used repetitively, created multispan bridges with very long center spans—as long as 750 feet (230 meters). Balanced cantilever construction also offers the advantages of eliminating ground-supported formwork over deep valleys, navigable channels, and congested urban or industrial areas; minimal disturbance to the surrounding terrain, allowing overhead construction while maintaining traffic; and construction that could proceed regardless of weather conditions.

The design of balanced cantilever bridges is complex and involves construction procedures that require specialized casting and erection equipment. But since most European governmental contractual arrangements permit design and construction firms to work together to develop innovative and cost-saving bridge designs, the casting and erection equipment designs were, and still are, considered at the same time as cost/design proposals. Thus, joint ventures between design firms with engineering excellence and construction firms with experience and equipment encouraged the development of in-house engineering expertise by contractors.

In such an atmosphere of engineering competition, new innovative designs and construction practices advanced rapidly. The state of the art of designing and constructing segmental prestressed concrete box girder bridges advanced in response to the need for more productivity in the construction of bridge structures. During the 1950s new prestressing systems and alternative methods for constructing segmental prestressed concrete bridges were developed, and design and construction procedures were refined. However, in the 1960s a major slowdown in the development of concrete bridge technology took place when some of the first structures built in the 1950s showed signs of distress. The Federal Republic of Germany took drastic measures banning several of the newly developed schemes.

Engineers, builders, and owners wanted to know what was happening, how to take care of the problems, and how to retrofit these distressed structures for service. Intensive investigations identified specific problem areas; and predictably, as in most technological developments, the problems occurred in areas where theoretical assumptions had substituted for factual experience. Time-dependent effects on concrete structures proved to be highly uncertain, with their true magnitude revealed only as the years passed. Furthermore, the newly introduced construction schemes overwhelmed existing analytical bookkeeping methods, which were both cumbersome and prone to computational errors.

For example, to mathematically simulate the construction and stress history of a balanced cantilever bridge, computations must be updated at each stage of construction to reflect changes in the structural behavior of the bridge. Every time a segment is added or a tendon is tensioned, the structural system must be reanalyzed. Moreover, time causes the concrete and prestress to creep, shrink, and relax. Furthermore, when the statical scheme changes or when two adjoining cantilevers are made continuous, stress redistribution takes place and must be recalculated. This redistribution is not only the effect of the staged sequence of construction but also of time-dependent effects that keep taking place in concrete and in prestressing steel long after construction ends.

In summary, the problem turned out to be a lack of understanding about time-dependent effects on concrete and prestressing steel. Repair procedures developed to restore the structural integrity of the distressed bridges also promulgated new and more conservative guidelines to regulate design and construction of bridges using this technology.

This major historical event clearly defines the transition to what has been called the second generation of modern bridges.

By the 1960s long-span, segmental bridges had established their competitive position. The industry addressed the problems encountered in the first generation of bridges through more research and advanced computational methods. Also, contractors started allowing better plant quality control in concrete production, plus getting the benefit of reducing creep and shrinkage in the finished structure, by letting segments undergo their natural and unrestrained strain changes during storage in the casting yards before they were finally incorporated in the structure.

Construction methods advanced rapidly, and segmental bridges were used in a wide range of erection schemes. In addition to the balanced cantilever method, span-by-span, incremental launching, progressive placing, and various combinations or modifications of these basic schemes were used.

Also, cable-stayed bridges in steel and concrete became the solution for spans longer than 700 feet, the limit for girder bridges. They were developed as an extrapolation of the balanced cantilever scheme: the cables replaced the post-tensioning tendons projecting outside the structure causing a gain in moment capacity due to the increased moment arm.

The development of cable-stayed bridges, inhibited at first by the analytical complexity and stay-cable fatigue, moved forward with developments in computer hardware and software and with developments in modem materials technology and advanced testing facilities.

Almost simultaneous with the development of the second generation of European bridges were major developments in segmental bridges and cable-stayed bridges taking place in South America, Canada, and Mexico. Today there are more segmental bridges in South America and Canada than in the United States, and more cable-stayed bridges in South America, Canada, and Mexico than in Europe. In South America the demands for new roads to develop virgin areas and to solve traffic problems in the growing and congested urban areas have produced record bridge construction rates and started important trends. Figures 1 and 2 show construction on a modem cable-stayed bridge and the final product.

THE U.S. EXPERIENCE

Based on recent bridge inventory and inspection programs, the United States has recognized the need to replace or to rehabilitate approximately 50 percent of the major bridges in the country. Funds for this work are being generated by new local and federal gasoline taxes, but the U.S. practice of bridge building has not advanced or incorporated the European formula, in which designers, contractors, and owner/agencies work together to meet the common objective of building bridges that are more time and cost efficient.

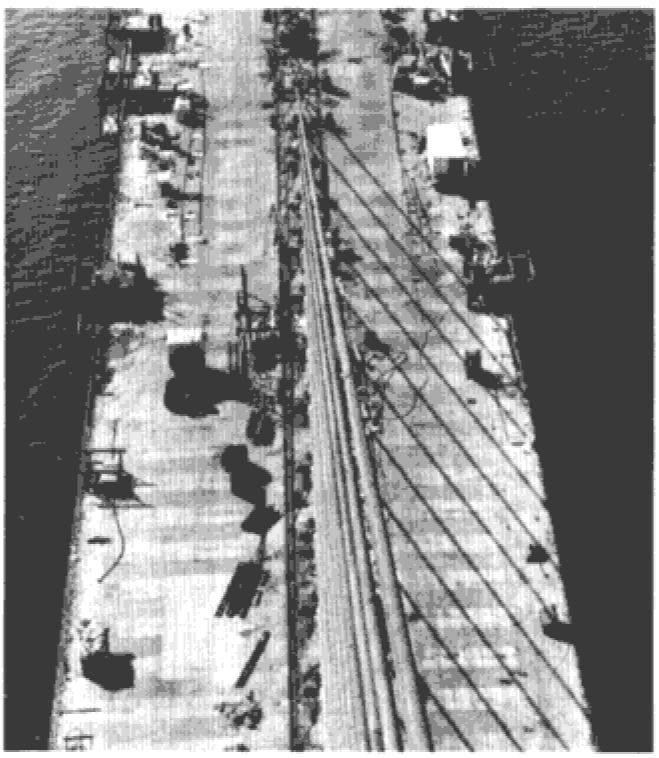

Figure 1 Construction in progress on modern cable-stayed bridge.

If we look back, it is clear that while major developments in European bridge technology were taking place, no similar efforts were under way in the United States. Not until the European market was saturated did the Europeans try to export their technology.

The major obstacle to ready acceptance of the European experience was the U.S. construction industry's structure, which clearly allotted the responsibilities and functions for design and for construction to two different services groups. Those in design services were not involved with contractors, and contractors in turn did not consider engineering functions as part of their construction contract responsibilities. The process that could have combined the two functions, the design/build bidding approach, particularly in the field of bridge and highway work, was not their standard practice. Figure 3 illustrates the traditional structure of the U.S. bridge design and construction industry.

The first bridges built in the United States based on the European experience were the Pine Valley Creek Bridge in California in 1969, using cast-in-place segmental construction with traveling forms, and the JFK Memorial Causeway Bridge, which opened to traffic in 1973. The precast segmental box girder for the JFK bridge was built in Corpus Christi, Texas, as a pilot project following a comprehensive model test program at the University of Texas at Austin. After concluding that the segmental bridge model safely carried the service and ultimate loads for all critical moment and shear loading configurations established by the analysis, a clean bill of health was given to proceed with actual construction. Moreover, theoretical calculations agreed with experimental results and actual construction.

After this major event took place, engineers, specialized material suppliers, and contractors saw the business opportunity and imported this technology when the economic conditions of the early 1970s demanded it. By the end of the 1970s the FHWA issued a directive requiring competing, alternative designs for steel and concrete for any major bridges in the United States. This opened the doors for major changes in bridge technology.

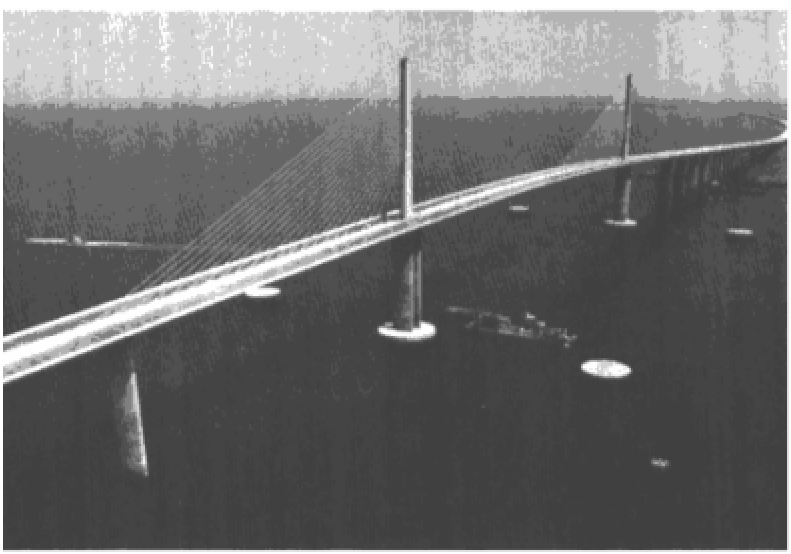

Figure 2 Cable-stayed bridge.

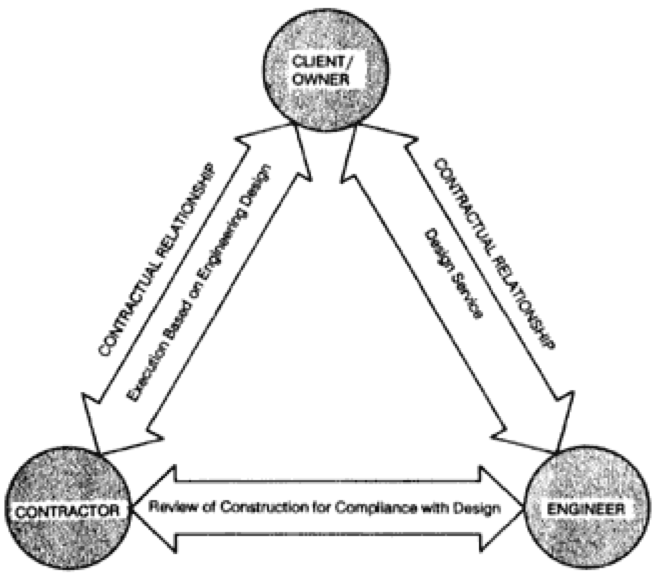

Figure 3 Functional relationship in a typical bridge project in U.S. Practice.

Successful innovation came first in concrete bridge construction, followed by improvements in steel bridges. Small and new design firms rushed in to take control of the concrete bridge design market. To match the competitive success of concrete bridges, the steel industry had to offer free engineering to design firms to improve the analytical capabilities of the designers. In addition, the steel industry had to lobby extensively to modify the established design criteria and the controlling policies to pave the way for new design concepts in steel. Research and development also improved national applications for steel, including the improvement of site assembly by using steel erectors that duplicated the effective erection schemes of concrete segmental construction, as in the case of the Baytown Bridge (bid in 1986).

The design competition requirement on all federally funded long-span bridges (discussed in the next section) forced changes in material use by demanding plant control tolerances, for example, reducing field fabrication to a minimum, which improves quality, and stricter quality controls overall. Such requirements enabled both more and better construction control and improved fabrication techniques, and encouraged engineers and contractors to use more sophisticated analytical methods under more demanding quality control conditions.

CURRENT TRENDS AND BARRIERS

Two aspects of the procurement process for bridge design and construction are especially important to stimulating innovation in the U.S. bridge industry. First, current bridge design and construction procedures require that alternative steel and concrete designs compete on all major federally funded projects. This is more than a legal requirement—the "extra" design opportunity is a challenge to incorporate the innovative techniques that can give a competing design an important advantage in performance or constructibility. Second, after the design phase is completed, but before a bridge project begins, price competition among contractors drives innovation in the means and methods of construction, factors that can dramatically affect the price tag on a structure.

The alternative design requirement also means that the contractor has alternatives under redesign clauses in the bidding documents. These redesign clauses fall into two categories: (1) the construction options allowed within the general design and (2) more significant changes made according to the "value engineering" concept.

Required alternative designs are not new. In the 1950s prestressed concrete structures had to be presented as alternatives to conventional designs. In 1978 the FHWA issued a directive to encourage alternative design on all major bridges in an attempt to fight inflation and a then-current trend of bids coming in 5 percent over the engineer's estimate. The alternative design program in effect since then has produced very competitive bids, reversing the previous trend and producing more projects bid at or under the engineer's estimate. In the state of Florida the cost of major bridges in the 1980s has remained constant because of improved designs and construction schemes that offset the rising costs of materials and labor.

However, innovation in bridge design and construction has been sharply curtailed by legal and professional liability concerns spawned by the openness in the contract documents that allows contractors more flexibility. Liability concerns have also changed the nature of the relationship between the engineer and the contractor.

The future success of alternative designs will depend on the structuring of the contract documents to preserve the contractor's high degree of flexibility in those optional details and construction methods that do not modify the design intent. Contracts must also keep the responsibility and liability separate and under the two distinct categories of design intent and contractor ways and means. Properly structured contract documents must provide clarity without restricting freedom; that is, true performance specifications must preserve the basic design requirements without precluding further development within these requirements. Also, policy changes in the functions engineers perform must be adequately addressed.

The current policy as laid out in "Alternate Bridge Designs" prepared by the U.S. Department of Transportation FHWA Technical Advisory (T 5140.12) is a start, but many of the other lessons learned still need to be incorporated. Even though the closer relationship between design and construction activities has clearly established closer ties between engineers and contractors, two major issues continue to inhibit the acceptance of contractor alternative design/build proposals.

These two issues are the twin questions of who pays for design modifications and who takes ultimate responsibility for the design. In a traditional design relationship the owner backs up the engineer unless he or she is found to be negligent. On a contractor design/build alternative the owner does not want to cosign on the design. In some instances, owners have taken the stand that if the contractor's alternative design will cost more than the client's or require modifications, then the contractor should abandon the alternative and build the client's design for the alternative bid price. From an insurance and legal point of view, this scenario has resulted in various claims, lawsuits, and increased professional liability premiums.

Antagonistic positions definitely do not help improve efficiency. Therefore, an effort to cooperate in defining mutually acceptable new standards will require a joint effort between the client's policymaking agencies, and the construction industry and design services groups. Clients must embrace better alternative design procedures to lift the barriers that restrict and could eventually stop the entire practice of alternative design. They need to look at alternative design as an economic solution to rising costs based on competitive bidding, and then be objective and open to reasonable changes required by state-of-the-art technology. Insurance carriers need to consider departures from existing practice as risks that are insurable, with broader premium ranges based on case-by-case assessments of practices developed to allow innovation. The current claims atmosphere is definitely an obstacle to innovation.

Current practices suggest that although the roles of engineers and contractors are still separate, alternative designs are emerging despite the reduction in the range of contractors' options for design modifications. For example, the engineer now produces schematic erection drawings that define the range of construction loads and the types of erection methods by specifying the structural behavior of the permanent structure; then the contractor, operating within this limited range, takes full responsibility for the design and sizing of equipment and temporary works. Technical specifications are written to be more performance-oriented, with strict tolerance requirements to achieve the trimmed-down designs.

The role of the engineering community in this field is clear. We need to further clarify the design intent and responsibility of all three parties involved in a bridge project in the contract documents. The current procedure for preparing these documents, for example, a linear process in which the designer drafts the contract according to the owner's procurement guidelines for construction contractors' bids, is still a big problem.

POLICY LESSONS FOR THE UNITED STATES

In an atmosphere where the cost pressures on public works projects such as bridge construction continue, efficiency gains are most likely to come only through innovation, and, in policy, from lessons learned from European experience. In particular, closer relationships between the design function and the construction function appear to be a better formula for innovation than the separation typically practiced in the United States. Although neither the legal nor insurance environment is likely to change much in the near future, it is the bridge owner's procurement guidelines—something that is amenable to modification—that currently force designers and builders into separate roles. There is much room for improvement in the owner's guidelines for considering innovation and technical advances in design and construction methods as an ongoing process. The designer, in turn, will also need to reflect this process in construction specifications that leave room for the contractor's innovations. Today, the major obstacle slowing the evolution in bridge engineering is the limited importance accorded innovation. This is reflected in current design specifications that do not recognize new design and construction schemes and standard construction specifications that fall short of the level of development in design, construction techniques, and materials. If we are to incorporate new bridge technology into U.S. practice, the owner, designer, and contractor must each be free to innovate and to work together toward achieving one objective rather than adopting adversarial roles that have made the most successful projects a lawsuit paradise.

Professional Services Firms and Information Technology: Ongoing Search for Sustained Competitive Advantage

Thomas H. Doorley, Alison Gregg, and Christopher Gagnon

The "professional services" comprise a diverse group of industries including law, management consulting, investment banking, accounting, architecture, and engineering. The professional services merit special attention because they represent the confluence of two of the most important trends in our economy: (1) growth of the services sector; and (2) emergence of the "knowledge worker."

These firms, as a rule, are high value-added enterprises that create economic value through the development of the skills of their employees leveraged through technical and organizational systems (Maister, 1982).

Yet, despite their importance, professional services firms (PSFs) have been among the least analyzed and understood in American industry. In fact, there are significant strategic and operational differences between PSFs and traditional product firms. For example, much of the traditional wisdom regarding strategy and competitive advantage was developed for goods firms and does not seem to hold for PSFs.

This paper will explore the nature of PSFs, especially in relation to the application of information technology for competitive advantage. We will

1. present some background about PSFs;
2. discuss relevant phenomena we have measured and observed;
3. explain these phenomena in relation to the way PSFs and technology interact;
4. estimate the more important ways in which technology will affect the professional services; and
5. suggest mechanisms managers and technologists should use in view of these changes.

TABLE 1 Professional Services Components of GNP for year 1986 (in billions)

GNP	$4,201.3
Business services	$ 162.8
Legal services	$ 52.3
Misc. professional services	$ 70.1
Total	$ 285.2
Total as a % of GNP	6.8%

SOURCE: U.S. Bureau of Economic Analysis.

PROFESSIONAL SERVICES—EXPANDING IMPORTANCE

The problems of measurement and the lack of reliable statistics in the services sector are well documented (Aanestad, 1987). However, we estimate that the professional services constitute 6.8 percent of U.S. GNP (Table 1). These industries have been exhibiting solid growth for the last 25 years. As shown in Table 2, rather than moderating, this growth has accelerated of late. But, the statistics are not all so positive. The importing of business services into the United States has grown faster than our exports of services to the point that our trade surplus in business services is almost nonexistent (McMeans, 1986).

INDUSTRY STRUCTURE—ECONOMIES OF SCOPE

Increasingly, a fundamental strategic theme is being sounded within PSFs, that is, the attempt to capture competitive advantage by broadening the scope of the enterprises's activities, using the leverage of economies of scope. Examples of this include the expansion of the Big Eight accounting firms into management consulting, the emergence of "one-stop" shopping for

TABLE 2 Growth in Professional Services (percent)

	1960–1984	1980–1984
Business services	6.8	7.2
Legal services	4.6	6.4
Securities and brokerage	4.6	16.9

financial or marketing services, and the breadth of services offered by architectural and engineering design firms.

SOURCE: Department of Commerce cited in *New York Times* 10/27/85 "Services: Where the Growth Is."

Table 3 shows the percentage of 1986 revenues Big Eight accounting firms had in different businesses. Although the accounting and auditing function represents the initial and traditional business of these firms, the growth of the other services will cause most of them to have more than half of their revenues outside the original services within 5 years.

The top engineering design firms also had their billings distributed across a variety of businesses (Figure 1). Many of these engineering firms believe that expansion into other areas will be facilitated by current experience. For example, CH2M Hill, a consulting engineering firm specializing in environmental design believes its experience in water and wastewater design gives the firm an advantage while competing for hazardous waste jobs.

The dilemma facing the firms will be the trade-off between concentration on a single line of business that may be subject to greater volatility versus the complexity of multiple activities whose diversity may dampen individual business fluctuations. Information technology's role in providing timely, accurate operating data increases as businesses move away from a single core activity.

INDUSTRY STRUCTURE—ECONOMIES OF SCALE?

In product-based businesses there is a demonstrable link between profit-ability and relative scale as was postulated by Bruce Henderson (1979) and others. Braxton Associates has observed this relationship in many of our clients' businesses as the following example from the passive component industry illustrates (Figure 2). Based on this and similar analyses, managers

TABLE 3 Diversification of the Big Eight

Company	Audit (%)	Mgmt. Consulting (%)	Tax (%)	U.S. Revenues
Arthur Andersen	47	32	21	$1350
Touche Ross	55	22	23	565
Peat Marwick Main	56	18	26	1080
Ernst & Whinney	58	20	22	900
Arthur Young	59	15	26	615
Coopers & Lybrand	61	18	21	855
Price Waterhouse	61	17	22	740
Deloitte Haskins & Sells	66	12	22	595

SOURCE: Public Accounting Report, 3/15/87. American Institute of Certified Public Accountants Annual Report, 1987.

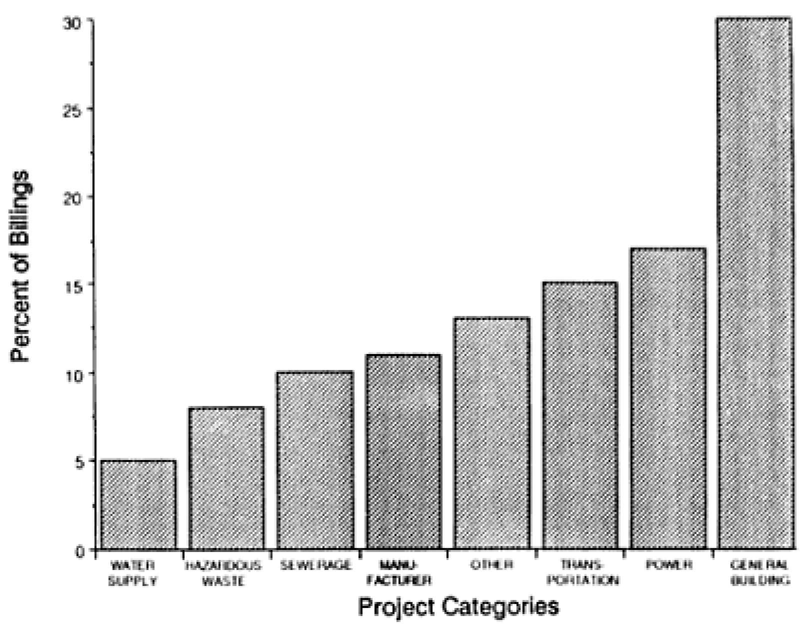

Figure 1 Top 500 design firms: billings distribution.

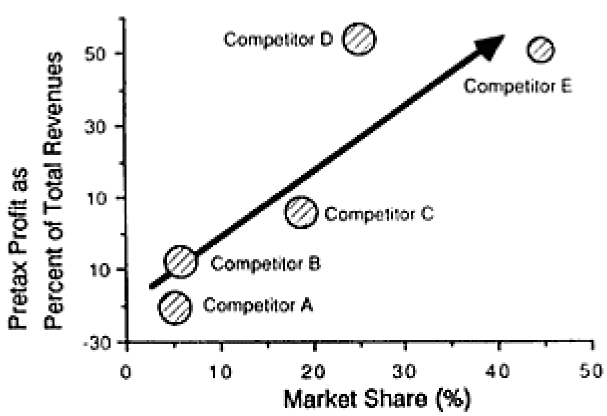

Figure 2 Profitability-scale relationship: passive electronic components. SOURCE: Braxton Associates.

in product-based firms formulate and implement strategies with the economies of scale relationship as a core, underlying concept.

TABLE 4 Concentration in Professional Services (Percentage of Total U.S. Industry Revenues)

	Top 4 firms	Top 8 firms
Management consulting	3.4	5.2
Advertising	2.0	3.7
Legal services	1.0	1.8
Executive placement	6.5	

SOURCE: Braxton Associates.

Managers in many PSFs develop large-scale oriented strategies as if similar laws of economic behavior apply to them as well. However, to date, absolute size or market share seem to have less inherent value in professional services industries than in others. For example, if large scale were of value, industry concentration rates would be high. But, concentration in these industries tends to be very low (Table 4); further, there seems to be little or no discernible trend toward increased concentration.

Also, if economies of scale applied for PSFs a profitability-scale relationship similar to Figure 1 should apply. Figures 3 and 4 indicate the opposite; there is no systematic relationship between the size of advertising and law firms and their profitability. Our research in other professional services sectors reinforces this conclusion.

Scale may not translate into other sources of competitive advantage either.

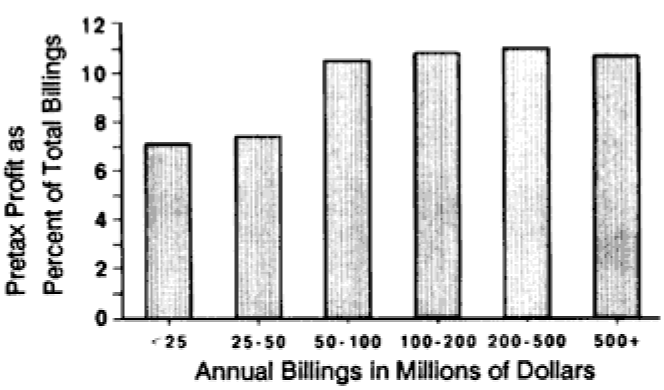

Figure 3 Profitability-scale relationship: U.S. advertising agencies.
SOURCE: Braxton Associates.

Figure 4 Profitability-scale relationship: U.S. law firms. The most profitable law firms in 1986
were ranked 11th and 29th in revenues.

For example, there do not appear to be marketing advantages of scale in the accounting and auditing business. A two-part study in *Accounting Review* concluded that there was no significant shift of revenue from smaller firms to the Big Eight during the period 1964–1980 (Danos and Eichenseher, 1986). That is, there was no increase in industry concentration and, therefore, no change in competitive status related to scale.

Increasingly though, managers of PSFs appear to be adopting strategies based on a belief in a direct profitability-scale relationship. For example, Saatchi & Saatchi is pursuing an acquisition-based strategy to build the world's largest advertising and marketing services firm. Whether this will yield profitability benefits remains to be seen. History is clearly against the success of the strategy on two counts: limited (if any) positive scale effects and the difficulty all organizations face attempting to capture synergy across multiple activities.

THE GROWTH-PROFITABILITY FUNDAMENTAL

Scale then has not generated profitability to the same degree in PSFs as in goods firms. However, if the value of scale is diminished, the role of growth is heightened. As Figures 5 and 6 demonstrate, scale and profitability for insurance brokers do not correlate, but growth and profitability do.

Figure 7 portrays similar data for a peer group of large Touche Ross offices. Again, the direct relationship between growth and profitability is depicted. Other sectors follow a similar trend.

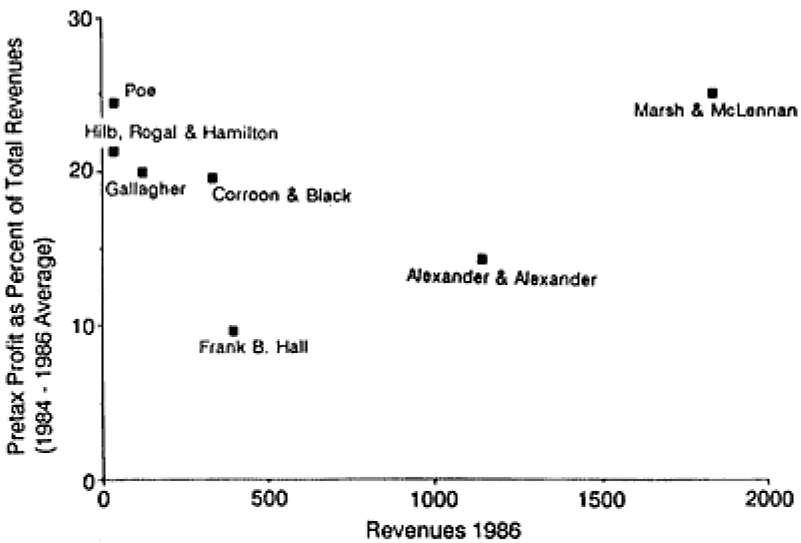

Figure 5 Size versus profitability: insurance brokers. SOURCE: Company annual reports.

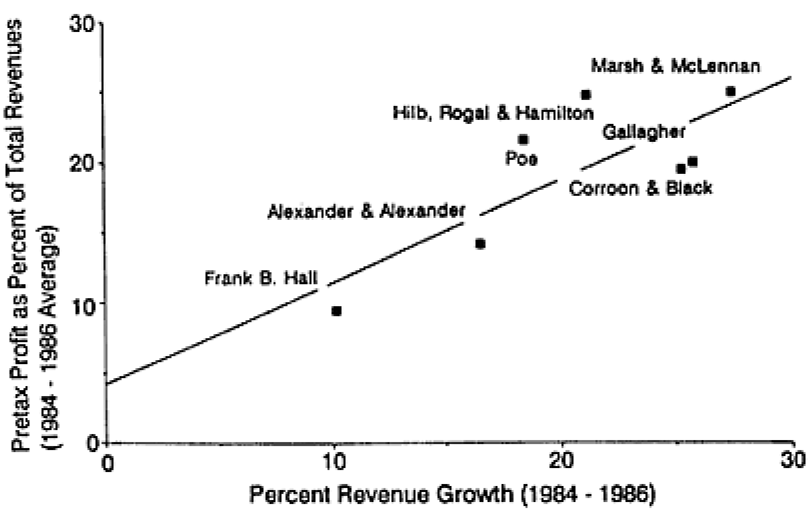

Figure 6 Growth versus profitability: insurance brokers. SOURCE: Company annual reports.

From these observations flows the following broad strategic direction for PSF managers:

- Growth and profitability are not independent, they are linked.

To complete the thought, our experience demonstrates two additional points:

- Growth must be market driven; and
- Organic (internally generated) growth has a greater expected value than growth through acquisition.

Thus, the debate about growth within a PSF should focus on "how to" achieve it (e.g., which markets? who will lead the charge?) not on "whether" to grow. Whether to grow is a given.

THE DIFFICULTY OF SUSTAINED GROWTH—THE "HOT-HAND" EFFECT

Another fundamental phenomenon operates that sheds light on the nature of professional services, namely, the "hot-hand" effect. The hot-hand effect concerns the relative difficulty PSFs have in sustaining higher than average growth rates in their industry.

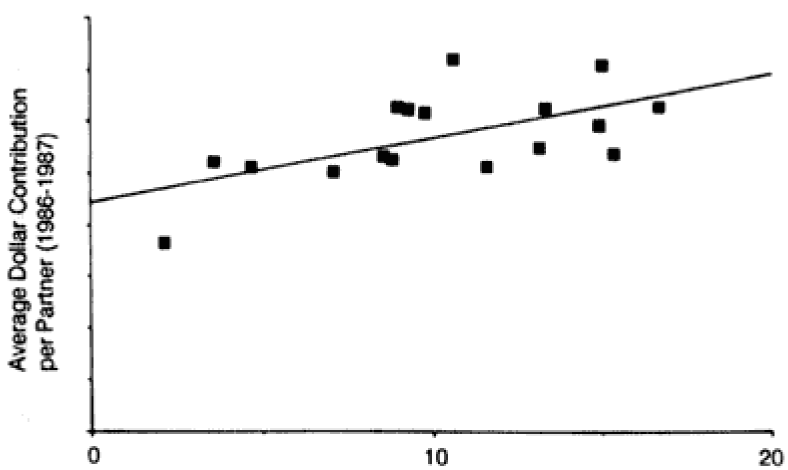

Figure 7 Touche Ross Offices: large office peer group. Actual dollar amounts are not given because that is proprietary information.

The hot-hand effect can be seen using the "sector graph," a display Braxton developed to capture competitive dynamics within a market segment. Firms that are sustaining higher than average growth move upward and to the left on the graph. For example, Figure 8 shows the U.S. advertising industry for the period 1979–1981. Growth for the industry as a whole was approximately 5 percent for the period but some firms, such as Doyle Dane and HBM Creamer had much higher growth. Note that in the next 3 years (Figure 9), these two firms had much lower than average growth rates while other players, such as Ted Bates and BBDO, grew much faster than the industry.

Finally, in the period 1983–1985 two different companies (Ogilvy Group and Grey) had outstanding performance, while the previous period's highest performers drifted toward the pack (Figure 10). Note that during this 7-year period no company was able to gain significantly in relative scale.

In many professional services industries not only is it difficult to gain abnormal share, but it is equally difficult to differentiate services in the clients' mind. Figure 11 shows that clients perceive no significant difference in quality of services among the Big Eight accounting firms.

Why is it that gaining advantage through scale or quality is so difficult for PSFs to achieve? One answer is that they are limited by their scarcest resource—quality people. As a firm tries to translate any near-term advantage into high long-term growth, it must compete to hire and retain qualified professionals. Professionals enjoy more mobility today than ever before and feel less loyalty to their employers if offered more money, freedom, or responsibility elsewhere. This issue is becoming even more acute as the "baby bust" generation enters the work force.

THE INDUSTRY IN TRANSITION

One of the ways PSFs may overcome the hot-hand effect is through the strategic application of information technology to leverage their scarcest resource, the talent of the individuals who make up the firm. Well-applied information technology tends to increase the knowledge, skills, and value of individuals. This enhances the firm's ability to pay them more. Information technology holds the promise of making centralized information, both data and expertise, more widely available, aiding communication within a PSF, and most importantly, contributing to the development of human capital.

But, technology as a route to securing a sustainable competitive advantage for a PSF is currently hypothetical. The facts of the recent past and the current setting do not yet support the hope. If an industry transition is under way, the competitor that makes a bold move and executes effectively can make an unprecedented gain. If no transition ensues, investment may be counterproductive.

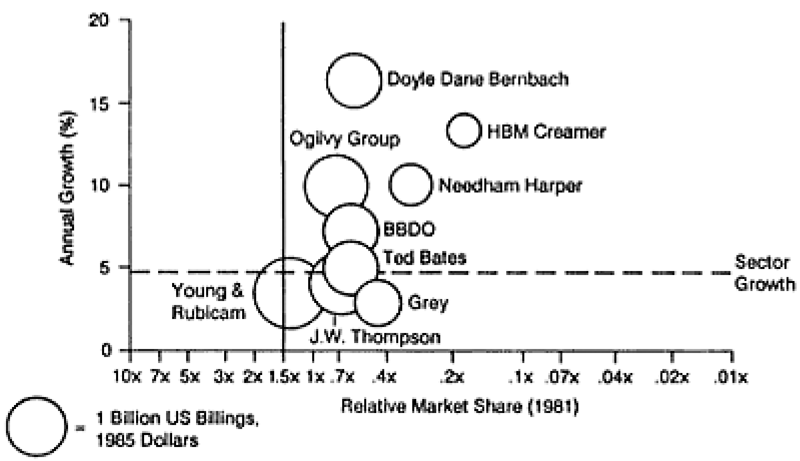

Figure 8 U.S. advertising agencies: U.S. billings sector graph 1979–1981. SOURCE: Braxton Associates.
The largest firm is compared with the second largest firm. All other firms are normalized on the largest firm. This display method emphasizes the difference between the largest and second largest firms.

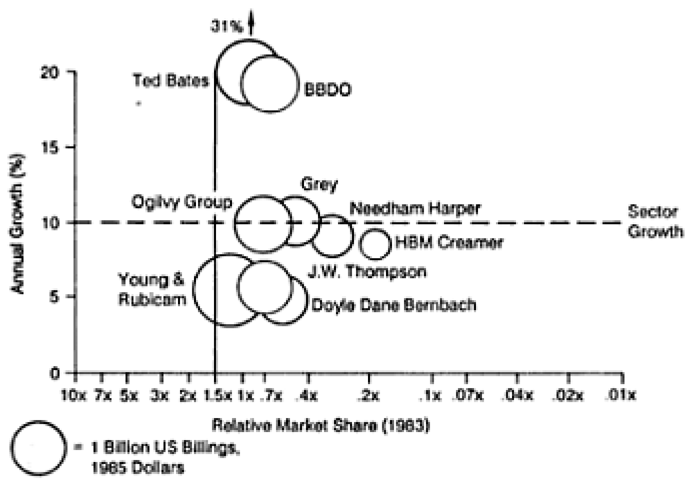

Figure 9 U.S. advertising agencies: U.S. billings sector graph 1981–1983. SOURCE: Braxton Associates.
The largest firm is compared with the second largest firm. All other firms are normalized on the largest firm. The display method emphasizes the difference between the largest and second largest firms.

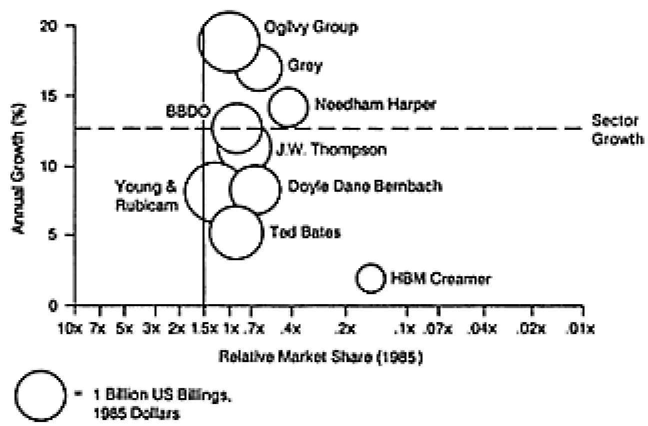

Figure 10 U.S. advertising agencies: U.S. billings sectorgraph 1983–1985. SOURCE: Braxton Associates.

The largest firm is compared with the second largest firm. All other firms are normalized on the largest firm. The display method emphasizes the difference between the largest and second largest firms.

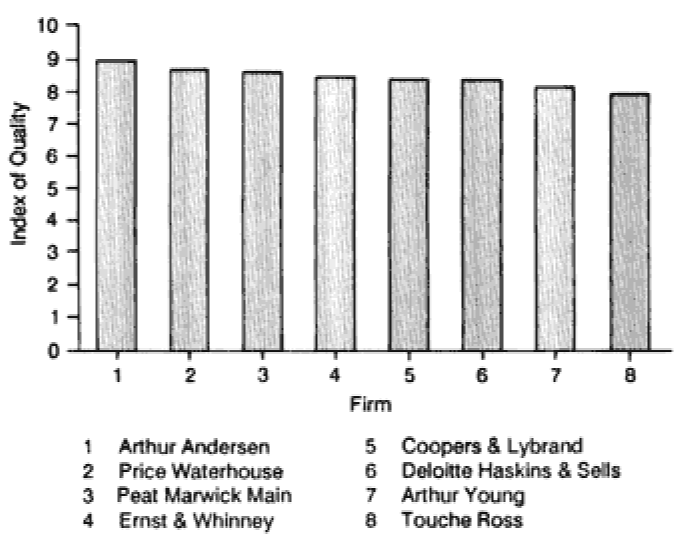

Figure 11 Service to clients: Big Eight firms. SOURCE: Survey of College Professors.

INFORMATION TECHNOLOGY

Increasingly, PSFs are aware of the potential to utilize computers in their business. For example, Peat Marwick Main & Co. believes that the future of its auditing and consulting practices is "inexorably tied to technology" (Spindler, 1986). Arthur Andersen has worked to develop an approach to "Information for Competitive Advantage" (Noling and Blumenthal, 1986).

PSFs have bought and are buying computers, communications equipment, and software. So much so, that in 1983 the capital endowment per worker in the "information sector" surpassed that in the traditional industrial sector (Roach, 1988). We estimate that accounting and legal firms spend an average of approximately 2 percent of billings (revenues) on information technology and software, not counting service fees or internal development expense.

Computer utilization in many professional services industries ranks among the highest. A study of 100,000 households, aimed at determining the incidence of personal computers in the U.S. workplace, determined that nine of the top ten industries in terms of percentage of workers using a personal computer are services industries. At least five of these industries have components that are incorporated in our definition of professional services (Figure 12).

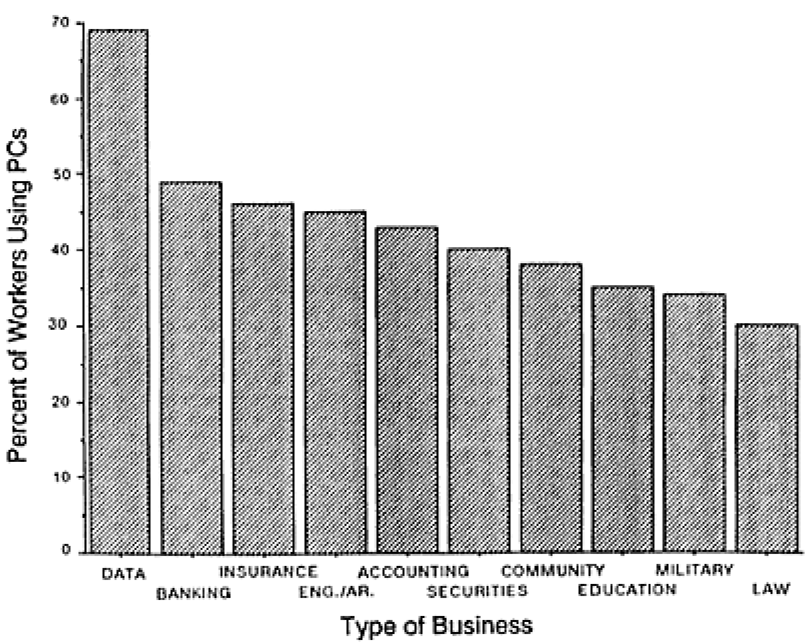

Figure 12 PC use by industry. SOURCE: Future Views, 10/28/1986.

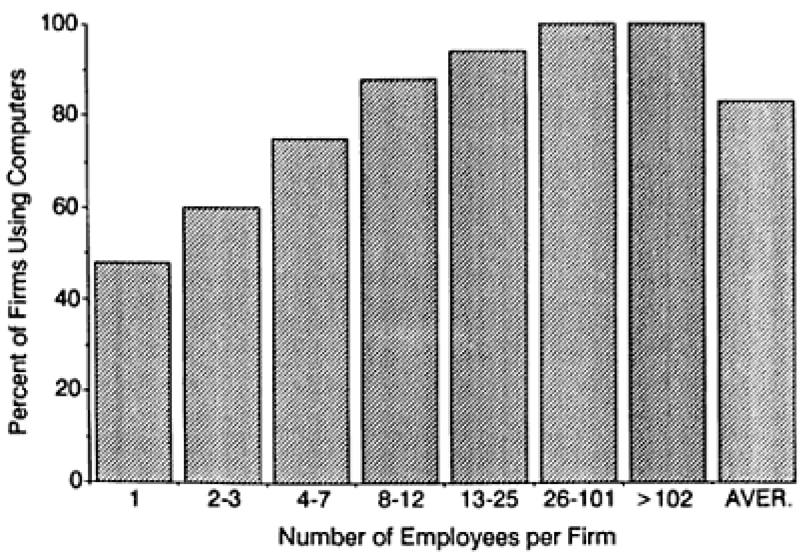

Figure 13 Consulting engineers: use of computers by size of firm. SOURCE: Computers and the Design Professions, American Consulting Engineers Council, 1985.

Traditionally an industry with high utilization of technology, consulting engineering firms, reflect this high incidence of computer use. In the most recent survey on computer use in consulting engineering firms the American Consulting Engineers Council found that more than 80 percent of all consulting engineering firms are computer users. All firms with more than 150 employees were computer users, whereas the average size of noncomputer users was just 7 employees. Figure 13 shows the distribution of computer users by firm size.

These data suggest that technology may be scale related, that is, the larger firms can afford to invest in software and hardware. If this is the case, scale economies in PSFs may be possible. If the greater ability and propensity to invest in technology can be converted into strong, sustained competitive performance (e.g., higher growth, better profitability), scale effects would be evident, a transition from "no-scale" to "high-scale" would be under way.

The top three computer applications are not engineering task specific, but rather are applied to common, shared activities, that is, basic business applications of accounting, management of information, and project management (Table 5). Pure engineering applications followed, ranked consistently with the most frequently offered services in the industry.

TABLE 5 Computer Applications: Consulting Engineering Industry

Application	Use (multiple responses)
Accounting	72%
Management information	69%
Project management	46%
Structural/building	40%
Geometry/surveying	39%
Hydrology/hydraulics	37%
HVAC/energy	32%
Bridge & highway	26%
Mechanical/plumbing	23%
Electrical/electronics	21%
Environmental	18%
CADD graphics	18%
Other graphics	15%
Traffic/transit	11%
Geotechnical	8%

SOURCE: American Consulting Engineering Council, 1985.

Several reasons explain the at first surprising result of nonengineering computer applications preceding engineering applications. The most obvious is that accounting, management of information, and project management are common activities shared by all engineering firms regardless of the type of services offered. Additionally, availability of software programs may also influence computer usage: 91 percent of the nonengineering software programs listed in the survey were purchased, whereas 60 percent of the engineering programs were purchased, 26 percent custom written, and 13 percent time-shared.

When asked about the ways they acquire engineering-oriented software, a majority of the consulting engineering firms responded that they used customer software, either written in-house (57 percent) or written by software consultants (36 percent). This implies that at the time of the survey, the commercially available software was inadequate to meet the industry's needs.

One would expect this intensification of information technology to show some significant productivity results. Many of the early signs, however, are not encouraging. Stephen Roach, a senior economist at Morgan Stanley has concluded that there have been no significant gains in white collar productivity since the 1960s. Measurements of output in the services sector are notoriously difficult to gather and usually fail to measure the quality, complexity, timing, and flexibility gains in services output. Still, it is not surprising that many professional services executives have "a growing dissatisfaction with high tech's productivity payoff" (Bowen, 1986). Furthermore, it is difficult to find examples of PSFs who have used information technology to establish a long-term sustainable advantage over their competitors. Thus, the payoff from investment in information technology is not obvious.

THE NATURE OF INFORMATION IN A PSF

PSFs compete in three basic ways according to David Maister (Maister, 1986)

- Experience (or the "gray hairs"). Firms that compete on the basis of experience sell the ability to perform certain kinds of work that has been performed before.
- Expertise (or the "smartest kid on the block"). Firms that compete on the basis of expertise hire a relatively small number of very bright professionals with specific expertise that can be sold as customized work at high margins. These firms tend to do ad hoc or creative work, attacking unique problems with a process rather than a canned solution.
- Execution (or the "cheapest—most efficient of a number of competent firms"). These firms offer a number of standardized services that they can perform at the lowest unit cost. These companies function within strict guidelines and well-defined bureaucracy. They tend to have many relatively junior staff performing "back-office" type work.

Each of these three strategies holds different implications for the application of information technology. Figure 14 shows a simplified model of information flows within a PSF. A firm has a body of knowledge (which

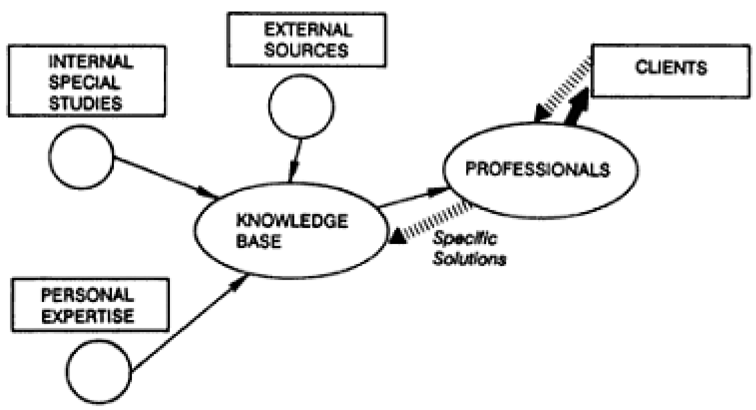

Figure 14 Information flows in professional services firms.

can be either subject knowledge or process methodology) that it resells to clients through its professional staff. There are three key stages in this model. First, is the storage of the firm's knowledge base in some form, be it human, paper, or electronic so it is accessible by the members of the firm. Second, is the means of communicating that knowledge or expertise throughout the firm.

Finally (this is where most information systems perform most poorly), there must be some feedback loop in which information from the client(s), delivery nodes, or other sources can be incorporated into the knowledge base of the PSF for the future. Information flowing into a firm-wide knowledge base, then out to the staff (arrows in Figure 14) works inconsistently within most PSFs.

The point where information technology can be used most effectively to leverage a firm's capabilities will be determined by which of the strategies described above it follows. All the technologies discussed below will be of at least some interest to most firms in the future.

The productive application of information technology in each of the three types of strategies will vary by type of firm as follows:

Experience. In an experience-based firm, two of the most critical success factors are (1) leveraging the knowledge of senior personnel, and (2) developing the human capital at the junior and middle levels.

Information technology allows firms not only to make information available to client services staff but also to educate the firm's professionals about both hard data and the firm's methodology, assets, and culture. Shoshana Zuboff (1985) of the Harvard Business School calls this process "informating" and describes the inculcation of the data base into organization structure. According to Zuboff, informated firms can and have "replicated" themselves, allowing firms to expand their markets geographically.

As this process takes place, firms can not only offer their services at lower costs but also free personnel to push their "state-of-the-art" by building on previous experiences. This development is a major contributor to the emergence of the multioffice law firm. Aided by internal and external data bases such as LEXIS, several firms with more than 500 lawyers have emerged (Glaberson, 1986) in the legal services sector. Many firms (e.g., Touche Ross among the Big Eight accountants) have successfully used a "Voice Box" concept where firm members can access the experience of members around the firm by entering their questions or problems over an electronic mail network.

Expertise. In a creative, expertise-based firm, computers are best used to aid professionals with more routine work and free their time for work that cannot be automated. David Maister relates that his clients typically feel that 50 percent of what professionals do in a given work day could be performed by more junior associates. For example, a Texas architectural firm found that its architects were spending too much time estimating and quoting proj ects. By automating this process, they estimated they were able to gain 10 times their investment and provide more accurate quotes (Wiseman, 1985). A Chicago law firm billed an extra $30,000/year per attorney by automatically tracking lawyers' phone time, much of which was going unbilled (Wiseman, 1985). Designers using CAD technology can now spend less time on drawing and modeling tasks, although the computer cannot substitute for the conceptual and creative ability of the designer. Similar benefits can be seen in investment banking where technology has made it possible for new financial instruments to be created, tested, and delivered to the market much more quickly than ever before.

Execution. Execution-based firms must try to lower costs and reduce personnel limitations by reselling rule-based skills. Many professional services are, in essence, rule based; new services developed at the expertise and experience levels can quickly be translated into standardized services. For example, a small accounting firm raised revenues 50 percent by offering on-line accounting services to small businesses. This increase required only three new staff members. Arthur Young has developed an expert system called ASQ that guides auditors through key decisions in the audit process based on partners' experiences. Arthur Young feels that in addition to lowering costs the system improves quality and speeds turnaround.

PSFS AND TECHNOLOGY—THE DILEMMA

Although these examples hint at substantial "potential" for information technology, at least three significant barriers exist that serve to retard information technology investment: (1) bias toward "personal tangible" investment; (2) traditional focus on (near term) partner earnings; (3) inability of information technologists to solve problems quickly, effectively.

Our experience at Braxton indicates that overhead investment in PSFs most often takes the form of personal tangibles, that is, individual purchases that have benefits that are direct, controllable, visible, and easily understood. For example, a PSF partner recognizes the need for telephones, typewriters, and individuals to operate them. The partner knows they perform a valued function. He or she can see it daily and, therefore, can support the reduction of direct compensation that results from increased overhead.

PSFs are large purchasers of information technologies, but their purchases are usually personal tangibles, frequently taking the form of personal computers. For example, in 1987, accounting and law firms were projected to spend more than $500 million on computer hardware (Professional Publications, Inc., 1986). Although more than 90 percent of the firms surveyed are planning on buying personal computers, only 13 percent are going to buy minicomputers, and only 25 percent are planning on investing in network technologies, which require more planning and a longer time horizon to justify.

The computer survey by the American Consulting Engineers Council, referred to earlier, likewise supports the concept of investment in personal tangibles. Seventy-seven percent of all computers used by consulting engineering firms are microcomputers, and frequently a firm owns computers from multiple manufacturers. The evidence from these three professional services industries indicates that the purchases of information technology are driven by individual needs and are not reflective of a coherent, unified technology strategy for the firms.

Further, according to Maister, the partnership structure of many PSFs orients their financial goals toward net income per partner (Maister, 1984), increasing the short-term, personal tangible bias. A dollar invested in systems with possible long-range impact is a dollar out of the partners' pockets today. Long-range planning implies equity ownership. The partnership structure of many PSFs seems ill suited to properly encourage that kind of behavior. Some fight this malaise through carefully developed "corporate cultures" that motivate partners to take pride in and a long-term view of their firm.

This partnership-induced bias helps us to understand the nature of the hot-hand effect. Because PSFs are poorly suited to making long-range technology plans, they have difficulty using technology to leverage competitive advantage and turn it into ongoing growth. Without technology, the limiting factor is the near-term personnel capacity of the firm. This hiring problem is exacerbated by the fact that many PSFs have difficulty forecasting their work loads and, therefore, future personnel requirements.

To stimulate more innovative managerial thinking, it may be that some PSFs need to think more about adopting some of the characteristics of corporations as long-range decisions become more crucial to survival. The California Society of Certified Public Accountants, recognizing this problem, recently installed a long-range planning system that will extend beyond changes in leadership.

Those firms that have been able to follow a technology strategy have had senior partners with a vision. George Gordon, founder of Gordon & Glickson, has led his law firm to a profitable practice largely through his personal commitment to technology and innovation. Gordon & Glickson leverages its understanding of high technology to sell to clients whose businesses involve computers.

There is a third barrier the information industry must overcome before information technology can truly become a weapon for competitive advantage in the professional services. The professional services are constantly changing as client needs become more sophisticated; the systems development process is currently too long to respond effectively to clients' needs. The system development process is typically a separate function that does not include the professionals who are delivering the service. The Arthur Young ASQ system, described above, required 2 years and 50,000 person-hours to develop (McKee, 1986). As one executive said: "We have a short window of opportunity. If it takes two-and-a-half years to develop a system, it's gone!"

PSFS AND TECHNOLOGY—THE PROMISE

Clearly, the cultural barriers described above constrain PSFs from making significant information technology commitments. For the promise of information technology to be realized, it must overcome these barriers by delivering on two fronts: (1) enhancing results, and (2) tailored technology that can allow a specific competitor to gain, then maintain an advantage.

Although it may be a platitude to state that all services must enhance results (e.g., most consulting firms go to great lengths to promote how much value is added by their service to clients), information technology must have as its focus allowing the PSF to perform more effectively. Arthur Young's ASQ system is aimed appropriately at enabling the firm to deliver a better product more quickly and less expensively. If this and other programs can deliver observable results, the tendency to invest will increase. Hopefully, PSFs will perform postinvestment analyses to measure whether or not desired results were achieved in order to justify similar future activities.

Finally, use of generic software (or other non-firm-specific technology) benefits multiple firms and multiple clients. As such, no individual firm obtains an advantage. Investments in such technology are defensive, that is, each firm must invest to keep pace, but no firm can differentiate itself. Information technology will assist PSFs through an industry transition phase primarily to the extent that firm's specific competitive barriers can be created. This is the enticing competitive promise that keeps PSFs at least dabbling, if not making enormous commitments.

CONCLUDING NOTES

Historically, PSFs have made investments in information technology hesitantly and usually oriented toward an immediate return. Such an investment strategy was appropriate given the limited leverage such investments generated.

In the future, if the rules of the game change, for example, if scale has greater value than in the past, there will be substantially higher investment leverage. If so, and if information technologists can deliver firm-specific, responsive, timely solutions, information technology investments may both multiply and be able to deliver competitive advantage to those with a clear vision of how to use technology and a concomitant capacity to sacrifice current earnings for that vision.

References

Aanestad, J. M. 1987. Measurement problems of the service sector. Business Economics (April).

Bowen, W. 1986. The puny payoff from office computers. Fortune (May 26).

Danos, P., and J. W. Eichenseher. 1986. Long-term trends towards concentration in U.S. audit market. Accounting Review (October).

Glaberson, W. B. 1986. MegaFirms are taking over corporate law. Business Week (November 17).

Henderson, B. 1979. Henderson on Corporate Strategy. Cambridge, MA: Abt Books.

Maister, D. H. 1982. Balancing the professional service firm. Sloan Management Review (Fall).

Maister, D. H. 1984. Profitability: Beating the downward trend. The American Lawyer (July/August).

Maister, D. H. 1986. The 3 E's of professional life. Journal of Management Consulting. Volume III: No. 3.

McKee, T. E. 1986. Experts systems: The final frontier. CPA Journal (July).

McMeans, D. 1986. U.S. international trade in services. U.S. Industrial Outlook.

Noling, M. S., and J. F. Blumenthal. 1986. Gaining competitive advantage as a professional service firm. New Management.

Professional Publications Inc. 1986. Study reported in Journal of Accountancy (October). P. 32.

Roach, S. 1988. Technology and the services sector: America's hidden competitive challenge. In Technology in Services: Policies for Growth, Trade, and Employment, B. R. Guile and J. B. Quinn, eds. Washington, D.C.: National Academy Press.

Spindler, P. 1986. Professional service firms face new challenges. New Management.

Wiseman, C. 1985. Strategy and Computers. Dow Jones-Irwin. P. 67.

Zuboff, S. 1985. Automate/informate: The two faces of intelligent technology. Organizational Dynamics (Autumn).

Advisory Committee on Technology in Services Industries

Chairman

James Brian Quinn, William and Josephine Buchanan Professor of Management, Amos Tuck School of Business Administration, Dartmouth College

Members

Donald N. Frey, Professor of Industrial Engineering and Management Sciences, Northwestern University, and former Chairman of the Board, Bell & Howell Company

Paul F. Glaser, Chairman, Corporate Technology Committee, Citicorp

Thomas R. Kuesel, Chairman of the Board, Parsons, Brinkerhoff, Quade & Douglas, Inc.

Frederick W. Smith, Chairman and Chief Executive Officer, Federal Express Corporation

Eric E. Sumner, Vice President, Operations Systems and Network Planning, AT&T Bell Laboratories

Contributors

JOHN H. DAVIS is executive director of the AT&T Consumer Products Laboratories and is responsible for product development of all of AT&T's telephone products. During 1982–1985 he was responsible for AT&T Bell Laboratories' development of cellular technology. He was educated as an electrical engineer, receiving his bachelor's, master's, and doctorate degrees from Georgia Institute of Technology, Massachusetts Institute of Technology, and the University of Pennsylvania, respectively.

THOMAS H. DOORLEY is managing partner and founder of Braxton Associates. The thrust of his work has been toward strategy management, that is, working to achieve integration between strategy definition and effective implementation. He has led the extension of Braxton's strategy management concepts and client base into consumer and service organizations and leads Braxton's work in conjunction with Touche Ross for retail companies. Prior to founding Braxton, he was a senior consultant on business strategy and organizational issues and a business unit manager at Arthur D. Little, Inc.

FREDERICK A. (TED) FELLOWES is director of business development for the Publication Systems Division of Bell & Howell. Since earning his M.B.A. at the University of Chicago, he has devoted nine years to various new product (or services) programs at Bell & Howell.

DONALD N. FREY is a professor in the Department of Industrial Engineering and Management Sciences at the Technological Institute of Northwestern University. He is the former chairman of the board, chief executive officer, and president of Bell & Howell Company.

CHRISTOPHER GAGNON is currently an Associate with McKinsey & Company in Los Angeles. He contributed to this volume while completing his M.B.A. at Dartmouth's Amos Tuck School of Business Administration.

PAUL F. GLASER is chairman of Citicorp's Corporate Technology Committee and a member of Citibank's Policy Committee. He is responsible for establishing technology policy and standards, for the introduction of new technologies, and for evaluating the quality and direction of systems efforts within Citicorp. From 1973 to 1984, he was chairman of Citibank's wholly owned subsidiary, Transaction Technology, Inc., and was instrumental in the design, development, and implementation of Citibank's automated banking centers. Prior to joining Citibank, he was with TRW, Inc. for 15 years. Mr. Glaser holds bachelor's and master's degrees in electrical engineering from New York University.

ALISON GREGG is a Senior Associate with Braxton Associates, spending her time in both Boston and London. She is a graduate of Dartmouth's Amos Tuck School of Business Administration and of Williams College.

ALLAN GRODY is president of Financial Intergroup, a consulting and joint venture marketing, development, and management company. Prior to founding Financial Intergroup, he was a partner in the Financial Services Consulting Division of Coopers & Lybrand where he was national director of the Investment Industry Consulting practice. Among his earlier assignments he worked in banking (Algemene Bank, N.V.), investment management (Neuberger & Berman), securities (Reynolds & Co.), and commercial and consumer finance (General Electric Credit Corp.). He has a B.A. degree in mathematics from City University of New York.

BRUCE R. GUILE is associate director of the National Academy of Engineering Program Office. Prior to joining the Academy in 1984, Dr. Guile worked as research associate with the Berkeley Roundtable on the International Economy and as a management consultant. Dr. Guile holds a bachelor's degree in English literature and computer science from Heidelberg College, a master's of public policy from the University of Michigan, and a Ph.D. from the University of California, Berkeley.

CHRISTOPHER KEITH is senior technology officer, senior vice president, and member of the Management Committee at the New York Stock Exchange. Previously he served as senior vice president at SIAC, where he was responsible for Development and Planning, and in that capacity was responsible for the development of CMS, DOT, ABS, ITS, the consolidated transaction and quote services, and OPRA among others. Earlier in his career he was founder and general manager of Data Network Services, ADP's financial communication service arm, and vice president of Systems and Operations at Market Research Corporation of America. He has a B.A. in psychology, did graduate work in physics at Massachusetts Institute of Technology, and has served on various industry boards.

RICHARD C. LARSON is professor of electrical engineering and computer science at the Massachusetts Institute of Technology. He served for nine years as codirector of M.I.T.'s Operations Research Center. His research has focused on developing applied operations research models of services organizations in both the public and private sector. He is president and cofounder of a Cambridge-based firm, ENFORTH Corporation.

JUAN A. MURILLO is currently vice president and director of concrete bridge technology with Parsons, Brinkerhoff, Quade & Douglas, Inc. in Tampa, Florida. He has been involved in bridge design and construction for the last 20 years and since 1974 in the introduction of new developments in concrete technology for bridge structures in the United States and abroad.

CARL NEHLS is a project engineer at Federal Express where he has played a lead role in the development of the COSMOS IIB custodial package tracking system. Carl holds a Bachelor of Science in mechanical engineering from Case Western Reserve University and a Master of Science in industrial administration from Carnegie Mellon University.

JAMES BRIAN QUINN is the William and Josephine Buchanan Professor of Management at the Amos Tuck School of Business Administration at Dartmouth College. Dr. Quinn earned his B.S. from Yale, an M.B.A. from Harvard, and a Ph.D. from Columbia University and joined the Tuck faculty in 1957. Professor Quinn is an authority in the fields of strategic planning, the management of technological change, and entrepreneurial innovations. He has held fellowships from the Sloan Foundation, the Ford Foundation, and the Fulbright Exchange Program. In addition to consulting with leading U.S. and foreign companies and publishing extensively on corporate policy issues, Dr. Quinn has the distinction of being the dean of a Japanese business school.

Index

J

N

O