NISTIR 6079

Advanced Technology Program
Motor Vehicle Manufacturing Technology
Public Workshop

J. C. Boudreaux
T. R. Lettieri

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Advanced Technology Program
Gaithersburg, MD 20899-0001

NIST

Advanced Technology Program
Motor Vehicle Manufacturing Technology
Public Workshop

J. C. Boudreaux
T. R. Lettieri

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Advanced Technology Program
Gaithersburg, MD 20899-0001

October 1997

Preface

The NIST Advanced Technology Program (ATP) provides competitive, cost-shared awards for industry to develop high-risk, enabling technologies with broad-based economic benefit. The ATP seeks to help industry fill the gap between basic research and product development, and to invest in technology that wouldn't be developed in a competitive time-frame without government cost-sharing. Along with general competitions which are open to proposals from all technical areas, the ATP is also funding a series of "focused programs" with specific business and technical goals.

Begun in 1995, the ATP Motor Vehicle Manufacturing Technology (MVMT) program focuses on technology advances that strengthen manufacturing capabilities along the entire automotive production chain. Extensive, and continuing, contacts with industry have made it clear that the major business challenge facing the U.S. light vehicle industry is the ability to respond to highly volatile and fragmented market demand by rapidly launching new product models in a manner that is profitable in low volumes.

Fifty percent of the value in American vehicles is added by suppliers, and that percentage is rising. MVMT concentrates on developing the technical bases of specific manufacturing processes, thereby leveraging American leadership in engineering and the physical sciences into a position of leadership in manufacturing; and on making processes reconfigurable and easily coupled with other processes in a complete product manufacturing system, thereby providing greater flexibility and higher utilization of plant capacity. This will in turn provide designers with new technical options and an enlarged, but predictable, process horizon which will enable innovative product designs, and encourage the effective re-utilization of capital equipment to produce a diverse family of high-quality products, and the rapid and accurate translation of designs into production.

Two MVMT competitions have been held: the first in 1995, and the second in 1997. The purpose of this workshop is to provide the manufacturing community with a technical overview of the progress of the projects awarded in the first competition and several closely related ATP projects, and to introduce the projects selected in the second competition. While the creation and protection of intellectual property is a core value of the Advanced Technology Program, broad-based economic benefits and spill overs can be obtained only if information about the technology is provided to those members of the technical community who could potentially benefit from having it.

The motor vehicle manufacturing sector has long been recognized as a prime mover in the development of manufacturing processes. The same technology is applicable to many other industrial sectors as well. For example, improvements in machining, tooling, and assembly are likely to directly impact the metal furniture and fixtures, primary metals,

fabricated metal products, electrical and non-electrical machinery, transportation, and precision instrument industries. Hence, the technology developed by this program has the potential for broad diffusion throughout American manufacturing, with ensuing wide-spread benefits. This public workshop is intended to be an important step in this technology diffusion effort.

J.C. Boudreaux
T.R. Lettieri

Contents

Neural Network Modeling of Abrasive Flow Machining
A.E. Smith and W.S. Slaughter, University of Pittsburgh

Computerized Modeling of Electrodeposition Between Conducting Boundaries Having Variable Geometries
J.R. Bush and M. El-Shazly, Abrasive Technology Aerospace

Experiences in the Grinding of Silicon Nitride on a Lower Cost High Speed CNC Grinder
T.W. Hwang, E.P. Whittenton, and C.J. Evans, NIST Manufacturing Engineering Laboratory

Nitrides in Materials Science: Some Thermodynamic Aspects
P.A.G. O'Hare, NIST Chemical Science and Technology Laboratory

Session V. Stamping

An Overview of the Agile and Precision Sheet Metal Stamping ("Near Zero Stamping") Program
E.V. Vahala, Auto Body Consortium; X. Wu and J. Shi, University of Michigan

The Status of the Springback Predictability ATP Project
D. VandenBossche, Chrysler Advanced Manufacturing

Internal NIST Research Complementing the "Springback Predictability" and "Agile Precision Sheet Metal Stamping" Projects
L.E. Levine, R. Thomson, G.G. Long, D.R. Black, and R.J. Fields, NIST Materials Science and Engineering Laboratory

Session VI. Surface Treatment and Material Processing

Cubic Boron Nitride Coatings for Cutting and Specialty Tools
A. Inspektor, A.S. Gates, U. Schleinkofer, and E. Oles, Kennametal; T. Lefkow, Northwestern University; W. Walch and R.L. Resnick, Extrude Hone

Industrial Process Driven System Requirements for PSII Applications
C.P. Munson, Los Alamos National Laboratory

Plasma-Based Ion Implantation: Surface Treatments for Automobile Manufacturing
J.V. Mantese, GM Research Laboratories

The Commercialization of PSII into Multiple-User Markets
M.C. Dudzik, Environmental Research Institute of Michigan

Session I. Introduction

The Motor Vehicle Manufacturing Technology (MVMT) Program: An Overview
J.C. Boudreaux, M.M. Daum, and T.R. Lettieri, NIST ATP

Lessons Learned from 2mm Program
F.E. Plonka and M.S. Ahmed, Wayne State University; D. Ceglarek and J. Hu, University of Michigan; B.G. Fromm, Progressive Tools & Industry Co.

The Motor Vehicle Manufacturing Technology (MVMT) Program: An Overview

J.C. Boudreaux
M.M. Daum
T.R. Lettieri
NIST/Advanced Technology Program

Introduction. The NIST Advanced Technology Program (ATP) provides competitive, cost-shared awards for industry to develop high-risk, enabling technologies with broad-based economic benefit. While the government provides the catalyst--and in many cases critical technical support--industry conceives, manages, and executes ATP projects. The ATP seeks to help industry fill the gap between basic research and product development, and to invest in high-risk technology that would not be developed in a competitive time-frame without government cost-sharing. Along with general competitions which are open to proposals from all technical areas, the ATP added in 1994 a new element to its investment strategy, namely, focused programs. Each focused program has well-defined research and business goals. Focused programs are not open-ended. A typical program runs about five years with an end-date defined up front, and involves $10 to $50 million per year in ATP funding and a comparable amount of private sector cost-sharing. ATP focused programs are selected by a competitive process, based on four major criteria: the potential for U.S. economic benefit, evidence of good technical ideas, strong industry commitment, and opportunity for ATP funds to make a difference.

Begun in 1995, the ATP Motor Vehicle Manufacturing Technology (MVMT) program focuses on technology advances that strengthen manufacturing capabilities along the entire automotive production chain. Extensive, and continuing, contacts with industry have made it clear that the major business challenge facing the U.S. light vehicle industry is the ability to respond to highly volatile and fragmented market demand by rapidly launching new product models in a manner that is profitable in low volumes.

MVMT Program Goals. The U.S. light vehicle industry consists of the Big Three automakers and their first, second, and third tier suppliers. Companies that produce the machine tools, dies, jigs, fixtures, and other related metalworking equipment used in motor vehicle manufacturing are considered to be second tier suppliers. The manufacturing processes are concentrated in the supplier community. Fifty percent of the value in American vehicles is added by suppliers, and that percentage is rising. Unfortunately, lean production strategies and price-based competition make funding for research and development an almost unheard-of luxury for suppliers. This program will mine the lode of process expertise in the Big Three, the motor vehicle supplier industry, and research community to

identify and develop innovative technical solutions to the business challenge of increasing productivity and reducing time-to-market of U.S. manufacturers in the highly competitive global marketplace.

The domestic oligopoly that existed for the U.S. automobile industry up through the early 1970's has evolved into a globally competitive open market. In order for the U.S. automobile manufacturers and their suppliers to remain competitive, they need to improve their capacity utilization, improve quality, reduce operating expenditures through improvements in manufacturing technology and productivity, and reduce their overhead expenses. The current competitive climate has led the automobile manufacturers to re-define their relationship with their suppliers. In the past, the relationship could be described as arms-length, market based, short term, and adversarial. The Big Three brought in suppliers late in the design process. Competing suppliers bid on the provided blueprints. At that point, suppliers could do little to improve the design, which may have been difficult and expensive to manufacture. Contracts were awarded on an annual basis covering the model year. The Big Three are now working with their suppliers earlier in the new product development process to lower costs and increase the speed to market. The trend now is for more of the engineering and systems development work to be pushed down into a smaller, but more technically capable, supplier base with specific goals being defined for cost, quality, timing, product features, and productivity increases. The contracts now tend to be longer term, usually awarded for the life of the vehicle model.

Business Goals. *Industry goals for cost sharing this program are to (1) reduce manufacturing and capital equipment cost of introducing a new vehicle model by an order of magnitude, (2) reduce the time to market from the current U.S. industry standard of 42 to 48 months to 24 to 36 months, and (3) increase the global competitiveness of U.S. firms by strengthening their ability to team with suppliers.*

While the program focuses on motor vehicle manufacturing, the manufacturing processes and technology to be developed are applicable to many other sectors as well. Improvements in machining, tooling, and assembly are likely to directly impact the metal furniture and fixtures, primary metals, fabricated metal products, electrical and non-electrical machinery, transportation, and precision instruments industries. Hence, the technology developed by this program has the potential for broad diffusion throughout American manufacturing, with ensuing wide spread benefits.

Technical Goals. *This program will develop (1) specific technical improvements in manufacturing processes and process monitoring and control, (2) flexible, reconfigurable equipment to produce diverse product families, and (3) agile manufacturing systems to permit rapid, low-cost product conversion and efficient equipment re-utilization.*

Providing designers with new technical options and an enlarged, but predictable, process horizon will enable innovative product designs. Agile manufacturing systems encourage the effective re-utilization of capital equipment to produce a diverse family of high-quality products, and the rapid and accurate translation of designs into production.

Economic Impact. The success of the Motor Vehicle Manufacturing Technology focused program will significantly improve the global competitiveness of the U.S. motor vehicle industry. The automotive manufacturing industry forms the core of the nation's industrial strength. In a typical year, the industry generates one-sixth of all U.S. manufacturers' shipments of durable goods and consumes 30 percent of all the iron, 15 percent of all the steel, 25 percent of all the aluminum, and 75 percent of all the natural rubber purchased by all industries in the United States. On average, every one dollar of manufacturing input in the United States allocated to producing motor vehicles adds two and one half dollars to the economy. At the retail level in 1995, sales of motor vehicles exceeded $259 billion, 3.6 percent of the nation's gross domestic product--the broadest measure of the nation's economic output.

The Commerce Department's Bureau of Economic Analysis (BEA) reports that in 1995, the automotive industry (Standard Industrial Classification 371 "Motor Vehicles and Equipment") accounted for 9 percent of all the private industrial employment provided by all manufacturers of durable goods in the United States. BEA data shows that industry firms as defined by SIC 371 directly employed 899,000 American workers in 1995. Employees in this SIC code earned compensation totaling $60 billion--equal to 12 percent of the total paid by all manufacturers of durable goods.

When all the establishments that make up the motor vehicle industry are added to establishments of all related industries, it becomes clear how pervasive the economic impact of this focused program will be on the U.S. economy. The American Automobile Manufacturers Association estimates that in 1992, there were a combined 589,000 manufacturing and service sector establishments within the U.S. that derive their business directly or indirectly to motor vehicles. Overall, these establishments provided jobs for an estimated 6.8 million workers--more than 7 percent of all U.S. private non-agricultural employment. Their workers earned a payroll worth $170 billion--7.5 percent of the nation's total.

The motor vehicle industry is a mature, cyclical industry. Annual growth for new passenger cars and light trucks in the U.S. on a long-term basis is predicted to be just over 1 percent. The number of motor vehicle models has increased over the years, with the average volume per model decreasing. The resulting trend is towards a more fragmented, niche-oriented market. A key element in gaining market share in such an environment is a manufacturer's ability to introduce new products to respond rapidly to changing consumer demand.

In the recent past, product development at the Big Three could be characterized as a lengthy, throw-it-over-the-wall process. The engineering and development cycle was, and still is, capital intensive and thus very expensive, due mainly to the need not only to redesign dedicated machining and tooling for each new model, but also to prototype and try out the redesigned equipment. According to a 1991 report on the U.S. motor vehicle industry and market by the Volpe National Transportation Systems Center, a "major" model changeover (20 percent change in content) costs about $3 billion and takes between two and four years to bring to market. Approximately half of the $3 billion ($1.2 - $1.8 billion) is related to manufacturing and capital equipment costs. An "all new" model changeover (50 to 70

percent content change) costs $6 to $9 billion, takes between three to five years to bring to market, and the changeover cost for converting the manufacturing facilities is an additional $1.5 - $2.9 billion. If these trends were to continue, then the model changeovers forecast for the turn of the century, with 70 to 90 percent changes in content, would cost $9 to $12 billion and take an additional two to three years to bring to market. Under these conditions, U.S. companies could not respond to changing consumer demands in a cost effective and timely manner. But assuming that reducing the equipment cost by an order of magnitude through the successful MVMT program would significantly lower changeover cost, a "major" changeover would cost $1.6 - $1.7 billion, an "all new" changeover would cost $3.2 - $6.3 billion, and forecasted turn-of-the-century changeovers would cost $5 - $7 billion. Considering that the Big Three are now involved in about 10 changeovers per year, and that this number is likely to grow due to the increased market fragmentation, the potential net savings are huge.

Reducing the time from vehicle launch to the first "all new" marketable vehicle from current industry standards to 24 months will provide the Big Three the flexibility to develop products to better respond to changing consumer demands. This time reduction will result in part from the closer relationships between the Big Three and their suppliers, the use of concurrent engineering, and designing vehicles for manufacturability. The manufacturing processes and flexible tooling technology proposed in this program will enable existing equipment to be modified quickly and with low incremental capital investment, allowing the Big Three and their suppliers to meet customer demand in a timely, cost-effective manner. The reduced cost of introducing a new product will lower the accounting break-even point, implying that smaller market segments could be profitably served.

This focused program will have an immediate and profound impact on U.S. metalworking industries. According to the U.S. Statistical Abstract 1995, 1994 U.S. consumption of metal cutting and metal forming tools, and special dies, tools, jigs and fixtures can be estimated to be about $19.3 billion. Motor vehicle or motor vehicle related industries typically account for between one third and one half of all machine tool shipments annually. To remain competitive, the first and second tier suppliers will need new metal cutting and metal forming tools, welding equipment, and special dies, tools, jigs, and fixtures that will meet their performance and productivity requirements. As the Big Three are competing globally, they will increasingly source globally for machine tools and other supplies. If the U.S. first and second tier suppliers are unable to provide components with suitable performance-to-price ratios, then the Big Three will look to foreign suppliers.

Evidence of Good Technical Ideas
The MVMT program focuses on manufacturing processes as the means to drive the competitiveness of American firms. Designed on the basis of industry input, the program targets technology advances that can strengthen manufacturing capabilities along the entire automotive production chain. The program will advance manufacturing processes in two ways. First, by concentrating on technical bases of specific manufacturing processes, it will leverage American leadership in engineering and the physical sciences into a position of leadership in manufacturing. Second, by making processes reconfigurable and easily coupled with other processes in a complete product manufacturing system, it will provide greater

flexibility and higher utilization of plant capacity. The technologies outlined below are clear areas of need that are recognized by the automotive industry and fall within the scope of the MVMT program.

Material Processes and Equipment

(A) *Stamping and metal forming processes.* Projects selected in the first MVMT solicitation targeted the precision and agility of stamped sheet metal parts on the scale of body components. The second MVMT solicitation will extend these advances to small precision parts, such as springs and fasteners. These industries, which consist of thousands of small family-operated businesses across North America, faces the same basic issues as the manufacturers in the higher profile industries: the need to increase productivity and quality, and to lower changeover times and overall costs. The equipment that produces these parts uses fundamental technology developed in the first half of this century. This technology will have to be enhanced significantly to meet future competitive global challenges in this industry. Targets of opportunity include technologies for imbedding sensors in tools and processes, increasing productivity (by an order of magnitude) of the more flexible CNC bending and forming technologies, closed-loop real-time monitoring and control systems for in-process self-corrective action, increasing productivity (by an order of magnitude) of press and slide forming processes, and the development of material handling systems to support these process improvements.

(B) *Advanced Machining.* The first MVMT solicitation targeted the machining of discrete prismatic parts, including such powertrain components as engine blocks, heads, transmission cases, and crankshafts. Newer automobile and aircraft engines, diesel engines, and hundreds of other products require exceedingly high tolerances in order to function properly. The targets of opportunity in machining for the second MVMT solicitation include reconfigurable machining systems with increased precision for producing high-volume parts. Competitiveness will be advanced by agile transfer lines using modular units that can be quickly assembled or disassembled based on market demand. These systems must support increased cutting speed and feed rates, real-time error correction capabilities, the generation and validation of cutting paths directly from computer-aided-design data, intelligent process monitoring and control capabilities, and responsiveness to changes measured *in situ* during the machining process.

(C) *Abrasive machining.* Grinding is a machining process often used in the finishing of parts, and there is a well-documented need for high precision, inexpensive grinding machines. Technology developments that will enable advanced grinding machines include advances in grinding tool materials, such as superabrasives, and improved understanding of the fundamentals of the grinding process. For example, in the centerless grinding of cylindrical automotive components, superabrasive technology is not cost-effective due to a lack of machine tool technology, particularly in emerging engine component technology using advanced ceramics such as cam followers and values. Previous research on the grinding of ceramics indicates that to achieve large removal rates while maintaining a small grain depth of cut, high wheel speeds and fine grit wheels must be employed. The targets of opportunity include process developments for superabrasives, machine tool development for increasing

static and dynamic rigidity, and control technology for greater precision in form and surface finish.

(D) *Constructive technologies for rapid fabrication of production tooling and functional parts.* In the automotive industry the longest lead time in producing a new product is the design and fabrication of production tooling, including such items as molds for plastic parts, dies for die casting, and stamping dies. The traditional approach is machining tool steel, a time-consuming process. Several methods have been developed for the rapid fabrication of prototype or limited-run tooling, but the tooling produced is typically not amenable to production molding or casting processes. In addition, the tool materials are usually unable to address demanding metal forming processes, such as forging, stamping, or casting, because of their limited temperature capability and hardness. Emerging constructive technologies that build up the desired shape rather than cutting it out of a blank offer a potentially revolutionary approach, not only for fixtures and tooling, but also for parts with features and geometries that cannot be obtained with conventional metal removal processes. These fabrication technologies permit miniaturized sensors and actuators to be embedded in both tooling and products, and thus can empower designers to create new designs not previously obtainable. The targets of opportunity include metal spraying, investment casting using rapid prototype models as patterns, vapor plating, direct metal deposition, three dimensional printing, droplet based manufacturing, free form fabrication, and free form powder molding.

(E) *Net shape forming of advanced materials.* Lighter weight advanced engineered materials, originally developed for the defense and aerospace industries, can reduce emissions and fuel consumption for the North American ground transportation fleet without significantly compromising vehicle package and safety. These engineered materials, primarily consisting of a matrix (polymer, metals, or ceramics) and synthetic fibers (glass, polymers, or ceramics), are now too expensive for adoption and widespread use in the automotive industry. The adoption of these materials depends on the development of manufacturing technologies which exploit their near-net-shape fabrication capabilities. Many alternative technologies for these advanced materials, such as vacuum die casting, semi-solid forging, precision forging, squeeze casting, metal injection, ceramic injection, plastic injection, reactive molding, and powder metal processing, need to be explored. The targets of opportunity in this area include dimensional repeatability, reduction in physical variation through real-time sensing and control, interface chemistry control, and (more generally) increased process reliability.

Assembly Processes and Equipment

This area has the overall objective of developing modular, rapidly deployable, flexible assembly systems capable of economically assembling any member of a product family in any desired quantity, and capable of being disassembled and redeployed to assemble members of another product family. This need will be addressed by developing standard, modular, and flexible workcells. Modular hardware will allow a better than 90 percent reutilization of system components in newly configured systems, resulting in a better than 90 percent recovery of the capital invested. Modularity will also facilitate assembly system configuration by allowing software to play a significant role in reducing the system

deployment time. These design modules will allow future assembly systems to be deployed in a matter of 4 to 6 months, a considerable improvement over the current 24 to 36 months.

(F) *Powertrain Assembly.* Powertrain assembly systems are responsible for all components of powertrains, including engine heads, short blocks, complete engines, transmissions, and any other powertrain subassembly. These components are characterized by edge dimensions less than one meter, and weights not exceeding 230 kilograms (about 500 pounds). Some components, such as gaskets, are deformable, requiring special handling. Assembly systems must have the capability to be both high-volume production units (capable of rates up to 40,000 units per month) and also agile and rapidly reconfigurable (having fast manipulators with lift capacities up to 230 kilograms). They also must incorporate dedicated part feeding systems, rapid sensing capabilities, and system performance monitors. Targets of opportunity include systems equipped with: (1) flexible feeders, (2) tool magazines for end effectors capable of driving screw and bolts, installing interference and snap rings, and delivering adhesives, and (3) 2D, or modestly priced 3D, vision systems.

(G) *General Assembly.* General assembly in automotive manufacturing plants comprises the entire assembly process after parts leave the paint area and up to the time the finished product leaves the plant for shipment to the dealer. This area includes installation of interior and exterior trim, instrument panels, seats, the powertrain assembly, steering assemblies, brakes, electrical, suspension, and, in the case of trucks, frame assemblies. The success of cost-effective and quality assembly is determined by the teams that design, build, and supply subassemblies to the general assembly area. Because component assembly is so diverse, general assembly has been left largely untouched by technological productivity improvements. Success in this area depends on solutions that impact suppliers of subassemblies, suppliers of production equipment, and the 64 North American general assembly plants of the automakers. Critical areas for technology development and deployment include material handling to and from the assembly line extending across the supply chain, new joining technologies to reduce the number of discrete fasteners in joining dissimilar materials, and inspection technologies to validate assembly processes and ensure the integrity of components both before and after assembly.

Information and Knowledge Processing

(H) *Intelligent process monitoring and control.* Today most manufacturing processes run open-loop, with only manual adjustment. Quality and productivity enhancements will require active control systems to reduce process variation. Such systems depend on improvements in sensors and process models. Closing the control loop requires improved sensor technology, including the ability to embed sensors directly in tooling, algorithmic and hardware support for smart sensors that can derive parameters of interest (e.g., injection molding pressure) from directly observable quantities, and the ability to fuse information from different sensory modalities (e.g., tactile and acoustic information in assembly). Interpreting this sensory feedback requires formal models of the process to predict behavior and to detect and remedy malfunctions. Efficient derivation and management of process models is a non-trivial problem, and will require innovative research and development in emergent control, including application of object- and agent-oriented software technologies to real-time

problems, genetic techniques that evolve control strategies, and applications of complex dynamics to take advantage of the enhanced adaptivity offered by systems on the edge of formal chaos. The increased process information available in a closed-loop environment raises the stakes for overall system integration, making possible much tighter coupling of different processes but also placing new demands on integrators. Open systems technologies must be extended and refined to ensure plug-and-play compatibility, rapid reconfigurability, and close teamwork between machines and human workers.

(I) *Integration of Product and Process Information.* Direct use of product data in production, and feedback of process information to process designers, can reduce the lead time and improve the accuracy of process tooling. Several manufacturing processes are tightly coupled to the product design. Examples include stamping dies, molds, assembly fixtures, inspection gages, packaging, material handling and other interfaces. Process designs (like stamping dies) are done at organizations that are suppliers to the product manufacturer, requiring data models to flow down the supply chain. Three hurdles require expensive manual intervention. The product model may be incomplete with respect to the details needed to generate tooling, requiring additional information. There may be errors in the product model (such as sliver surfaces or line fragments) that do not show up in drawings but clog CAM software. The accuracy of process design systems may be limited and not compensate for actual production results, such as springback or shrinkage. Process information needs to be captured in a way that not only supports continuous improvement within a single process, but also closes the process loop with upstream and downstream processes. For example, an assembly operation should be designed on the basis of dimensional information from the process that manufactured the parts to be assembled, rather than relying only on product models of those parts that do not take account of manufacturing variability.

Industry Commitment. The U.S. light vehicle manufacturers and their suppliers were heavily involved in the initial development of the MVMT focussed program and in the modifications made to the original program white paper's technical thrust areas. Due to light vehicle suppliers growing importance in vehicle design and manufacturing, suppliers are the program's chief focus and their participation in creating the MVMT focussed program was deemed essential for the program to be successful. Today, these firms account for about half of the value added in light vehicles. That proportion is expected to grow as automobile manufacturers assign an increasing share of engineering and development work to suppliers and look outside for components that they once made themselves. The 17 white paper submissions that formed the basis for the first solicitation of the MVMT focussed program were generated by more than 150 firms, representing manufacturing capabilities along the entire automotive production chain. A program planning workshop held in Ypsilanti, Michigan in October 1994 allowed the over 250 industry and academic participants to provide comments on the draft MVMT program white paper. During the winter of 1994/1995, regional bidders conferences were held in Ypsilanti, MI; Champaign-Urbana, IL; Pittsburgh, PA; Kansas City, Kansas; Alberquerque, NM; and in Gaithersburg, MD.

Overall, a total of 199 companies were involved in the 61 proposals industry submitted to the first MVMT competition. Of the 61 proposals, 15 received ATP awards. The 15 awards

include 93 total participants, with 31 of the awardees being small businesses. Particularly for smaller automotive industry suppliers, the time and cost of having their key personnel participate in writing a proposal for a competitively awarded program, given the political uncertainty surrounding the ATP in the winter and spring of 1995, provides strong evidence of the industries commitment to this program. Abstracts of the awarded proposals are provided as an appendix to this document.

In 1992 the United States Council for Automotive Research (USCAR) was formed by Chrysler, Ford, and General Motors, following cooperative research and development that began in 1988. The mission of USCAR is to facilitate, monitor and promote precompetitive cooperative research and development. On September 29, 1993, President Clinton and Vice President Gore joined with the chief executive officers of the Big Three U.S. automakers to announced the formation of a new partnership aimed at strengthening U.S. competitiveness by developing new technologies for motor vehicles. This Partnership for a New Generation of Vehicles (PNGV) identified three specific but interrelated goals: (1) significantly improve national competitiveness in manufacturing, (2) implement commercially viable innovation from ongoing research on conventional vehicles, and (3) develop a vehicle to achieve up to 3 times the fuel efficiency of today's comparable vehicle. This ATP focused program is most closely aligned with the first goal.

Significance of ATP Funding. Even though the motor vehicle market has been strong in the past few years, it is highly unlikely that without ATP support the Big Three and their first-tier suppliers would undertake a motor vehicle technology program even beginning to approach the scope and emphasis being proposed here. The high technical risks of this program would not be matched by suitable near-term benefits. The industry has adopted a product-oriented outlook and has concentrated its research and development resources in areas that consumers can clearly identify and use for product differentiation. The fraction allocated to process-oriented R&D tends to focus on shorter term, incremental improvements, in contrast with the major gains in performance and capabilities that the ATP focused program will foster. More fundamentally, the Big Three are served by essentially the same supplier base. Numbering about 3,500 companies, U.S. automotive suppliers tend to be small and medium-sized firms. Most spend little or nothing on process-oriented research, leaving them ill-prepared to anticipate and respond to major shifts in manufacturing technology and automobile concepts. An example of such a turning point is an end-of-the-decade transition to lightweight aluminum components for most body and powertrain parts now made with cast iron. If domestic suppliers are slow to respond to this transition, U.S. auto makers will be forced to look abroad to meet their needs for machine tools and parts. Furthermore, from a very practical standpoint, if any one of the Big Three funds a supplier to develop process-oriented technological improvements, many competing motor vehicle manufacturers would also benefit as "free riders". The funding company would not be able to gain a competitive advantage from their investment, and subsequently not receive an adequate return for the high technical risk nature of program.

Beyond the major technological advances that it will spur, the new program is expected to foster a more cooperative and more constructive relationship between auto manufacturers and

their suppliers, resulting in additional competitive advantages. This benefit is already obvious from previous ATP funded projects involving this sector and industry participation in proposing this new focused program.

In order to ensure that the first MVMT solicitation did not conflict with other programs, discussions were held with representatives of several Federal agencies, including the National Automotive Center of the   U.S. Army Tank and Automotive Command, the National Science Foundation, the Department of Energy Defense Programs Technology Transfer, and the Advanced Projects Research Agency. Two conclusions were clear at that time and have remained so. First, as the PNGV canvass correctly predicted, there is no indication of either direct conflict or redundancy. Second, an inventory of manufacturing-related programs, conducted for the PNGV, revealed only a small collection of activities devoted to factory-floor technologies, despite their well-recognized importance to accomplishing PNGV goals.

Having been briefed on the scope of the first and second MVMT program, other federal agencies remain strongly supportive, viewing it as augmenting their own initiatives. The focus of MVMT is the development of specific manufacturing *process* technology, whereas almost all of the other Federal programs concentrate on product design and high-level systems and enterprise integration.

Without the collaborative efforts that the ATP aims to marshal, U.S. auto makers and their suppliers would not mount and sustain the range of activities needed to achieve the major advances in technology, manufacturing practices, and industry performance that are the objectives of the new program.

Lessons Learned from 2mm Program

Frank Plonka
Muhammad S. Ahmed
Wayne State University

Darek Ceglarek
Jack Hu
University of Michigan

Brian Fromm
Progressive Tools & Industry Co.

Abstract:
This paper highlights the knowledge gained from the Auto Body Consortium's 2 mm program. Funding from a National Institute of Standards and Technology's Advanced Technology Program supported this program. The paper describes lessons learned in the areas of, collaborative research, variation reduction methodology, characterizing sources of variation, development of enabling technology, and capturing and communicating knowledge. The collaboration that exists between government agencies, industry and education institution is the key to the success of the 2mm program. The methodologies for reducing variation and the enabling technology were not only developed but were also applied to the participating plants. Capturing and communicating knowledge helped in communicating with the partner companies and developing a production system. This production system is the key to communicate knowledge and lesson learned throughout the production process. The major achievements of the 2mm program can be listed as: (1) development of a methodology for body-in-white dimensional variation implemented as the "Process Navigator", a process control and a diagnostics software; (2) obtained for the first time in the USA a world class 2mm. level of body-in-white dimensional variation; (3) developed a methodology for fixture design for non-rigid parts; (4) compiled a projects analysis of major dimensional variation contributors in the stamping and assembly process; (5) developed in line measurement systems calibration methodology..

1. 0. Introduction

This paper communicates the knowledge gained from the Auto Body Consortium's 2 mm program sponsored by the National Institute of Standards and Technology's, (NIST), Advanced Technology Program (ATP). The 2mm program was a three-way partnership among industry, universities, and government. Industry partners include the Chrysler Corporation, General Motors Corporation and the Auto Body Consortium (ABC), which consists of the following companies at that time: APX International, ASC Inc., Classic Design Inc., Detroit Center Tool, Inc., ISI Automation Products Group, Inc., Modern Engineering, Perceptron, Inc., and Progressive Tool & Industries Co. Research was provided by The University of Michigan and Wayne State University.

The overall objective of the 2 mm Program was:

> *To investigate and develop methods and products that will assist the domestic automobile manufacturers in designing and producing bodies to achieve world-class body-in-white dimensional quality of 2mm, variation.*

The 2mm program was divided into the following research areas and projects:
I: Dimensional Measurement Technology
1. Computer aided design and setup for in-line optical coordinate measuring machines
2. Visibility analysis and sequencing simulation for tooling certification
3. Measurement, modeling, and real-time NC path generation for free form surfaces

II: Process Control Methodology
4. On-site measurement and process monitoring for stamping
5. Variability characterization and tolerance budget analysis for body manufacturing
6. Process navigators for automobile body assembly

III: Technology Base for Future Body Assembly Systems
7. Information feedback for tooling and process design
8. Optimal non-rigid sheet metal part holding
9. Robust design of work-holding fixtures
10. Optimization in multiple panel fitting

IV: Technology Transfer Models
11. Decision support, planning and analysis for adoption of advanced manufacturing technology.

During the three years of the 2mm program domestic automobile manufacturers saw a significant impact on vehicle quality and customer satisfaction. This paper will highlight the efforts of the 2mm program in the form of lessons learned. Figure 1 indicates the five areas of lessons learned in the 2mm program that resulted from the research.
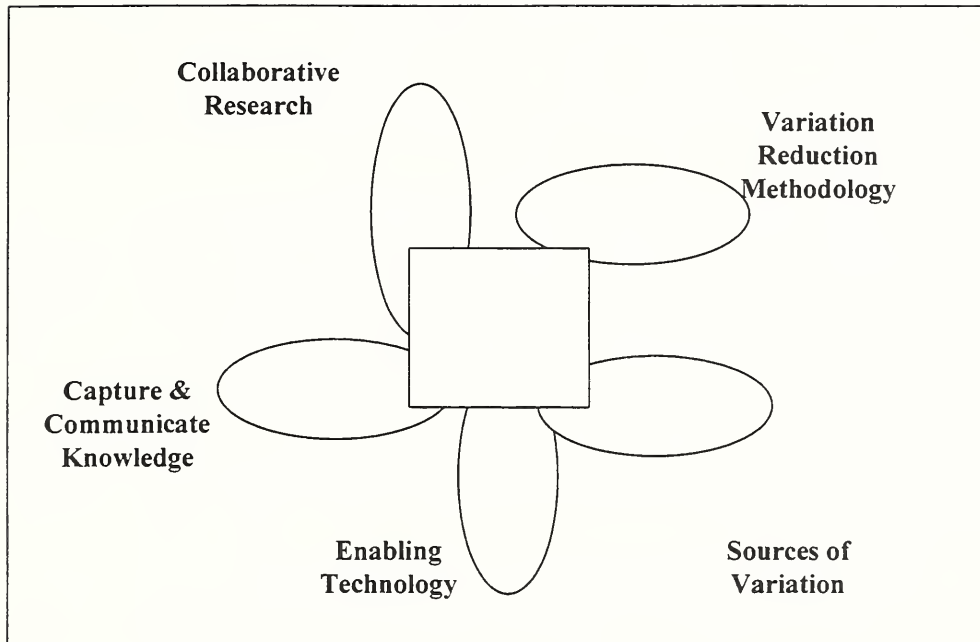
**Figure 1. Knowledge Captured and Lessons Learned
from the 2mm Program**

These areas include:

Collaboration between government, industry and university partners.
Methodologies to reduce automobile body dimensional variation to sub 2mm level.
Characterized the sources of variation.
Enabling technology developed in the program that facilitates in the reduction of
variation.
Capture & transfer of knowledge, from the partner companies, about variation through
the development of the ABC production system.

2.0. Collaborative Research

The 2 mm Program has established a very positive image for interaction between the
government, industry, and university communities. The 2 mm Program has been featured by
both the National Science Foundation and the National Institute of Standards and Technology as
a successful model for these interactions. This collaboration funded by government has resulted
in collaboration between parties that would not normally engage in concentrated and systematic
research. This initiative encouraged collaboration between competing automobile manufacturers,
supplier companies and universities. Examples of interactions between competitors involving
OEM, suppliers and universities include:

Chrysler, General Motors and ABC representatives attended quarterly project meetings to review the research results. Progress on variation reduction was reported, as well as, root causes and plans to improve. It was a common place to see suggestions for improvement from a competing company's representative in a competitor's plant. The end result of these efforts was to raise the strategic and technological level of competition thereby strengthening both companies against other global competitors.

Suppliers also shared problems and suggestions for research in an open form with the assurance that as partner companies they would have access to any products or knowledge derived. Meetings held in the OEM plants with suppliers resulted in suggestions for improvements being passed freely between them, thus benefiting the end customer.

The projects being reported in this paper was the result of collaboration between universities that normally do not occur. Figure 2 shows the collaborative efforts between Wayne State University and The University of Michigan.

Some of the lessons learned captured from conducting collaborative research are:

    Align projects with core business goals
    Research and development projects need clear goals.
    Period of three years is a short time for research and development but too long for an industrial project.
    Establish demonstrable project milestones to confirm accomplishments and to bring new member up to speed
    Define project deliverables early
    Challenge conventional methods with next-generation technologies
    Enhance organizational learning
    Perform continual bench-marking to challenge the need for the project
    Hold quarterly review meetings to emphasize sharing of results.

2 mm Update Newsletter

Lessons Learned
from the Research

Michigan
Industrial
Exposition

ABC
Production
System

Lessons Learned
Collaboration

Five Plant
Case Study
Review

Buy-Off
Procedure

Process Navigator
Inline Measurement
Gage Calibration
Fixture for Non-rigid
Parts
Pareto Analysis for
Variation Sources

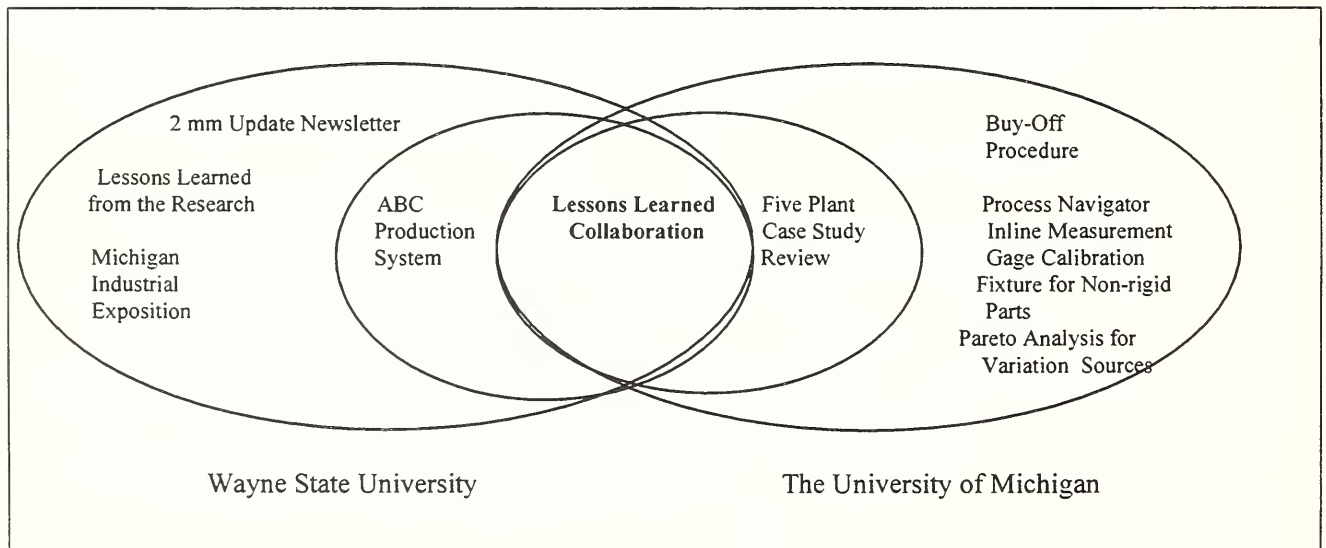Wayne State University   The University of Michigan

Figure 2. Collaborative research model between educational institutions.

Benefits of collaborative research:

> Suppliers entered into this program to gain new process knowledge.
> Some had no expectation for a product. However, a marketable product was a desirable outcome.
> Research results that reduce the length of time tools sit on the floor before the customer accepts them increase the throughput of tooling suppliers.
> Suppliers realize financial benefits by becoming more efficient and marketing their capability.
> Suppliers were challenged internally about things they took for granted, which led to changed methods.
> Next-generation technologies challenge suppliers and force them to look at better, more efficient, and more economical ways of certifying tooling to increase the accuracy and functionality of measurement tools.

3.0. Variation Reduction Methodology:

The 2mm program researchers had worked with domestic assembly plants to reduce dimensional variation with a variation reduction methodology. A systematic approach is used to identify, solve, and document occurrences of variation in the body shop. Below is a brief description of the variation reduction methodology. The 2 mm Methodology has five steps. Each occurrence of variation is documented in a case study, [Ceglarek and Shi,1995, Hu and Wu,, 1992].

1. Problem identification

Through OCMM and CMM data and control charts, process data are monitored for any process change, such as mean shifts, sporadic jumps, or variation changes. Six sigma variation is calculated from OCMM data or a valid CMM sample size for all measurement points and plotted in ascending order from left to right. The value under which 95% of the six sigma values fall is taken as the Continuous Improvement Index (CII). The CII is compared to the predetermined threshold for the plant and plotted on a CII run chart that tracks the Continuous Improvement Indices from the beginning of the variation reduction efforts. An analysis is conducted for those points above the plant threshold.

2. Data analysis

First targeting the points with the largest variation, statistical techniques, such as correlation analysis and principal component analysis, are used to determine any behavior variation patterns and whether any sensors and areas of the vehicle are moving together. The aim of this step is to group correlated points into variation problems, and thereby case studies, to narrow the focus to a particular station or assembly process area and locate the root cause.

Pareto charts are based on the frequency of problem occurrence and not necessarily on the

severity or contribution of the problem to variation. Because variation reduction efforts are data-driven and conducted so that each time a problem is identified efforts focus on major contributors to variation, case studies *do* reflect the root cause severity.

3. Action plan formulation
People with product and process knowledge localize the problem to the area in the process most likely causing the variation. Strategies are developed for short- and long-term solutions that could include visual inspection of the station, design of experiments, stamping plant data collection, or design reviews.

4. Implementation
The action plan developed in the previous step is implemented and any changes are tracked with data.

5. Evaluation
Short-term solutions are evaluated by comparing data collected before and after changes were made.

The length of time a case is open depends upon how quickly a successful solution is reached.

The project, Variation Reduction for Body-In-White Assembly, provides a software product, the Process Navigator, figure 3, which enables quality engineers and plant floor personnel to analyze the variation data from Optical Coordinate Measuring Machine, OCMM, to determine the root cause of dimensional problems and to document their findings as case studies, [Hu and Wu 1995]. The tool permits the use of advanced statistical methodologies by plant floor personnel to systematically reduce sources of variation.
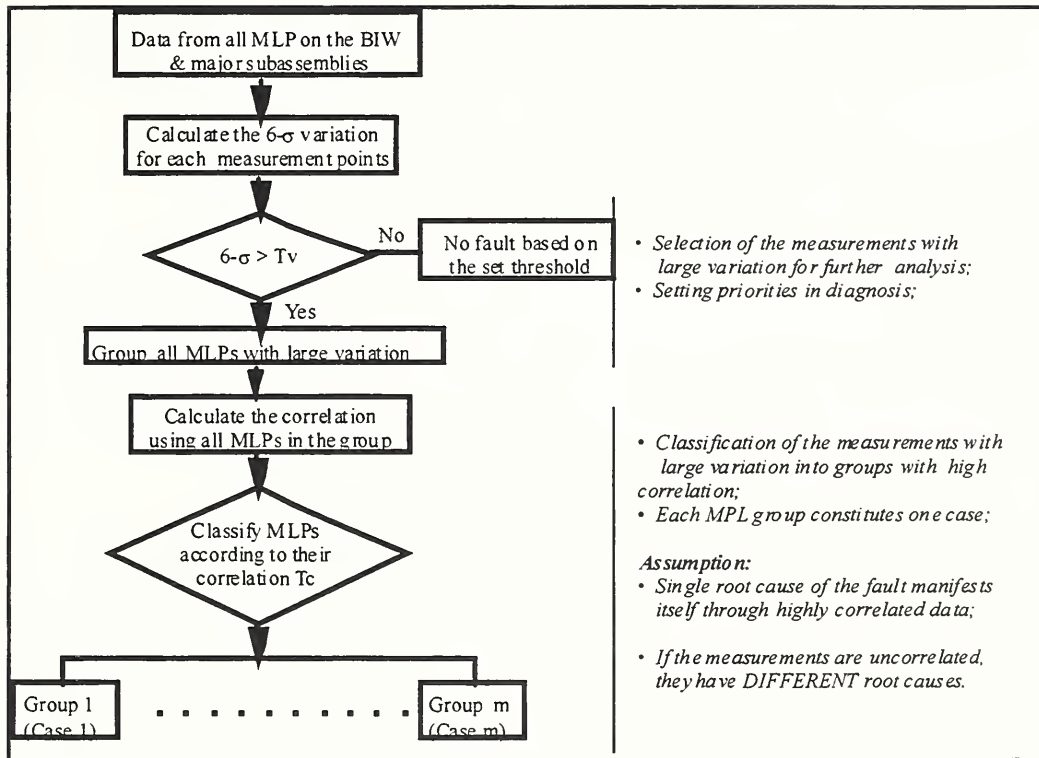
Figure 3: The flow chart of diagnostic reasoning used in the Process Navigator

4.0. Sources of Variations

A major outcome of the 2mm program research was identifying the different sources of variation inherent in the stamping and the assembly processes. This was done using two approaches:
The first conducted two research projects, which characterized sources of variation by experiments in the stamping plants and computer simulations of the assembly process, respectively. The second approach characterized the contributors to variation by analyzing case studies conducted in six assembly plants and holding lesson learned conferences, with body assembly operations personnel, to develop strategies for their elimination.

4.1.   Research Projects

The following two research projects describe the experiments and simulation studies conducted to identify sources of variation.

4.1.1. On-Site Measurement and Process Monitoring for Stamping Variation Reduction.

This project provided knowledge derived from case studies that used theory, experimentation, and simulation. Sources of variation inherent within a stamping press (i.e., within a stamping run) and the total stamping process, which includes run-to-run and material handling systems, were systematically identified. This project quantified the following effects on part variation: press maintenance condition, material handling, material properties, material thickness, and part gauging. A comprehensive on-site measurement system was developed to measure stamping press characteristics critical to dimensional variation. In addition, case studies and experiments

quantified sources of variation due to run differences, from other press lines and from material handling elements. This knowledge will be helpful in improving part designs as well as process setup and part movement.

4.1.2. Variation Simulation for Deformable Sheet Metal Assembly
This project provided knowledge regarding the key elements of variation in existing automobile body manufacturing processes (stamping, subassembly, body framing and panel hanging). The influences of product geometry, joint configurations, clamp design, and the welding process on dimensional variation for the body-in-white and major subassemblies were identified. The results show that variation is not necessarily stacked up but rather is a function of gauge, shape, process and tooling. Knowledge gained from this project helps the product and process engineer preclude variation in subassemblies at the design stage.

4.2. Analysis of Case Studies
A project entitled Information Feedback for Tooling & Process Design, organized knowledge of root causes of variation determined from analyzing 2 mm variation reduction case studies and holding meetings with assembly plant body-in-white personnel and their process, equipment, and parts suppliers. Case studies from six assembly plants were reviewed. Figure 4 represents a fishbone diagram that was developed to represent the identified factors that contribute to variation.

4.2.1. Lessons Learned Approach
Four lessons learned conferences were held at a local assembly plant. The following tooling elements were discussed, one per conference: pins, locating blocks, clamps and weld guns. A diversified team from different areas participated in these conferences. The following steps outline the approach used:

a. Initial meeting with plant
Sit down with plant management and based upon the plant's experience and dimensional variation areas of concern, select a topic area, case studies and supporting evidence (data and physical exhibits) to share with participants without product and process knowledge.

b. Call a meeting with plant and supplier personnel
Hold interactive meetings with representatives from assembly plant and upstream contributing activities concerned with quality, operations, and the manufacturing process and from suppliers to share knowledge of variation contributors and involve participation in lessons learned development and permanent preventative solutions to reduce variation.

c. Lessons learned conference
Each conference looks closely at various documented case studies at participating plants to determine root causes, undesirable effects and costs of problem, generalized fundamental principles, general fixes for the problems, and most importantly, determine ways of preventing future similar occurrences through experimentation and updating standards. The conference aim is not to point blame but rather a search for solutions and to transfer any knowledge gained in the form of lessons learned upstream to other corporate members, directly to the work force and with

suppliers.

d. Verify lessons learned

Because lessons learned resulted purely from discussions, to obtain their full benefit, validate findings through additional engineering analysis, experimentation and research to incorporate results into procedures.

e. Knowledge documentation—package lessons learned

Documented knowledge must be accessible to locate and retrieve and be understandable to individuals with different perspectives and analytical backgrounds. The final aspect in this process includes selecting and packaging lessons learned for presentation to plant maintenance personnel, corporate standards staff, platform process design engineering groups and supplier tooling designers.

4.2.2. Examples of Lessons Learned.

The project also provided a new method for evaluating the current state of the process during tooling buy-off, of which tooling is one process component. Using the tooling buy-off procedure provides direction for problem solving and feedback for continuous improvement while tooling is at supplier site. Following are some of the lessons learned:

> Local distortion of unimportant features on parts with acceptable tolerances can cause unforeseen variation in assembly due to the lack of cross sectional coordination and fit of parts.

Figure 3

Contributors to Dimensional Variation

Copyright © 1996 2mm Program Inc.,
F.E. Plonka, Wayne State University, and
The University of Michigan

Product — Incoming Material (Locate)
Transport — Carriers (Transport)
Locate — Rough Locators (Locate)
Locate — Pins (Locate)
Locate — NC Blocks (Locate)
Measurement Process (Control)
Miscellaneous (Machine & Line Configuration)
Multiple Tool (Machine & Line Configuration)
Robot Programming (Joint Assemble)
Welding (Joint Assemble)
Clamping (Locate)

Coordination of parts is complicated if produced by multiple sources resulting in a need to verify coordination procedures

Look for PLP hole damage and verify coordination of carrier rough locators to fixture PLPs

Pin to hole size coordination checks required

Verify PLP scheme to satisfy local fixture dimensional requirements

Look for pin wear

Look for NC block wear or damage in weld field

Do not adjust weld gun tip location using equalizing mechanism

Use totally enclosed clamps in weld environments

Examples of root causes for pins clamps, principal locating point (PLP) blocks and weld guns are contained in Plonka and Fromm, [1995] and Plonka, Ceglarek, Fromm et. al. [1996].

5.0. Capture & Communicate Knowledge:
Following are two approaches that were developed in order to capture and communicate knowledge:

1. ABC Production System
2. Lessons Learned from Partners.

5.1. ABC Production System
A generic model of the activities performed in designing and building automobiles from capturing the voice of the customer, at early product definition through customer ratings of delivered vehicles, was developed. The flowchart in Figure 5 shows the ABC Production System. Public presentations and non-proprietary documents by companies were reviewed to test the framework and to ensure its completeness. Figure 5 also show how different 2mm projects fits into the ABC production system. ABC Production System, provides a road map for technology and knowledge transfer in the following ways:

- It identifies and documents the fundamental activities, principles, and methodologies employed throughout the supply chain to achieve dimensional control and variation reduction
- It facilitates the organization and sharing of the continuous improvement and variation reduction knowledge gained throughout the supply chain that should be used in future car programs.

- It can be used to show the contribution from each research project to variation reduction.

The method we have followed in 2mm Lesson Learned book, [Plonka, Ceglarek, Fromm, et.al. 1996] is to arranged lessons learned according to the following six major ABC building block activities to provide an easier route for the readers to reach their desired destinations: Quality Deployment, Define Product, Define and Validate Process Concepts, Design and Construct Tools, Launch and Run Production, and New Model Product and System Improvement. The production system activities are color-coded, as shown in Figure 5. Similarly the fish bone diagram that deals with the variability contributors are also color coded in the similar fashion. This color coding allows the user to link a particular root cause/lessons learned to an ABC production system activity.

5.2. Knowledge from Partners.
ABC companies were interviewed to gain their insight on the Production System and lessons learned in working with the OEMs. The information was captured from individuals concerned with design, quality, operations, and the manufacturing process. They shared their thoughts and experiences regarding sources of variation and methods to reduce it. Their comments were added to the production system framework. In addition, several concepts where approaches differ widely, such as using a single tool versus multiple tool assembly processes, were identified. These present a best-practice challenge to the industry.
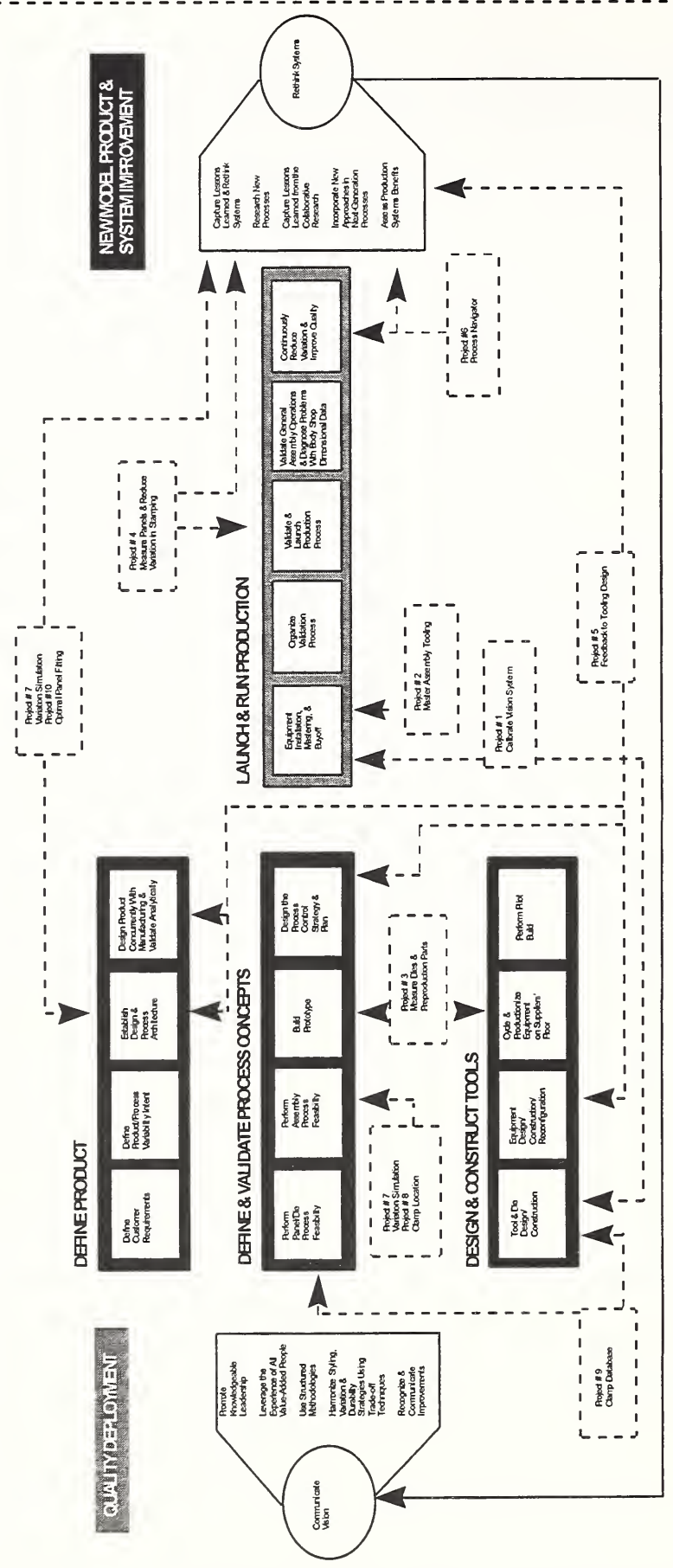
With in quality deployment discussions with many of the tooling suppliers indicate that they take a life-cycle approach to the design of their products. Structured quality operating and continuous improvement systems are used by all. They are changing tactical responses to problems encountered into strategic design objectives and programs to reinforce the deployment of quality. As more attention is placed on variation reduction, efforts to improve throughput will have greater impact. Some of the observations made are listed below:

- Full-service tooling suppliers are positioned to improve process equipment life cycle performance and product quality.
- Tooling suppliers employ Total Quality Management processes and risk management techniques.
- Major emphasis is still on throughput.
- The car companies share warranty performance and improvement goals with suppliers.
- Partners emphasized the importance of architecture and the costs associated when changes are made to overcome sub-optimal decisions in design.
- Trade-offs must be made between technology and budget.
- Die tryout experiences are being captured and archived using video technology.
- The involvement of certification engineers at prototype and pilot is a positive step in improving process accuracy and learning.
- Partner companies emphases the need to focus on the robustness of tools to achieve a more agile system during designing and constructing tools activity.

The suppliers realize the advantage of total system knowledge and system improvement. They are moving towards becoming full-service suppliers with an increasing scope of responsibility.

Figure 4

Suppliers work with all three of the domestic OEMs as well as the transplants and feel a need to define a common way of operating with their customers. Toyota is held up as a model. The building blocks for a production system have potential for training the existing and emerging work force on new fundamental technologies and effective ways of improving automobile body quality. Although consortium members feel that technology transfer and knowledge sharing between member firms could be improved, all were positive on the benefits and value of doing research and challenging the status quo.

6.0. Enabling Technology:

In this section we describe six projects that provide enabling technologies for reducing both variation and lead-time. The first two projects described below addressed the certification of assembly tooling and the calibration of Optical Coordinate Measuring Machine, (OCMM), sensors, respectively. The third project developed technology for rapid scanning of prototype panels and die surfaces. The next two projects support the design of assembly fixtures and developed new clamp technology, respectively. The last project applied variation simulation analysis to evaluate improved strategies for locating door assemblies in automobile bodies.

6.1. Visibility Analysis and Sequencing Simulation in Tooling Certification

This project tested new dimensional measurement technologies and strategies to provide tool certification personnel with new capabilities. These capabilities enable them *certify* that the locating blocks and pins are *in* design location and *if not* to support any necessary adjustments. For theodolite technology, a methodology to measure contact surface directly was proposed and tested. Algorithms to optimize the setup of theodolites were developed and tested. As an alternative to theodolite technology, this project performed tests to validate the applicability of other measurement technologies, such as laser tracking interferometers, real-time and near real-time photogrammetry, and an articulating arm coordinate measuring machine.

6.2. In-Line Measurement System Design and Calibration

This project developed two technologies that aid gauge station designers in designing optical coordinate measurement machines and enable certification engineers to calibrate sensors without the need for scribed sheet metal masters. The methodology utilizes a CAD system that supports designing the mounting structure (or tunnel) around the sensors and mounting hardware for optimal sensor placement relative to the required field of measurement. The calibration of sensors and any needed adjustments are accomplished by using technologies that track the position of the sensors in the gauge coordinate space.

6.3. Measurement and Modeling of Free-Form Surface by Laser Line Scanning System

The technology developed though this project permits the measurement of prototype panels and die surfaces using a laser line scanning sensors mounted to a robot to reach any point and angle. Time to measure complex surfaces and features can now be reduced with an accuracy of 50 mm at three sigma.

6.4. Optimal Fixture Design for Deformable Sheet Metal Parts

This project provides a computer-aided engineering (CAE) tool that can be used during the product and process design stage to aid in the selection of sheet metal datum surfaces and the location of their corresponding assembly fixture principal locating points (PLPs). The program evaluates alternate clamping locations and weld sequences, based on expected detail part errors, in order to minimize resulting distortion and variation. This project can be used by product and process engineers to reduce variation by the optimization of clamping and weld sequences at the design stage. Production lead-time can be reduced by minimizing the modification of assembly tools during pilot and launch. By reducing the number of fixture locations, a corresponding reduction in the cost of producing the fixture can be realized.

6.5. Robust Design of Work Holding Fixture

This project developed two technologies that supported the design of assembly fixtures. The first was a clamp database and retrieval system to help the engineer select clamps and determine the effective clamping forces. The benefit of using the clamp selection tools is the reduction in variation caused by improper clamp application. The second technology was a two-stage adaptable clamp that permits controlled movement of sheet metal during the locating cycle and sufficient clamping force to hold parts in a fixed location during welding to reduce distortion. The use of two-stage adaptable clamps will ensure positive location of panels against locators and result in reduced air gaps and distortion.

6.6. Optimization in Multiple Panel Fitting

This project provided knowledge regarding variation in the body panel fitting process. Recommendations were provided for product architecture, process locating features, and joining technology based on experimental results using variation simulation analysis (VSA) modeling. Applying these recommendations can improve body panel flushness and gap variation, eliminate the need for "quality" adjustments, and lower locating hardware investment.

7.0. Conclusion:

The 2mm program was a major success for all its participants. It not only helped the participating OEMs to reduce auto body variation but also helped their suppliers to work closely with universities in the search for new ideas and technologies. It demonstrated a perfect example of collaboration and coordination between government, industry and educational institutions. The knowledge gained in the 2mm program helped in the successful launch of two additional NIST ATP programs; Near Zero Stamping and Intelligent Resistance Welding.

8.0. References

Ceglarek, D. Shi, J, 1995, *Dimensional Variation Reduction for Automotive Body Assembly*, Manufacturing Review, vol. 8, No. 2, pp. 139-154.
Hu, S. and Wu, S.M, 1992, *Identifying Root Cause of Variation in Automobile Body Assembly Using Principle Component Analysis*, Trans. of NAMRI, vol. 20, pp. 311-316.
Plonka, F and Fromm, B.,1995, *Information Feedback for Tooling and Process Design*, IBEC, Oct..

Plonka, F, Ceglarek, D, Fromm, B, et.al. ,1996: *The Capture and Communication of Knowledge: A Lessons Learned Approach*, The 2mm Program Inc. Ann Arbor MI.

.

Session II. Advanced Machining (1)

The Development of a Precision Piezoelectric Micro-Positioner for Line Boring Bar Tool Insert
G. O'Neal, B.-K. Min, Z.J. Pasek, and Y. Koren, University of Michigan; P. Szuba, Lamb Technicon

Non-Circular Turning for Camshaft Machining--Fast Response Actuators, Sensors, and Motion Control for Cam Profile Tracking
T.-C. Tsao, University of Illinois at Urbana-Champaign

Development of a Machining Diagnostic System
W.A. Kline and R. Sriram, Montronix; R.E. DeVor, University of Illinois

The Development of a Precision Piezoelectric Micro-Positioner for Line Boring Bar Tool Insert

G. O'Neal, B.-K. Min, Z. J. Pasek, Y. Koren
University of Michigan, Ann Arbor, MI

Phil Szuba
Lamb Technicon, Warren, MI

Abstract

The paper describes an effort to develop a prototype smart boring tool supporting agility at the station level by enabling automated tool change and providing an advanced on-line compensation mechanism. The tool utilizes a compensation mechanism powered by a piezoelectric microactuator driven by an advanced controller, compensating for the position error of the boring bar and effectively rejecting cutting force disturbances. The design methodologies for the actuation mechanism and the controller are presented. Experimental data were used to verify the theoretical and simulation results indicating successful operation of the tool with the bandwidth of up to 50 Hz, providing the tool tip compensation range of ±25 µm with a resolution of 0.2 µm.

I. INTRODUCTION

1.1 Background

Current high volume production of automotive engine is based on dedicated transfer line technology. Depending on the produced volume, a transfer line is typically composed of 60-70 machine stations for cylinder heads and about 140-150 stations for engine blocks. Highly dedicated tooling is used for long-term, steady production of a single model. Heavy initial capital investment can only be justified by an annual volume of 300,000 units or more and a product lifetime of 10-15 years for a single model.
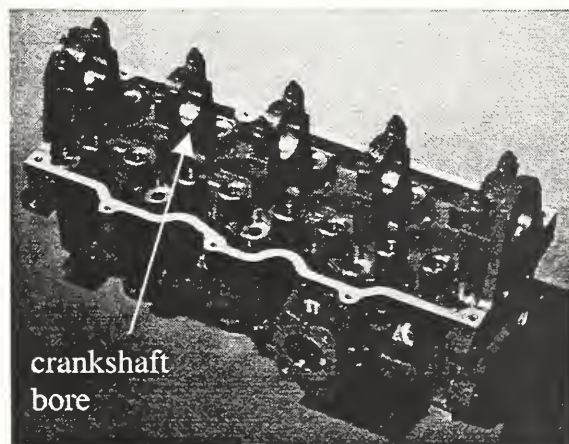

Figure 1.1 View of a typical car engine block

33

The highly competitive international market is now plagued by an overcapacity of global automobile production. Simultaneously, a customer-driven environment requires greater variety of products and the quick introduction of new products. In addition, recent government regulations on fuel economy and emissions have affected the design and manufacture of powertrain components, but dedicated tooling is impeding the quick introduction of new designs.

These changes in the automotive market drive an urgent need for new powertrain machining technologies that provide flexibility and upgradability at an affordable cost. With the introduction of CNC machines into powertrain machining systems, the level of flexibility in engine production has grown considerably in the recent years, but with one single exception, that is machining of long bores. For machining of long bores a process called line boring is employed; the process' name refers to a consecutive machining of bores with the same diameter (such as in bearing nests). This operation, however, still requires use of dedicated manufacturing equipment, and hence impedes achievement of full flexibility of machining systems. A recently conducted study [Pasek, et al., 1993, 1994] showed that a fully flexible precision line boring station is still not available. This is quite surprising in light of the critical role played by flexible line boring in the agile manufacturing of engines.

1.2 Problem Statement

Precision line boring is a very demanding application in terms of both quality and production rate requirements. Achieving increased flexibility without compromising quality, production rate or reliability (in a station that operates at least 16 hours a day) is difficult and challenging, and requires a systems approach that integrates advanced control with precise and intelligent machine hardware. To achieve this goal, leading edge technologies in machine design, tooling, and advanced control need to be applied in a coordinated systems approach.

Boring tools are recognized as a major bottleneck in increasing flexibility and precision of the line boring processes. This paper describes the effort to develop a new concept of the line boring tool, supporting agility at the boring station level. This task will be accomplished by the following means:

- enabling automated tool change of the long boring bars,
- an on-line compensation mechanism relying on sensing and intelligence built into the tool itself.

One of the major obstacles to increasing flexibility of the boring tool is that current tooling designs require outboard and intermediate support bushings. If these bushings are eliminated, automated tool change can be readily accomplished. Also, support bushing elimination increases system reliability by simplified design and elimination of contact components that wear (bearings).

However, elimination of supports makes the boring bar more compliant thus leading to vibrations and precision problems. To enhance the design of a long boring bar with vibration resistance, a number of process sensors need to be built into the tool itself. The signal processing of the sensor signals is carried out in a computer embedded in the spindle. The same computer performs a compensation function by controlling the motion of a piezoelectric actuator to which the tool tip is attached. This concept is called a "smart tool."

1.3 State-of-the-Art

1.3.1 The Line Boring Process

A number of products, such as, for example, engine block cam and crank shaft bores, gearbox bearing houses or screw pump case bores, require boring of long holes by utilizing a line boring process (explained below) to ensure tight diameter and concentricity tolerances. Currently, all line boring operations performed on these parts on an industrial scale are carried out using specialized (dedicated) equipment.

In powertrain components where bearing support is necessary for cam and crank shafts, a number of precision bearing nests are located in-line. When the distance between two such bores is not too large (less than 5-10 bore diameters), line boring is employed (Figure 1.2). Line boring ensures that the holes are of the same diameter and concentricity error is minimized. The process is traditionally carried out by one of the three methods for high volume applications.

When transfer lines were first implemented, line boring bars were supported at both ends to obtain adequate rigidity. Therefore, all the journals were machined simultaneously by including multiple cutting inserts down the length of the boring bar (Figure 1.2).
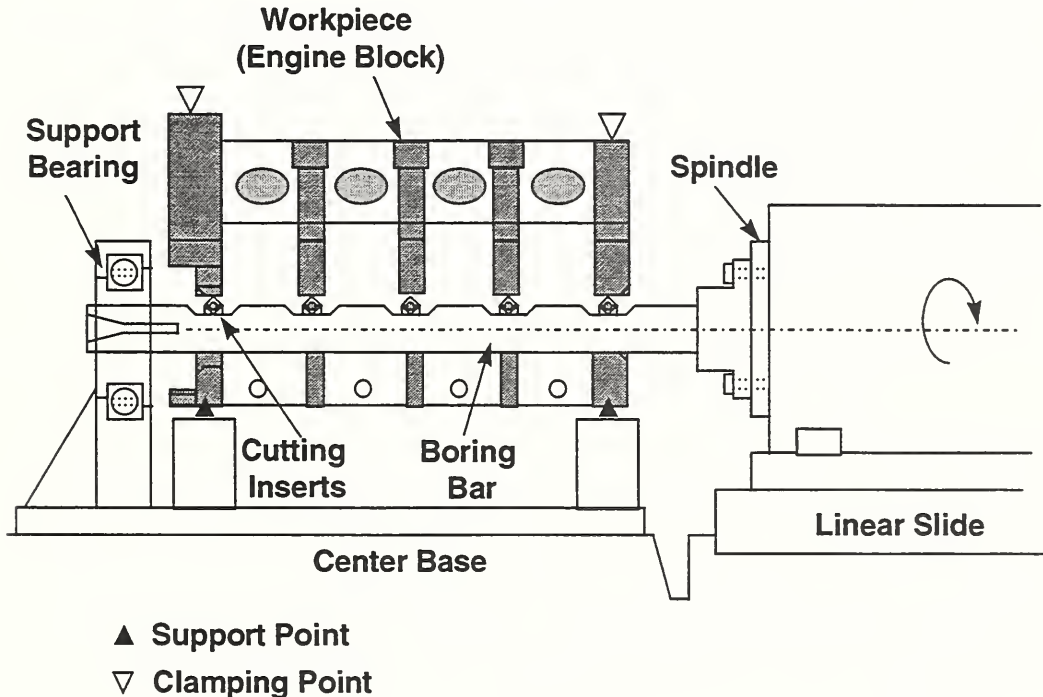
Figure 1.2 Conventional line boring with outboard support and multiple Cutting Inserts

The next evolution was a boring bar with longitudinal carbide guide pads [Pond, 1987], which are ground to a diameter slightly smaller than the desired final bore diameter. In this process, a single-point dedicated tool bores the first hole, and is subsequently fed into the part until machining of all journals are completed. The need for the outboard support is eliminated by using the part itself to support the tool as subsequent journals are machined. Because the length of the boring tool projecting from the spindle is constant (the part or the spindle itself moves), the deflections of the boring bar are also expected to be constant and the holes will have nearly identically shaped centerlines (assuming adequate rigidity of the part and fixturing).

Hence when the bearings are pressed into the bores, the shaft they support can be supported with minimum geometric congruence.

The newest line boring development being used in manufacturing today is an extension of the guide pad concept. This method moves the tool in the feed direction relative to the spindle. By doing so, the spindle can be brought very close to the first journal of the part, eliminating the need of an inboard support. The tool is then pushed through the spindle as subsequent journals are machined. This technique minimizes tool deflection by minimizing tool overhang and maximizing tool support (Figure 1.3).
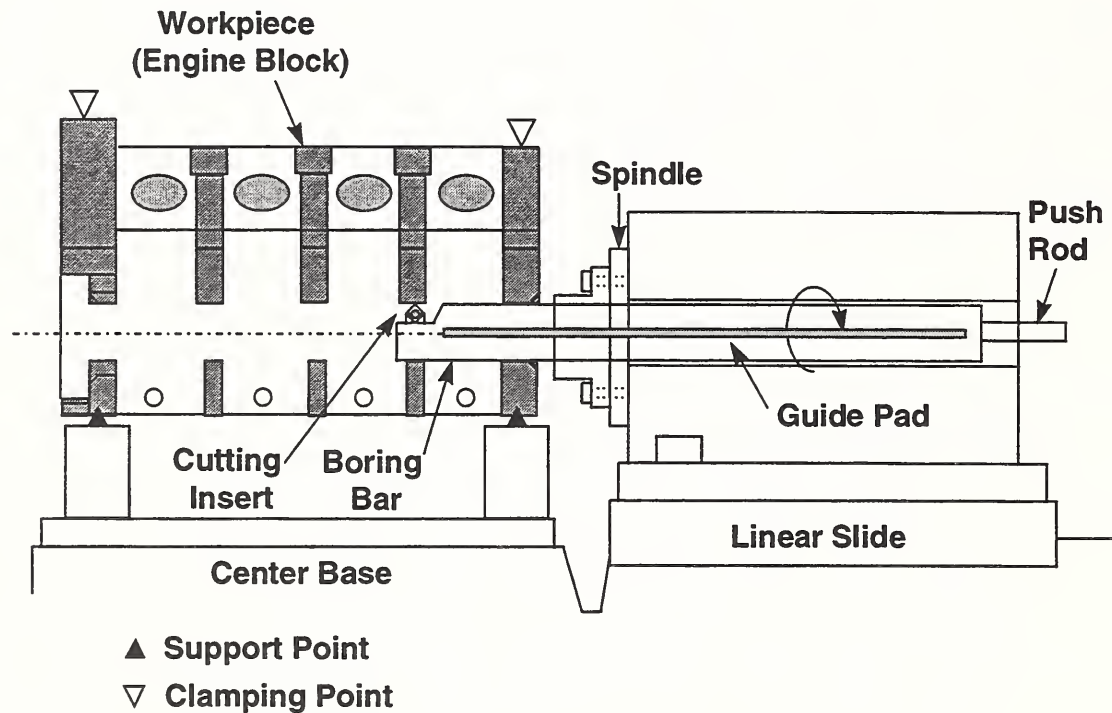
Figure 1.3   Line boring with through-the-spindle guided tooling.

In typical automotive applications the range of machined bore diameters is D = 30 - 75 mm (1.5" − 3.0"); the required cylindricity of the bore lies within 10 - 25 μm, and the surface finish is R_a = 2 - 50 μm. In extreme cases boring bars can be as long as 1350 mm (53") and weight up to 120 kg (250 lb).

The line boring station is the most dedicated station in powertrain machining transfer lines. In a typical line boring station the spindle is mounted horizontally and can travel only in the horizontal direction, which affects the length of the machine. Attached to the spindle is a single dedicated boring bar which is responsible for machining the bore through the entire length of the engine casting. Due to the length of the bore the boring bar is usually supported by bushings at a number of points to ensure high precision of the operation. To facilitate entry and exit of the boring bar into the bore, the workholding fixture is provided with a raising and lowering motion. The boring operation is performed sequentially, under control of a hard-wired system. Material handling is carried out by an automatic transfer mechanism (such as a "walking beam") which continuously cycles and moves the engines serially from one station to the next.

1.3.2 Current Line Boring Technology

The state-of-the-art in flexible line boring technology has been assessed in a recent benchmarking study [Pasek et al., 1993], consisting of interviews with machine tool suppliers and other experts, and through literature and patent searches. Results of the searches turned out to be relatively limited, most likely because proprietary information is not released by manufacturers, non-proprietary research is very limited in this specific area, and academic research is almost non-existent in this area.

While most of the engine producers still rely on dedicated transfer lines, solutions providing a certain degree of flexibility are gaining more popularity. A number of relevant patents propose particular devices aimed at improving the boring process. These devices were invented to overcome some typical shortcomings of the boring process in production of long holes, such as the low rigidity of the boring bar. Several solutions were proposed, such as pretensioning of the boring bar [Fujiwara, 1991;Schechter, 1993] or application of additional support [Johne, 1991]. Some devices dealt with the issue of boring bar centering in a number of coaxial holes. A patent by Johne [1991] innovated a machine tool precision boring head which uses flexible parallel springs to serve as an end support. Another patent [Tajima, 1988] enhances the productivity and machining accuracy by automatically exchanging a line boring bar and support bushings and by automatically adjusting the position of supports, the boring bar and a workpiece with the use of an NC machine tool. Another patented solution [Hashimoto, 1987] proposes a convertible line boring station, utilizing common features throughout the product line. The design enables the simultaneous change of both machine head (boring tool) and the jig (fixture); the system is relatively complex and expensive. From these technology reviews it can be concluded that a system including a fully flexible line boring station is not currently available and no method has been patented to machine cam and crank journals in a flexible manner.

Reviews of the recent developments in the general-purpose boring technology [Aronson 96, Mason 97] reveal new trends in the area of traditional boring applications. The most prominent trend observed is an increased presence of instrumentation and control in boring (tools integrated with electronics) in response to increasing precision demands.

Much of the research in the boring area has been focused on vibration reduction related to either one of two sources: forced vibration (due to cutting forces and mass unbalance) or self-excited vibration (chatter). Studies to eliminate or reduce vibration explore either passive or active methods. Passive methods include use of: (i) materials with high elastic modulus, (ii) passive dynamic vibration absorbers.

Materials with high elastic modulus, such as machinable tungsten alloys and sintered carbide, directly improve stiffness of the boring bar. They are, however, quite expensive and may potentially cause a serious unbalance problem during bar rotation. Rivin [1993] proposed a combination bar made of several materials to improve rigidity and reduce the problem of unbalance. In addition, Rivin [1983 and 1992] proposed use of passive dynamic vibration absorber (DVA) to increase the damping of the bar structure. Lee [1988] adopted a bar of graphite epoxy composite to improve both stiffness and damping. Slocum [1994] introduced the concept of the replicated internal viscous damper in the interior core of the bar to improve damping. Andreassen [1995] devised a adjustable, cylindrical damper for boring bars, in which ring-shaped, elastic spring elements and damping oil provide the damping action. Nevertheless, all passive methods can only improve boring accuray to a certain level; further improvement can be achieved only by using active vibration reduction approaches.

A number of research papers deal with boring bar design involving active control. Ulsoy and Koren [1993] reviewed controllers for servo, process and supervisory control levels. Kim et al. [1987] used Forecasting Compensatory Control (FCC) to construct an autoregressive (AR)

stochastic model of the cylindricity error by laser measurement. Based on a the forecasted error signal, a feed forward controller generated a control command to compensate. The compensation was carried out by the piezoelectric actuator working in open-loop, and susceptible to hysteresis and drift in the piezoelectric material, with no ability to actively reject cutting forces.

Tewani et al. [1988] introduced the idea of an active vibration absorber. It was further developed by Stephens [1989]. The optimal state feedback was used to control the active dynamic absorber in Tewani et al. [1991a]. Tewani et al. [1991b] used the distributed model of a boring bar by finite element analysis instead of a lumped system model to account for the continuous bar structure. It was shown that the transient response of the lumped system degenerates as the length-to-diameter ratio of the bar becomes larger. In Marra [1995] the H-infinity optimal control was used for the active dynamic absorber and it was shown that it had better performance in external disturbance rejection than the optimal control strategy. However, the uncontrollable vibration in the transverse direction still remained.

Another solution for an active damper was proposed by Tanaka et al. [1989] and was further improved for practical use in Tanaka [1994]. An adaptive control scheme was used to generate compensatory signals according to the fluctuating chatter frequency which was monitored.

Hanson and Tsao [1994] developed variable-depth-of-cut machining approach which uses a fast tool servo to eliminate the error between the desired tool position and the actual one by feedforward or repetitive controller. In addition, an analytical model was used for designing the tool servo. Since the boring machine was not available, the experimental tests were conducted by turning the workpiece on a vertical milling machine.

It has to be stressed, however, that most of the reviewed works focused on standard boring processes for which L/D ratio of the bar did not exceed 5:1, or addressed similar applications, including turning.

II. SYSTEM OVERVIEW

Overall Smart Tool system layout is presented in Figure 2.1. The boring bar is located in the motorized, high-precision spindle (10 HP, max rotational speed 6,000 rpm) and is connected with it through the standard taper mechanism enhanced with additional electrical connectors. The tool body (see Figure 2.1) contains a number of sensors and the actuation mechanism. The necessary instrumentation supporting the sensors and the actuation mechanism is located in the package attached to the rear end of the spindle and rotating with it. The instrumentation package contains the following: (i) on-board computer, (ii) laser electronics, (iii) position sensor electronics, (iv) wireless transmitter and antenna for communication with the main controller of the machine, (v) power electronics for the piezolectric actuator. Electrical power necessary to run the piezoelectric actuator (125 V, 2 amps) is transferred via a set of slip rings located behind the instrumentation package.

The housing for the instrumentation was designed in such a way that it can work with multiple boring bars thus eliminating the need for frequent disconnections.
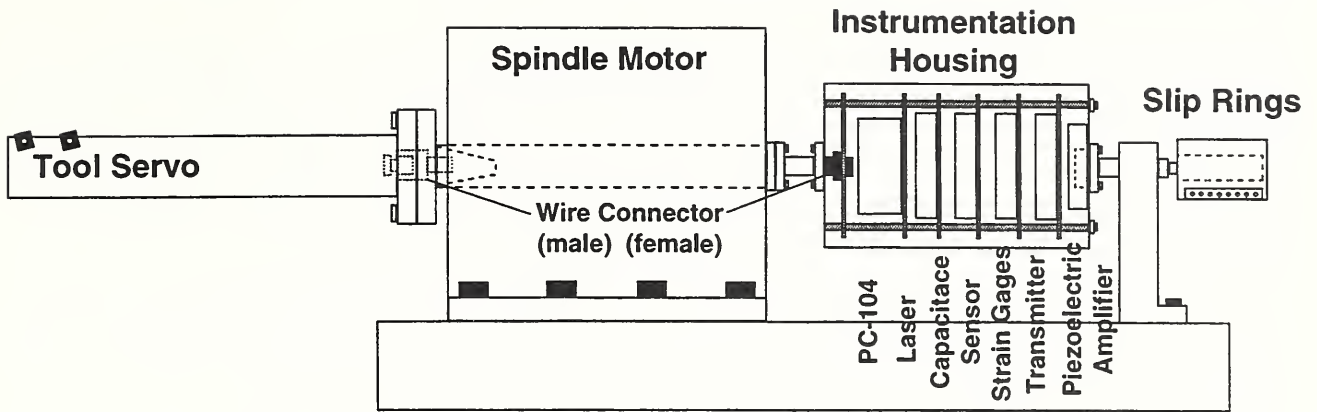
Figure 2.1 Overview of the Smart Boring Tool System

The inner details of the boring bar are shown in Figure 2.2. Functionally, the bar contains three subsystems:
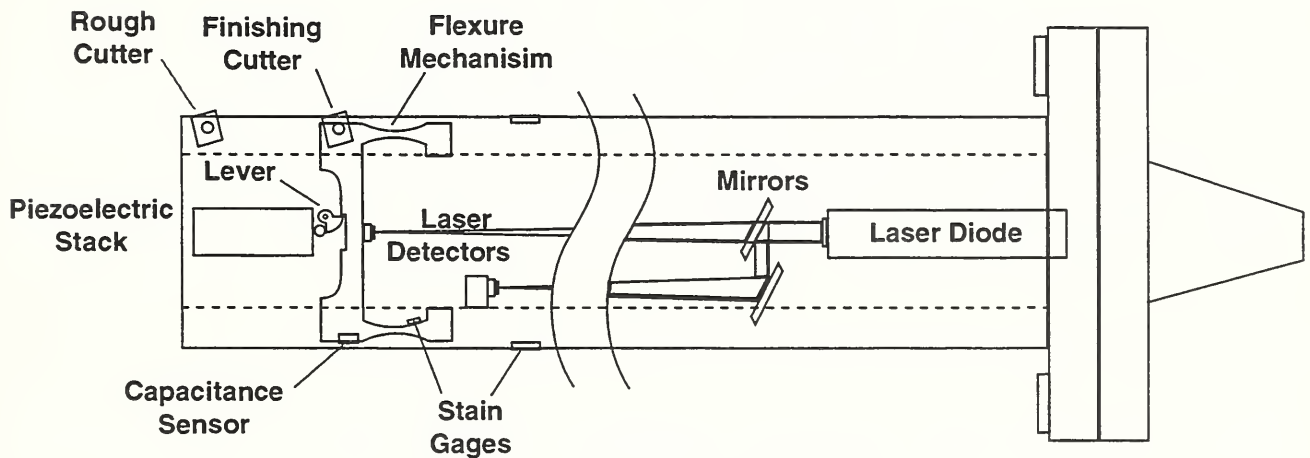
- sensors,
- actuation mechanism,
- cutting inserts.

Figure 2.2 The Smart Boring Tool

III. PIEZOELECTRIC ACTUATOR

The fast tool tip servo requires an actuator with a high cut-off frequency to compensate for the vibrations in the boring bar, while generating the forces necessary to accelerate the tool tip translation mechanism. This actuator must be compact enough to fit inside the boring bar, and light enough not to unduly affect the vibration resistance of the boring bar. Four possible technologies were considered for this application: solenoids, electric motors, magnetostrictive material, and piezoelectric material. Of these, a piezoelectric stacked actuator was determined to be the best solution for the system.

Piezoelectric material has the property of changing shape in an electric field, allowing it to be used as an electromechanical transducer. Piezoelectric actuators have high operating frequencies, and high output energy per unit volume and mass. However, they have the disadvantage of small strokes, often requiring mechanical amplification. Since the piezoelectric material is brittle, a tensile and shear stress should be avoided. Therefore, a preload force should be applied that assures the piezoelectric material has only compressive stress.

The mechanical behavior of a piezoelectric actuator is analogous to a distributed mass spring with variable free length, where the free length is approximately proportional to the applied voltage. A mechanism to translate the tool tip (Section IV) can be modeled as a mass-spring-damper second order system. To magnify displacement from the piezoelectric actuator, a lever is used to connect with the tool tip translator. If the mass of the piezoelectric is negligible compared to the mass of the translation mechanism, then a lumped model can be used to describe the system dynamics as shown in Figure 3.1.
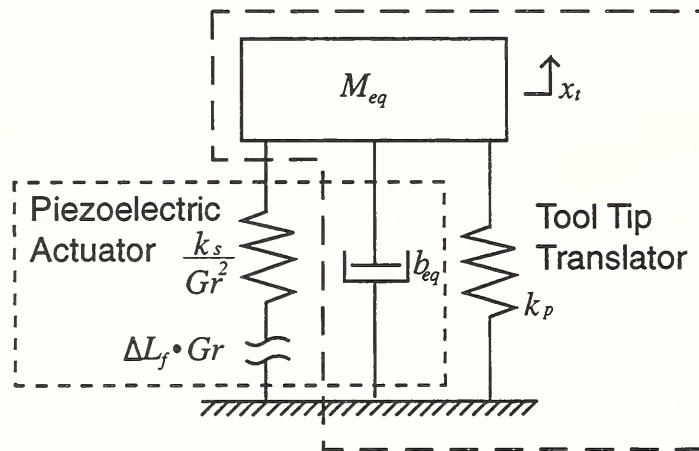
Figure 3.1 Lumped Mass Model of Fast Tool Servo

Where ΔL_f is the change in free length of the actuator, which is approximately proportional to the applied voltage. This relationship does not account for the non-linearity associated with hysterisis or drift. k_p is the stiffness of the piezoelectric actuator, and k_s is the stiffness of the tool tip translation mechanism at the tool tip. x_t is the position of the tool tip. M_{eq} is the equivalent mass and b_{eq} is the equivalent damping coefficient, for the system at the tool tip. Gr is the mechanical leverage ratio of the lever.

The maximum steady state range of the tool tip can be derived from this model.

$$Range = \left(\frac{k_p}{\dfrac{k_p}{Gr} + k_s \cdot Gr} \right) \cdot \Delta L_m \tag{3.1}$$

Where ΔL_m is the maximum change in the free length of the piezoelectric material, which is usually specified by the manufacturers.

The natural frequency of the system can also be derived from the model.

$$\omega_n = \sqrt{\frac{k_s + \dfrac{k_p}{Gr^2}}{M_{eq}}} \tag{3.2}$$

The capacitance value usually supplied by the manufacturer is for the piezoelectric actuator operated below the natural frequency with no load. In the current system, this stiffness of the tool tip translator (k_s) was kept small relative to (k_p) to increases the range of the actuator (Eqn. (3.1)). The preload is set to half the blocked force allowing similar accelerations in both direction of the tool tip movement, while always keeping the piezoelectric material in compression. The natural frequency of the actuator is designed to be more than two times maximum operating frequency for control purposes. Under these conditions, the manufacturer's supplied capacitance is a good estimation of the actual capacitance (c). The maximum frequency (f_{max}) the system can be driven at is given by the following expression.

$$f_{max} = \frac{i_{max}}{\pi \cdot c \cdot V} \tag{3.3}$$

Where (V) is the full range of supply voltage, and i_{max} is the saturation current of the amplifier. This equation shows the importance of the linear amplifiers saturation effects. The maximum frequency is a function of the amplitude of the vibration.

When choosing a piezoelectric there is a trade-off between its stiffness, maximum free length, and capacitance. For a given piezoelectric material with specified maximum electric field, the following proportional relationships are helpful in selecting the size of the actuator to be used.

$$\Delta L_m \propto (height) \tag{3.4}$$
$$(Blocked\ Force) \propto (area) \tag{3.5}$$
$$c \propto (volume) \tag{3.6}$$

Where height, area and volume are the physical sizes of the actuator.

IV. TOOL TIP TRANSLATION MECHANISM

The tool tip translation mechanism enables actuation of the cutting insert in the thrust direction, while simultaneously rejecting cutting and feed forces. Only thrust forces are transmitted to the piezoelectric actuator. In the ideal case, the translation mechanism should constrain the cutting insert exclusively to rectilinear motion. Parasitic motion in the axial direction affects the feedrate, which in turn may increase surface roughness. The translation mechanism should also be designed to provide the preload to the piezoelectric material, thus the connection to the piezoelectric actuator only needs to transfer compressive loads, allowing it to be a simple point contact. If this preload is applied internally by the actuator itself then the translation mechanism will need to be pulled on by the actuator when contracting, requiring a connection capable of transferring tension loads. Since the piezoelectric material is brittle, it needs to be isolated from shearing or tensile forces. The translation mechanism should be dynamically balanced about the axis of spindle to reduce the effects from rotational motion.

To expedite the design process, a tool tip translation mechanism similar to one that appeared in a paper by Hanson *et al.* [1994] was used in the tool tip servo, while simultaneous work was done on an alternative design. The initial mechanism has been used in all of our experiments (Section VI). This design has the advantage of being simple in construction and design, but it is inherently unbalanced and has too much parasitic motion in the feed direction to be suitable for the flexible boring machine.

The proposed final design is based on a parallel leaf spring (Figure 4.1). This design is symmetric about the spindle axis and hence dynamically balanced. The symmetry of the two parallel leafs reduces the tendency of the base to pitch, limiting the parasitic motion in the feed direction. One disadvantage of this design is that an additional mechanism is required to convert and magnify the piezoelectric actuator motion from feed to thrust direction.
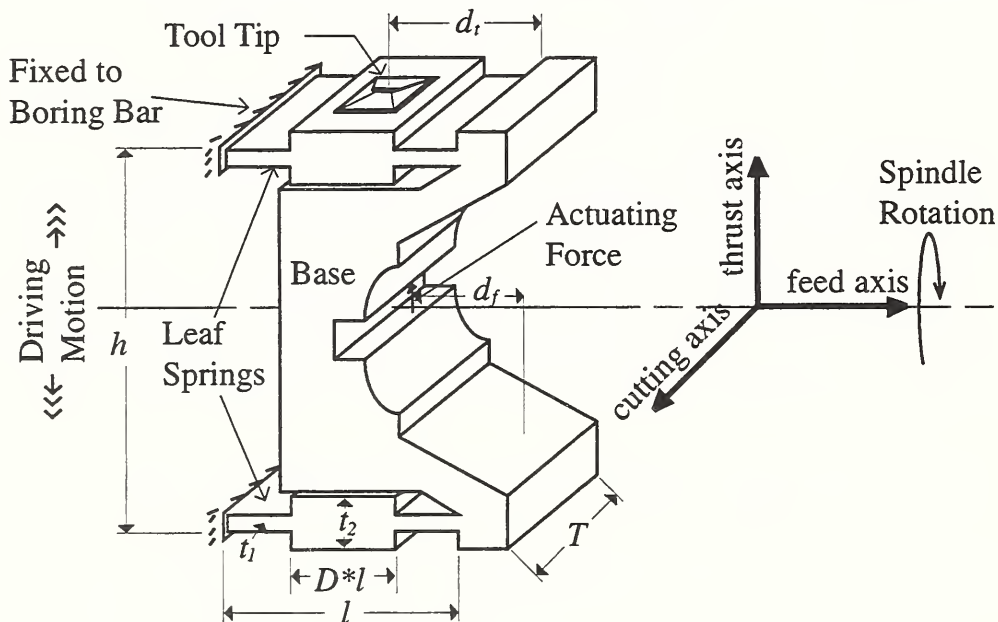

Figure 4.1 Proposed Final Design (Parallel Leaf Spring)

4.1 Design Methodology

There are many factors, which must be considered in the design of this mechanism. The following design methodology uses equations derived from solid mechanics principles to find a balance in these trade-offs. The designer can specify the desired width of the leaf springs (T), stiffness to an actuating force (c), and safety factor for fatigue (Sf). The values for d_f, h, l, d_t, t_1, and D are then determined to meet these desired specification in an optimum manner.

The line of action of the actuating force should pass near the center of the leaf springs ($d_f = l / 2$), and the height of the parallel leaf spring should be chosen as large as possible ($h = \sqrt{(D_b^2 + T^2)}$). These measures reduce both the parasitic pitching motion of the base, and the strain induced in the leaf springs during actuation.

The value for t_1, l, and D are determined by simultaneously solving Eqns. (4.2), (4.3) and a fatigue formula, which may require and iterative approach since the equations are non-linear. Eqn. (4.2) insures that the parallel leaf spring has the desired compliance (c) to actuating forces.

$$c = \frac{l^3}{4 \cdot E \cdot T}\left(\left(\frac{1}{t_2^3} - \frac{1}{t_1^3}\right)D^3 + \frac{1}{t_1^3}\right) \tag{4.2}$$

The ratio of the compliance of the leaf springs in the drive direction to that in the cutting directions can be increased by thickening a middle section of the leafs. Setting the derivative of the ratio of these compliances with respect to D to zero results in Eqn. (4.3). The roots of this equation are possible optimum values of D, in the sense that this compliance ratio is maximized.

$$6\left(\frac{t_1^3}{t_2^3} - 1\right)D^3 + 3\left(\frac{4\left(\frac{t_1^3}{t_2^3} - 1\right)}{\left(\frac{t_1}{t_2} - 1\right)} - 1\right)D^2 - 3 = 0 \tag{4.3}$$

The compliance in the cutting direction is for a force applied to the end of the leaf spring. The compliance in the drive direction is for a force applied rigidly to the base end of the leaf spring through a rigid link, such that its line of action is through the center of the leaf spring. This force location results in strain geometry similar to what the complete parallel leaf spring experiences during actuation. Therefore, the relative compliance of the entire mechanism is addressed by optimizing the compliances of the leaf springs.

Any fatigue formula that relates the safety factor to the alternating and mean stresses can be used in the simultaneous equations. The designed mechanism has large strains due to actuation forces, and small strains due to cutting and feed forces. Therefore, the following analysis only accounts for the prior strains. The critical location for fatigue failure is on the surface of the leaf springs at either end. Therefore, the alternating (σ_{pa}) and mean (σ_{pm}) components of the principle stress at these locations should be used in the fatigue formula. The alternating stress (σ_{pa}) arises from the piezoelectric movement about the preload offset position.

43

$$\sigma_{pm} = \frac{6 \cdot E}{t_1^2 \cdot l^2 \cdot \left(\left(\dfrac{1}{t_2^3} - \dfrac{1}{t_1^3}\right)D^3 + \dfrac{1}{t_1^3}\right)} \cdot (\text{Range of Actuator}) \qquad (4.4)$$

The mean stress (σ_{pm}) is due to the piezoelectric preload, which is set to be half of the blocked force, which was described in Section III, of the piezoelectric actuator.

$$\sigma_{pm} = \frac{3 \cdot l}{4 \cdot t_1^2 \cdot T} \cdot (\text{Blocked Force}) \qquad (4.5)$$

Since these three equations are non-linear, multiple solutions exist when they are solved simultaneously. Only solutions that have values of D between 0 and 1 are physically meaningful. Further, it is possible that D equal to 0 or 1 gives the maximum relative stiffness. Therefore, D should be set to these boundary values and t_1, and l, solved with the remaining two equations. Designer can then choose from this finite set of solutions.

There are several feasible locations for the tool tip to be placed along the feed direction on the base of the mechanism. The cutting direction compliance of the leaf spring is a function of this location (d_t). But, the drive direction compliance is not, because the actuation force is applied to the base and not the tool tip. Further, the feed direction compliance is also independent of the tool tip location, since this force's line of action is unaffected by this placement. The optimal location for rejecting cutting forces was determined by examining the compliance of a single leaf spring as a function of the point of application of the cutting force along a rigid link fixed to the base end of the leaf spring. This compliance was found to be minimized when the tool tip is located in the middle of the leaf spring ($d_t = l / 2$).

This tool tip location requires a through hole in the middle of the thickest section of the leaf springs. For dynamic balance both leaf springs have symmetric holes. These holes have limited effect on the stiffness in the cutting or feed directions because most of the deformation occurs in the thinner sections of the leaf springs. Further, the moment of inertia associated with bending in the cutting direction is least dependant on this central mass. Of course, any reduced stiffness in the drive direction is advantageous.

V. MEASUREMENT SYSTEMS

In order to provide a feedback signal to a piezo-actuator control loop for tool tip adjustment, a laser measurement system is used. Two position sensitive optical detectors, one for measuring the position of the cutting tool and a second for measuring displacement of the end of the boring bar. The detector measuring the position of the cutting tool is a single axis bi-cell detector. Such a detector has been chosen because the cutting tool motion is constrained to one degree of freedom and the information for the piezoelectric actuator is required only for the same degree of freedom. Motion in any other direction have negligible effect on the accuracy of the cut. The detector measuring the displacement of the end of the boring bar is a two-dimensional continuous position sensitive detector. It provides supplemental information about the motion of the end of the bar body.

The sensitivity of the position measurement is limited by noise of the electronic components (amplifier). Under ideal conditions in a screen room environment and using a filter to reduce the noise level, a resolution of about 0.1 μm has been achieved.

The use of position sensitive optical detectors is based on the assumption that the vibration of the end of the boring bar held in the precision spindle is negligible. Any motion observed by the detector is therefore due solely to motion of the cutting tool or the cutting and of the boring bar with no contribution from the end of the rod located in the spindle. If repeatable errors are produced by the spindle motion (spindle error motion), these can be mapped and corrected.

Two optical beams are produced by splitting the beam from diode laser using a beam splitter (See Figure 2.2.). The laser and the focusing optics are located on the axis of rotation. One beam passes through the axis to the detector attached to the cutting tool. The second beam is directed to an off-axis detector attached to the end of the support bar.

While laser measurement system is the primary source of the feedback signal for the actuator control, two other sensors have also been employed at various stages of experimental work. These are (i) a capacitance probe to measure the tool tip position, and (ii) a set of strain gages mounted on the outside surface of the boring bar.

VI. TOOL TIP CONTROLLER

6.1 Introduction

The purpose of tool tip controller is to move the tool tip to compensate the position error of the boring bar and to hold the tool tip still relative to the spindle when disturbances occur. The prototype tool tip actuator has a capacitance type gap sensor to provide the position feedback and a laser system is being designed to provide the controller with the displacement information for the tool tip relative to spindle. Therefore, the control algorithm has been designed to track a reference signal. A digital linear quadratic optimal controller has been designed to implement in the controller computer. Computer algorithms run at 5kHz sampling frequency.

6.2 System Identification of PZT Actuator Prototype

The parameters of the difference equation for the fast tool tip servo have been determined experimentally. Figure 6.1 shows the tool tip servo experimental setup. The equation represents the transfer function between the voltages into the piezoelectric actuator amplifier and the calibrated signal out of the proximity sensor. It is important to note that these equations do not account for the cutting forces, because the system identification was conducted without cutting.

From the physical model of the tool tip servo described in Section III, it was determined that the system could be fitted to a second order difference equation. The damped natural frequency, damping ratio and DC gain were determined from the open loop step response. To validate the accuracy of the model, the simulation results by Matlab Simulink and real step responses were compared. This second order model best reproduced the dynamic behaviors of the system. However, the DC gains of real system varied with respect to different input voltage due to non-linearity in the tool tip servo. Eqn. (6.1) is the difference equation of the tool tip actuator with 0.0002 [sec] sampling time.
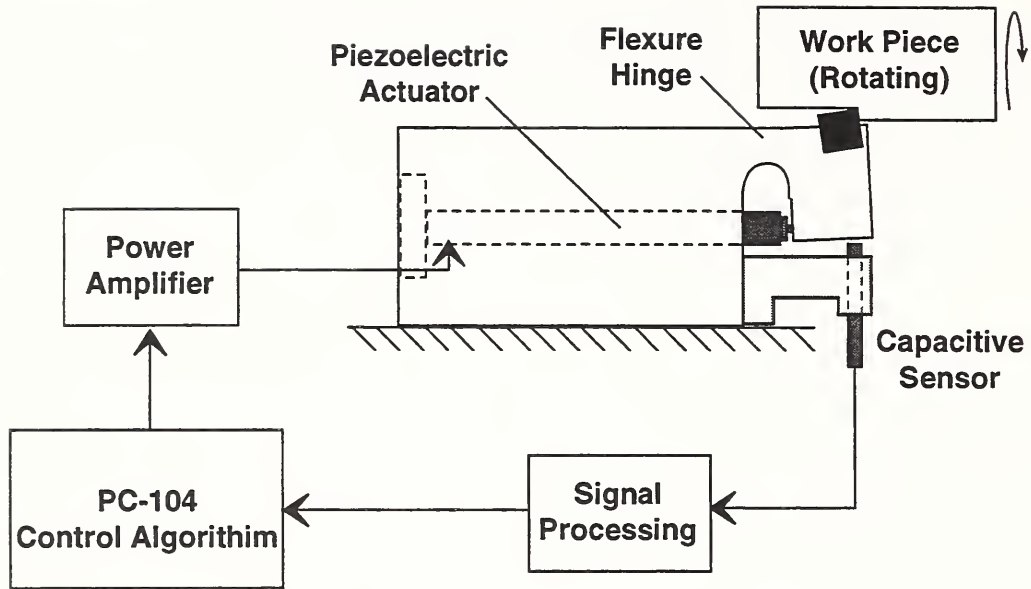
Figure 6.1 Tool tip servo experimental setup

$$y(n+1) + 1.6100\,y(n) + 0.8645\,y(n-1) = -8.4786u(n) - 7.9603u(n-1) \qquad (6.1)$$

Where, $y(n)$ is the present position of the tool tip in [μm] and $u(n)$ is amplifier input voltage in [V]. Next, a state space representation was found for this difference equation, where the first state variable (x_1) is the tool tip position (y).

$$x(n+1) = \mathbf{A}x(n) + \mathbf{B}u(n) \qquad (6.2)$$
$$y(n+1) = \mathbf{C}x(n)$$

where,

$$\mathbf{A} = \begin{bmatrix} -0.6711 & 1.9873 \\ -0.1179 & -0.9389 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} -8.4786 \\ 0 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

6.3 Controller Design

6.3.1 Observer design

A full state feedback control law requires knowledge of all the current states of the system. A state space realization was chosen in such a way that the first state is current tool tip position (detected directly from the proximity sensor). However, in this state space realization, the second state is not a previous value of tool tip position, which requires the $\mathbf{A}$ matrix has the form of Eqn. (6.3).

$$\mathbf{A} = \begin{bmatrix} * & * \\ 1 & 0 \end{bmatrix} \tag{6.3}$$

Since the second state can not be derived directly from the proximity sensor signal, a Kalman filter observer [Phillips and Nagle, 1995] has been designed to estimate the states in real time for the full state feedback. An observer was designed of the form

$$\hat{x}(n+1) = [\mathbf{A} - \mathbf{L} \cdot \mathbf{C} \cdot \mathbf{A}]\hat{x}(n) + [\mathbf{B} - \mathbf{L} \cdot \mathbf{C} \cdot \mathbf{B}]u(n) + \mathbf{L}y(n+1) \tag{6.4}$$

$\mathbf{L}$ was chosen according to the Kalman filtering method. This Kalman filter observer attempts to minimize the expected values for the square of the errors in the estimated states. These errors by the observer are assumed to come from plant disturbances and measurement noise, which have white Gaussian distributions. Thus the state equation is assumed to be of the form of Eqn. (6.5).

$$x(n+1) = \mathbf{A}x(n) + \mathbf{B}u(n) + \mathbf{B}_1 w(n) \tag{6.5}$$
$$y(n+1) = \mathbf{C}x(\mathbf{n}) + v(n)$$

Where $w(n)$ is the plant disturbance and $v(n)$ is the measurement noise. By assuming various values for the covariance of $w(n)$ and $v(n)$ the controller can be properly tuned to achieve a compromise between fast estimator dynamics and filtering of the sensor signal. It is important to note that this observer is a current observer, using the latest proximity sensor reading to calculate the estimated states.

6.3.2 Feedback gains

Without considering reference input, a time-invariant full state feedback controller (Phillips and Nagle, 1995) which has form of Eqn. (6.6) has been designed.

$$u(n) = -\mathbf{K}x(n) \tag{6.6}$$

The gain vector $\mathbf{K}$ is chosen in the optimal manner to minimize the quadratic cost function J in Eqn. (6.7)

$$J = \sum_{n=0}^{\infty} \mathrm{x}^\mathrm{T}(n) \cdot \mathrm{Q} \cdot x(n) + u^T(n) \cdot \mathrm{R} \cdot u(n) \tag{6.7}$$

The J is a weighted sum of the states and the control inputs. The weighting factors Q and R decide the controller characteristics by a compromise between low control effort and small

position errors. These weighting factors were tuned first from the simulations and then refined by experiments.

6.3.3 Integral controller

In order to have zero steady state error in the tool tip position, in the face of cutting forces and modeling errors, the integral of the position error is fed back in parallel with the full state feedback. The integral control has been accomplished within the framework of the optimal controller designed above, by including the integral of the position error as an extra state of the system, when the full state feedback gains are selected. The augmented state matrix representation is given by

$$\tilde{x}(n+1) = \begin{bmatrix} A & 0 \\ C \cdot Ts & 1 \end{bmatrix} \tilde{x}(n) + \begin{bmatrix} B \\ 0 \end{bmatrix} u(n)$$

(6.8)

$$y(n+1) = [C \quad 0]\tilde{x}(n)$$

Where, $\tilde{x}$ includes the integral of position as an additional third state and Ts is the sampling time. Hence the gain for the integral of position error were chosen optimally.

6.3.4 Feed forward gains

The controller described above was designed to optimally drive all the states to zero. Feed forward control is used to adapt this controller to tracking problems by introducing the gains N_x and N_u. The N_x and N_u convert the reference into state and command references respectively, which are the steady state values required to achieve the reference position. The difference between the state references and the actual states is optimally minimized, as well as the difference between the command references and the actual command.

By the condition that output of system y should be same as reference input r at steady state, the gain N_x and N_u can be easily decided by following Eqn. (6.9).

$$\begin{bmatrix} A-I & B \\ C & 0 \end{bmatrix} \begin{bmatrix} N_x \\ N_u \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

(6.9)

$$\begin{bmatrix} N_x \\ N_u \end{bmatrix} = \begin{bmatrix} A-I & B \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Figure 6.2 shows the block diagram of the controller, which includes feedback, feed forward, and integral controller.
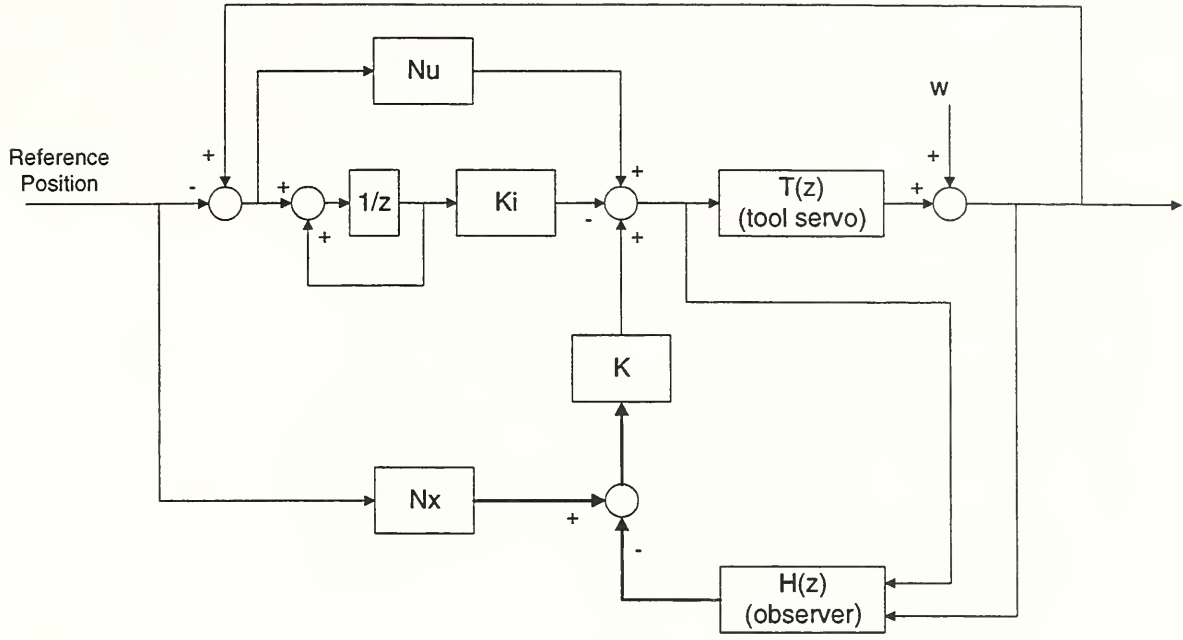
Figure 6.2 Controller block diagram

6.3.5 Theoretical performance of the controller

Essentially the controller was designed to create a stiff tool tip, which could reject the dynamic cutting forces. Through Matlab simulation, the weighting factors of the optimal control laws resulting in the gains as in Eqn. (6.10).

$$K = [\; 0.104504 \quad 0.262090 \;],$$
$$K_i = -366.571716, \quad\quad\quad\quad (6.10)$$
$$N_u = -0.211359,$$
$$N_x = [1 \quad -0.0608311]$$

Figure 6.3 shows the Bode plot of the transfer functions between cutting forces and tool tip position. The dashed line is open loop transfer function obtained experimentally. The solid line is generated from the theoretical closed loop transfer function. Theoretically, the closed loop stiffness at the tool tip of actuator is infinite, as long as the actuator does not saturate. The compliance has a maximum at around 2,330 Hz, which is near the first natural frequency of the tool tip servo. At frequencies above this natural frequency of the servo, the compliance drops off significantly, for both the open loop and closed loop cases. This is because the mass of the servo acts as a low pass filter, attenuating the frequency components of cutting forces, which are higher than the natural frequency of the servo.
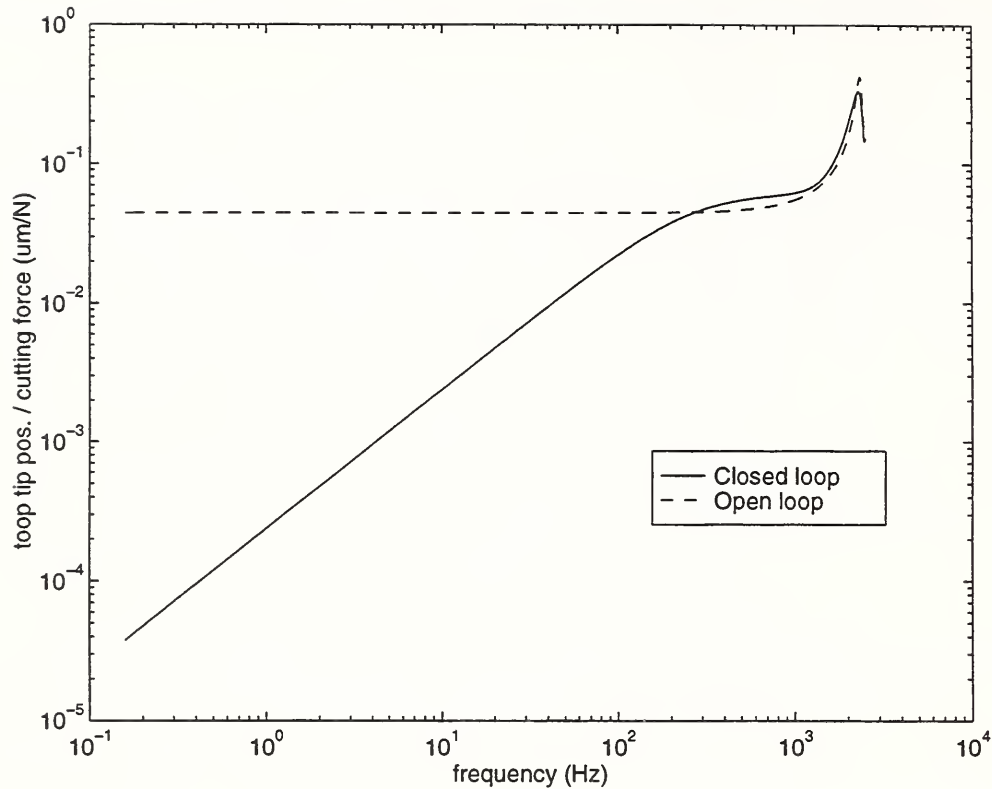
Figure 6.3 Theoretical dynamic compliance of servo

6.4 Experimental Results

6.4.1 Cutting experiment

In order to evaluate the performance of the tool tip servo under real machine tool operating conditions, preliminary cutting tests have been conducted using a stationary boring bar and rotating workpieces. The tool tip servo was attached to the tool holder of the turning machine. An aluminum tube with a 7-inch inner diameter and 1-inch wall thickness was placed in the chuck. The tool tip position measurement was recorded by the PC controller. Spindle speed was 513 rpm, corresponding to cutting speeds of 188.0 inch/sec. Feedrate was 0.0083 inch/rev. The depth of cut was 0.015 inches.

Two different cutting tests: step response and sine wave tracking, have been conducted. The purpose of the step response experiment was to find what tolerances could be maintained in the position of the tool tip relative to the fixture, as well as transient characteristic of step response of the tool tip servo. The purpose of the sine wave tracking experiment was to verify the ability of the system to follow references during cutting. This is similar to the objective of isolating the tool tip from boring bar vibration.

6.4.2 Cutting experiment results

As can be seen in Figure 6.4, 10μm step movement of tool tip during cutting has been applied. The controller maintained the tool tip position relative to the fixture to within $\pm 1\mu$m. The actuator could perform step changes in position in about 0.6 msec. The open loop response of the tool tip movement under the same cutting condition as the closed loop response shows steady state offset and larger overshoot.

The experimental results for sine wave tracking are shown in Figure 6.5. Like the result of step response, the error of closed loop control can be maintained within ±1μm, while open loop result shows larger errors. Table 6.1 shows the errors of both cutting experiments.

Since the step and sinusoidal reference result in variation of depth of cut, the experimental results shows that the actuator with designed controller performs variable depth of cut operation within 1μm error level, as well as position tracking control.

Table 6.1 Controller experimental results

Reference type		RMS error (μm)	Overshoot (μm)
Step	Closed loop	0.21 (steady state)	0.95
(10μm)	Open loop	0.23 (steady state)	2.33
		0.68 (offset)	
Sine wave	Closed loop	0.40	-
(Amplitude = 10μm)	Open loop	1.46	-
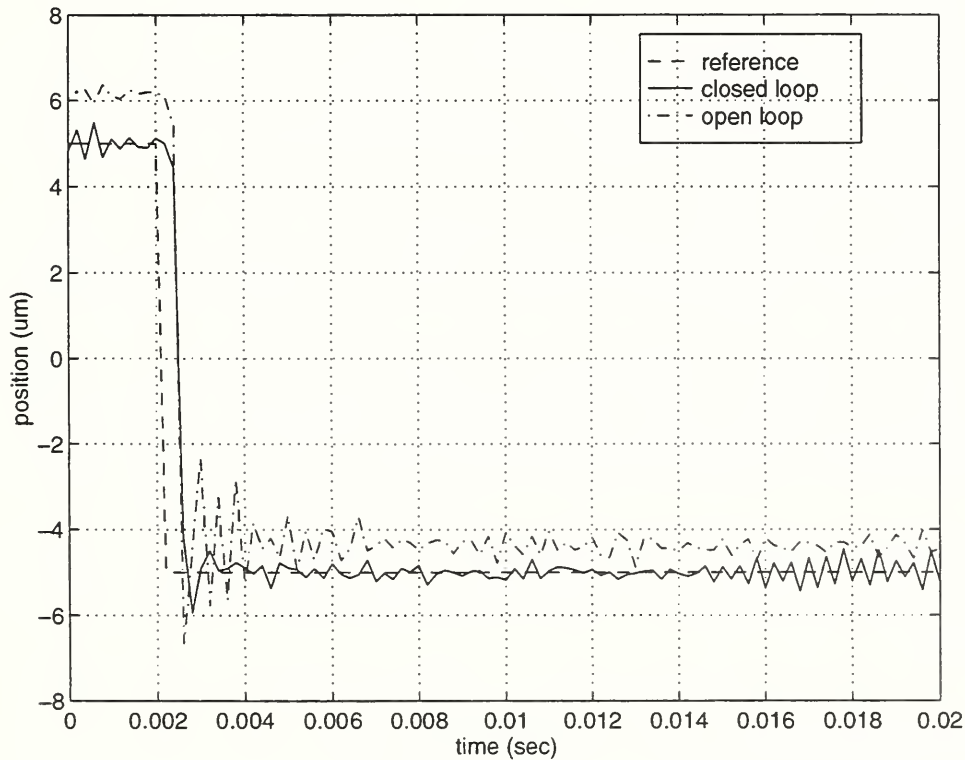
Figure 6.4 Step response
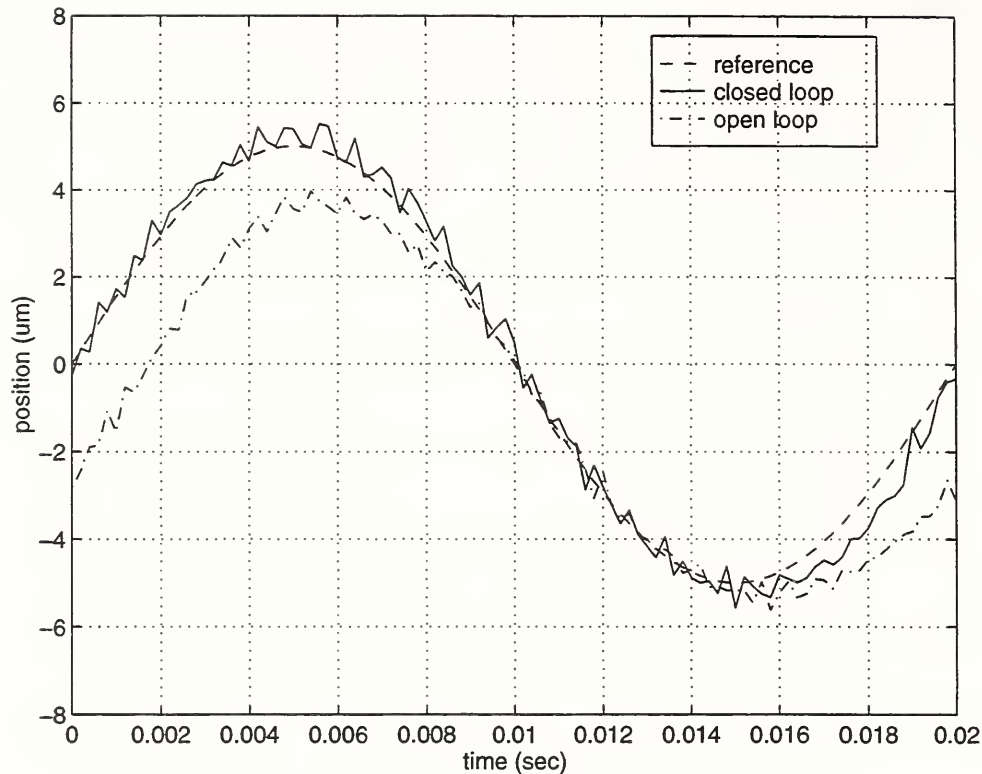
Figure 6.5 Sine response

VII. CONCLUSION

An agile line boring station does not exist today. Its development undertaken in this NIST/ATP sponsored program is a high-risk effort leading to a new, high-tech piece of equipment, relying on a number of existing and new technologies and their integrations into a working prototype - overall a technological quantum leap.

One of the critical bottlenecks in the development of an agile line boring station is the tooling. The paper focused on an effort to develop smart boring tool supporting agility at the station level by enabling automated tool change and providing an advanced on-line compensation mechanism.

It has been shown through simulations and experiments that the tool can successfully operate with the bandwidth of up to 50 Hz, providing the tool tip compensation range of ± 25 μm with a resolution of 0.2 μm. The compensation mechanism is fully balanced, assuring that the actuation does not affect the feedrate and the reference signal can be tracked within the error band of ±1 μm.

The paper reports on the work in progress. Future efforts will explore extensive cutting tests (including cutting of bi-metallic workpieces), performance assessment and system endurance evaluations.

ACKNOWLEDGEMENTS

References

Andreassen , L., 1995, "Means for Damping Vibrations for Example Self-generated Oscillations in Boring Bars and Similar," U.S. Pat. 5413318

Aronson, R. B., 1996, "The Whole Boring Business," Manufacturing Engineering, May, pp. 59-68

Fujiwara A., Yokoyama Y., 1991, "Tension-Giving Type Line Boring Machine," Jap. Pat. 3026407 A

Hanson, D.R., Tsao, T., 1996, "Compensation for Cutting Force Induced Bore Cylindricity Dynamic Errors - A Hybrid Repetitive Servo/Iterative Learning Process Feedback Control Approach", Japan/USA symposium on Flexible Automation, Vol. 2

Hanson, D.R., Tsao, T., 1994, "Development of a Fast Tool Servo for Variable-Depth-of-Cut Machining", ASME, DSC-Vol. 55-2, Dynamic Systems and Control: Vol. 2

Hara, Y., 1990, "A New Micro-Cutting Device with High Stiffness and Resolution", Ann. CIRP, 39/1

Hashimoto, H., et al., 1987, "Boring Process and Apparatus," U.S. Pat. 4,701,081

Johne, F., 1991, "Machine Tool Precision Boring Head," DE 3926025

Kim, K., Eman, K.F., Wu, S.M., 1987, "In-process control of cylindricity in boring operations", Journal of Engineering for Industry, Vol. 109, November, pp. 291-296

Kouno, E., 1982, "A Fast Response Piezoelectric Actuator for a Servo Correction of Systematic Errors in Precision Machining", Ann CIRP 33 (1), 369

Lee, D. G., 1988, "Manufacturing and testing of Chatter Free Boring Bars," CIRP Annals, Vol. 88/1, pp. 365-368

Liu, K.J., Rouch, K.E., 1993, "An Application of Active Dynamic Absorber to the Milling Process and its Experimental Verification", Vibration and Control of Mechanical Systems, DE-Vol. 61, ASME

Marra, M.A., et. al., 1995, "H_∞ Vibration Control for Machining Using Active Dynamic Absorber Technolgy", Proc. of the American Control Conference, June

Mason, F., 1997, "Smart Tools for Boring and Tapping," Manufacturing Engineering, May, pp. 84-93

Mason, W., 1935, "An Electromechanical Representation of a Piezo-Electric Crystal Used as a Transducer", Proceedings of the Institute of Radio Engineers

Odazaki, Y., 1990, "A Micro-Positioning Tool Post Using a Piezoelectric Actuator for Diamond Turning Machines", Precision Enginering, July, Vol. 12 No. 3

Paros, J., Weisbord, L., 1965, "How to Design Flexure Hinges", Machine Design, Nol T-27

Pasek, Z.J., et al., 1993, "Flexible Line Boring Project Report: Benchmarking Study", University of Michigan, Ann Arbor, September

Pasek, Z.J., et al., 1994, "Flexible Line Boring Project Report: New Concepts, Technologies and Their Evaluation", University of Michigan, Ann Arbor, August

Phillips, C. L. and H. T. Nagle, 1995, Digital Control System Analysis and Design, Prentice Hall

Pond, J. 1987, "New Boretools Combine Features of Single-Blade Reamer and Conventional Fineboring bar," Metal Working News, Vol. 14, No. 56, p. 30

Rivin, E. I., Kang, H., 1992, "Enhancement of dynamic stability of cantilever tooling structures", Intl. J.of Machine Tools & Manufacture; Vol. 32, No. 4, August, pp. 539-561

Rivin, E.I., 1987, "An extra-long cantilever boring bar with enhanced chatter resistance", Proc.of the 15th North American Manuf. Research Conf., pp. 447-452

Schechter, I., "Toolstrain,", 1993, informational materials from Pegard Productics

Slocum, A. H., Marsh, E. R., Smith, H. D., 1994, "A New Damper Design for Machine Tool Structures: the Replicated Internal Viscous Damper," Precision Engineering, v. 16, n. 3, July, pp. 174-183

Stephens, L. S., 1989, "An Active Dynamic Absorber for Suppressing Machining Chatter," M.S. thesis, University of Kentucky

Tajima, Y., 1988, "Machining Device Using Small Diameter But Long Line Boring and Machining Method," Jap. Pat. JP 63185505 A

Tanaka, H. et. al., 1994, "Active Chatter Suppression of Lensder Boring Bar Using Piezoelectric Actuators," JSME Intl. Journal, Series C, Vol. 37, No. 3, Sept. pp. 601-606

Tang, P., 1995, "Combined Piezoelectric-Hydraulic Actuator Based Active Vibration Control for Rotordynamic System", Journal of Vibration and Acoustics, Vol. 117, July

Tang, P., 1997, "Electromechanical Modeling of Hybrid Piezohydraulic Actuator System for Active Vibration Control", Journal of Dynamic Systems, Measurement, and Control, Vol. 119, ASME

Tewani, S. G. et al., 1988, "Control of Machining Chatter by Use of Active Elements," Proc. Symp. on Advanced Manufacturing, Univ. of Kentucky, Lexington, KY, pp. 31-36

Tewani, S. G., 1991, "Active Optimal Vibration Control Usign Dynamic Absorber," Proc. 1991 IEEE Intl. Conf. on Robotics & Automation, Sacramento, CA, pp.1182-1187

Tewani, S. G., Rouch, K. E., Walcott, B. L., 1991, "Finite element analysis of a boring bar using an active dynamic absorber for active vibration control"; ASME WAM, Atlanta, GA, AD v. 24. Publ by ASME, New York, NY, USA. p 179-188

Thomas, M.D., Knight, W.A, Sadek, M.M., 1970, "Comparative dynamic performance of boring bars", Proc. 11th Intll. Machine Tool Design and Research Conf., Birmingham

Ulsoy, A. G., Koren, Y., 1993, "Control of machining processes", Transactions of ASME Journal of Dynamic Systems, Measurement, and Control, Vol. 115, No. 2B, pp. 301-308

Yokoyama, Y., 1991, "Line Boring Machine," Jap. Pat. JP 3208507 A

Non-Circular Turning for Camshaft Machining -- Fast Response Actuators, Sensors, and Motion Control for Cam Profile Tracking

Tsu-Chin Tsao
University of Illinois at Urbana-Champaign
1308 West Main Street, Urbana, IL 61801
t-tsao@uiuc.edu

Introduction

Non-Circular Turning Process (NCTP) technology is based on the use of dynamic variable-depth-of-cut machining in a single-point turning environment to produce non-circular shapes using rapid actuation of the tool slide and high-speed, real-time digital signal processing and unique motion control algorithms. NCTP technology enables the generation of a wide variety of cam profiles in software and reduces or eliminates the amount of lobe grinding. This paper presents the critical motion control development for NCTP.

Requirement of NCTP Cam Profiling Motion Performance

<u>Velocity and Acceleration Requirement</u> For a particular cam lobe profile, the velocity coefficient v and the acceleration coefficient α represent the profile's maximum spatial velocity and acceleration, respectively. There are four cam types used in engines, namely, the roller tappet pushrod (R/PR) type, the overhead cam (OHC) type, the flat tappet pushrod (F/PR) type, and the direct acting (D/A) type. Among the numerous lobe shapes used in engines, the maximum numbers (velocity coefficient and the acceleration coefficient), for each of the four cam types, as shown in Table 1, represent the most demanding profiling motion. For these coefficient numbers, the cumulative volume percentage for the present aftermarket production and for the projected year 2000 Big 3 production are also listed in Table 1. As an example, the table shows that 89% of the cams, which include all the RP/R, OHC, and F/PR types, and some of the D/A type, produced by year 2000 will have an acceleration coefficient less than 0.0312 mm/deg^2.

For cam profile tracking motion in camshaft turning, the required maximum speed V and acceleration A of the tool profiling motion are related to the cam coefficients and spindle speed n by

$$V \text{ (M/sec.)} = v \text{ (mm/deg.)} \times n \text{ (rpm)} \times 0.006$$
$$A \text{ (M/sec2.)} = \alpha \text{ (mm/deg}^2.) \times n^2 \times 0.036.$$

The tool slide performance, which includes the actuator and the feedback sensor dynamic performance, is characterized by its maximum speed and acceleration capability. Figure 2 shows the actuator/sensor speed and acceleration requirement for turning the listed four points of the cam coefficients at 300, 500, 1000 and 2000 rpm, respectively. As an example of using this figure, to track all cam profiles at 1000 rpm spindle speed, the actuator and sensor hardware must have capability of generating motion at speed of 2.6 m/sec and acceleration of 7000 m/sec^2.

However, to track all cam profiles except for the Direct Acting type, the maximum velocity and acceleration are 1.5 m/sec and 1200 m/sec^2 respectively.

	F/PR type	OHC type	RP/R type	D/A type
Maximum v (mm/deg.)	0.2032	0.2286	0.2413	0.3937
Cumulative Volume Present U.S. Aftermarket	66%	77%	99%	100%
Cumulative Volume Year 2000 BIG 3	1%	56%	88%	100%
	RP/R type	OHC type	F/PR type	D/A type
Maximum α (mm/deg^2)	0.0097	0.0218	0.0312	0.1468
Cumulative Volume Present US Aftermarket	21%	32%	99%	100%
Cumulative Volume Year 2000 BIG 3	32%	88%	89%	100%

Table 1. Maximum Cam Velocity and Acceleration Coefficients of the Four Types Used in Engines and Their Cumulative Volume Percentage
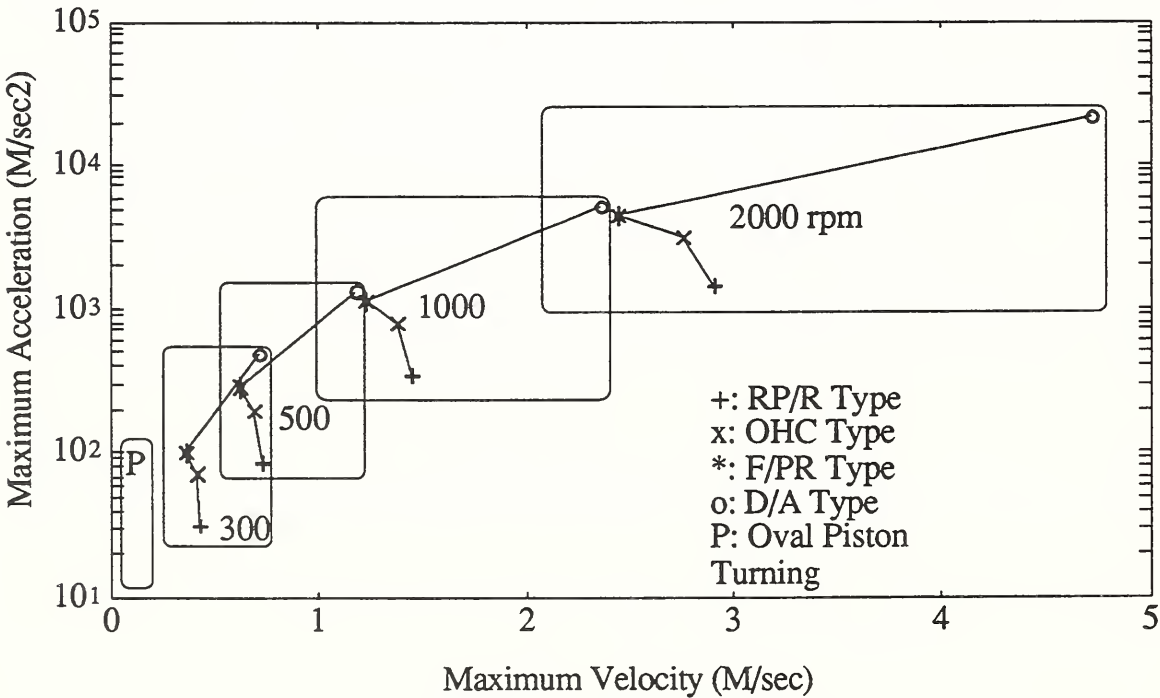
Figure 2   Maximum Speed and Acceleration Requirement for Cam Profiling

<u>Linear Actuator Bandwidth Requirement</u> The above speed and acceleration of cam profiling motion poses significant challenges to the present actuator and sensor technology. Furthermore, typical cam profiles require at least the first 20 Fourier series harmonics to approximate the original profile with less than 25 micron truncation errors. This implies that the servo control

should have tracking bandwidth higher than the 20th harmonic frequency, which easily goes beyond the performance of current industrial servo control technology. Consider that 600 rpm spindle speed renders 200 Hz. for the 20th harmonic frequency. Furthermore, each harmonic wave must be precisely tracked without magnitude or phase distortion. The camshaft timing performance requirement typically translates to cam profile error tolerance in the range of 25 to 75 microns. Unique custom designs of an electrohydraulic actuator and a voice coil linear motor made it possible to meet such unusual performance requirement.

Servo Control Technology

Most advanced commercial motion control hardware implements PID feedback with velocity and acceleration feedforward. This scheme cannot compensate for the high order system dynamics and high frequency signals. A position loop magnitude bandwidth of 100 Hz. is about the maximum that the state-of-the-art CNC servo controller can achieve, not to mention its significant phase distortion even well below 100 Hz. The high bandwidth precise tracking requirement coupled with actuator/drive dynamics and cutting force disturbances makes the servo control the most challenging technical barrier of camshaft turning.

Figure 3 shows the hardware system configuration of the servo control system. Due to the anticipated high sampling rate of the digital control implementation, a high performance digital signal processor has been used.
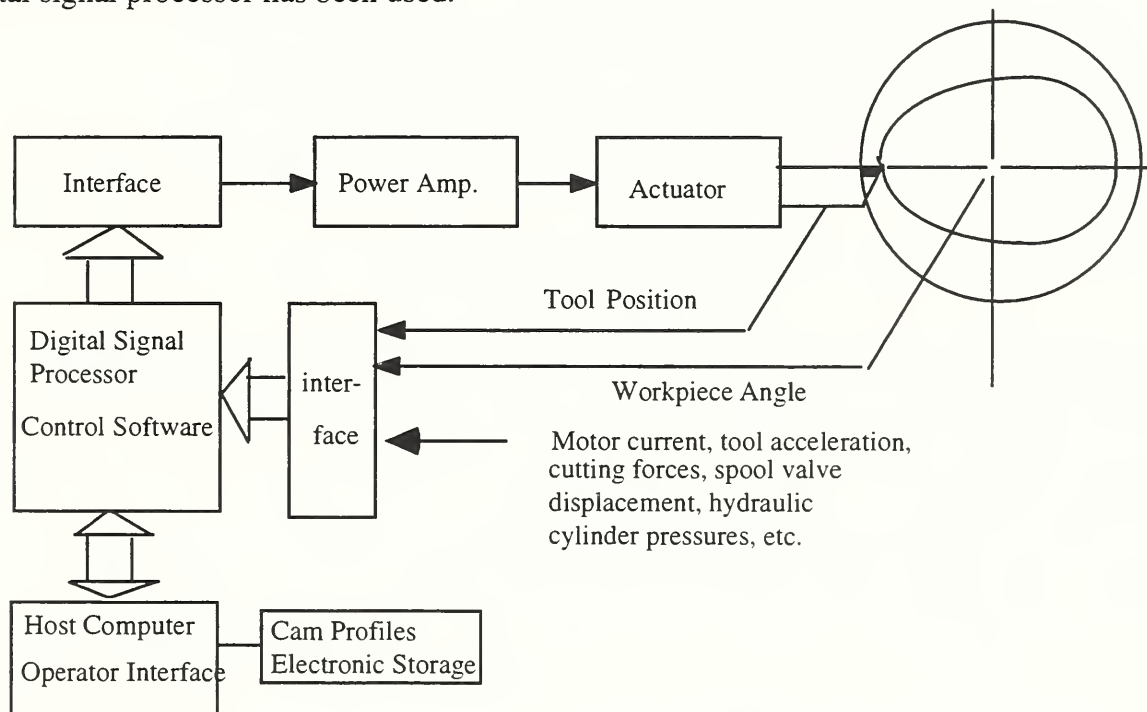

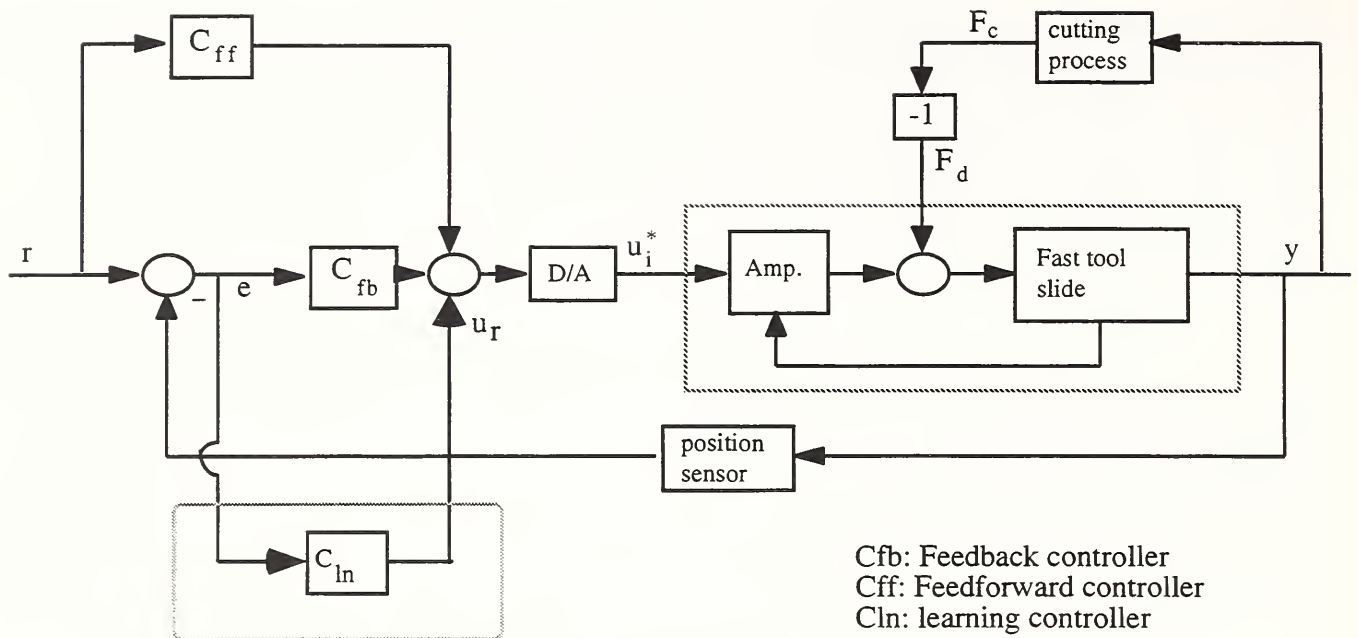**Figure 3 Control System Hardware Configuration**

Figure 4 Control System Block Diagram

The overall servo control system implemented for cam profile tracking consists of optimal robust feedback and feedforward, and a special cam profile learning tracking controllers, as shown in Figure 4. The performance objective is to achieve the best possible disturbance rejection/robust stability, and cam lobe tracking. This control component establishes maximal tool slide dynamic stiffness that ensures stable cutting and minimal vibration for adequate machined surface finish.

Experimental Results

Electrohydraulic Actuator The implementation of the servo control on the electrohydraulic actuators for a spindle speed of 600 rpm generally results tracking errors less than 25 microns. Figure 5 shows typical experimental results. The tracking errors while in machining are similar to those of air cutting.

The tracking performance at spindle speeds of 900, and 1200 rpm shown in Figure 6 and 7 respectively has maximum errors increased to approximately +/- 60 microns and +/- 200 micron respectively. The higher speed machining at its current status can be used for rough machining while the lower speed (600) rpm can be used for the cam lobe finishing, where errors must lie within 25 to 75 microns.

It is anticipated that every hydraulic servo valve and actuator will have variations in dynamic and static characteristics, which may make a fixed controller designed for a particular unit invalid for other units. Motivated by this an adaptive control system has been developed to account for this. Since there is only one actuator unit at hand, the adaptive control was tested by changing the hydraulic supply pressure. The fixed controller was designed for supply pressure of 3000 psi. Figure 8~10 shows the comparison between fixed and adaptive control at supply pressure of 3000psi, 2000 psi, and 1500 psi, respectively. As can be seen, the adaptive controller was able to maintain tracking performance at 1500 psi while the fixed controller could not maintain the performance.
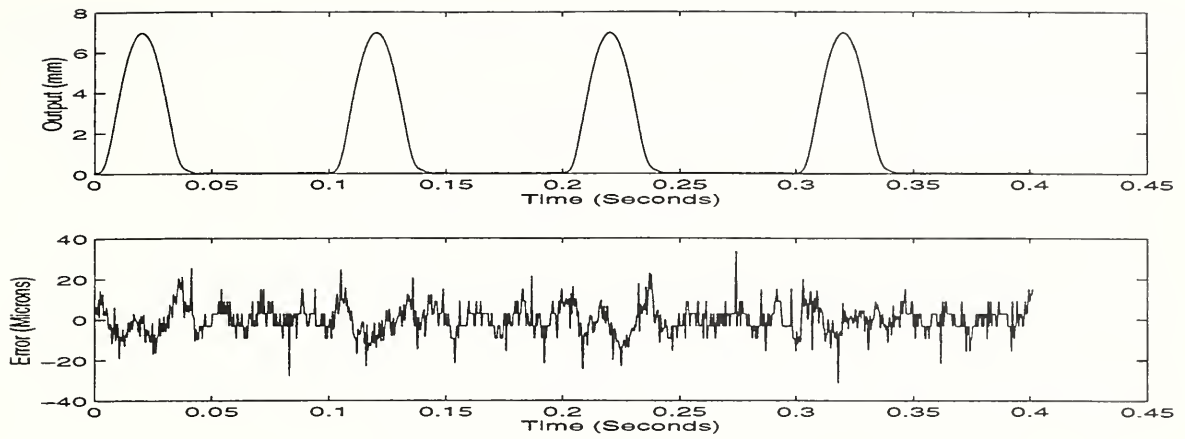
Fig. 5 Electrohydraulic Actuator Tracking Performance for Spindle Speed at 600 rpm
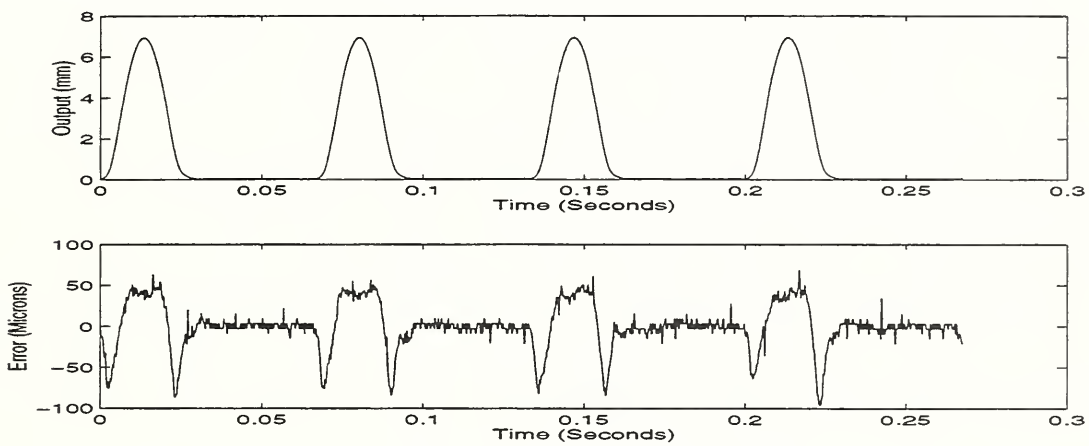
Fig. 6 Electrohydraulic Actuator Tracking Performance for Spindle Speed at 900 rpm
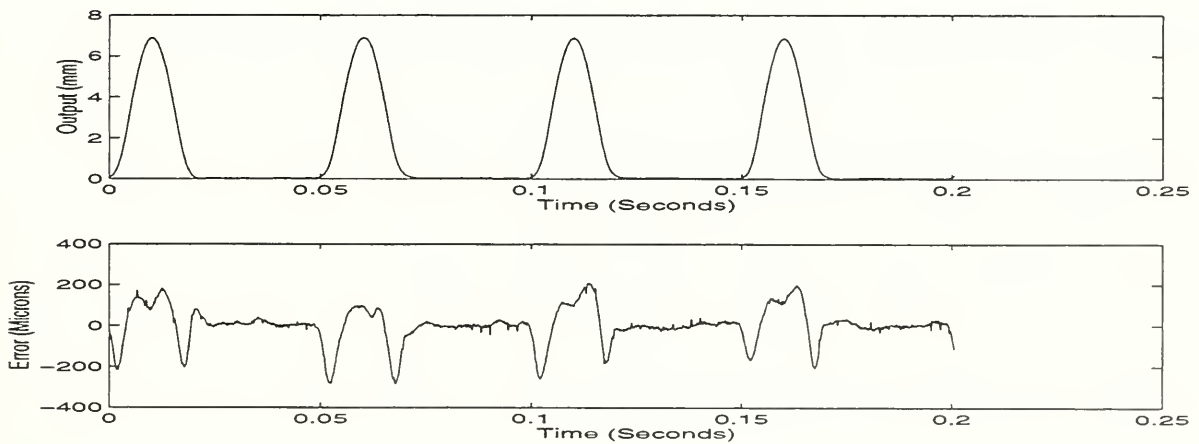
Fig. 7 Electrohydraulic Actuator Tracking Performance for Spindle Speed at 1200 rpm
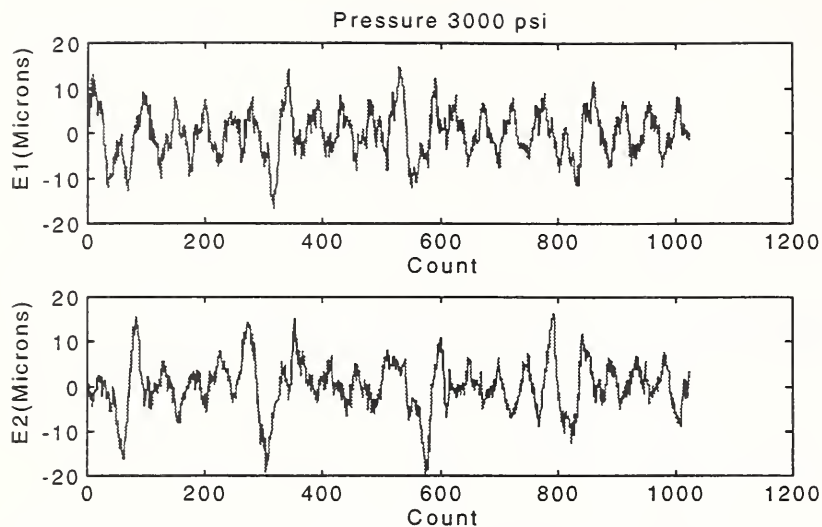
Fig. 8 Errors of Non-adaptive (top) and adaptive (bottom) repetitive controller (3000psi)
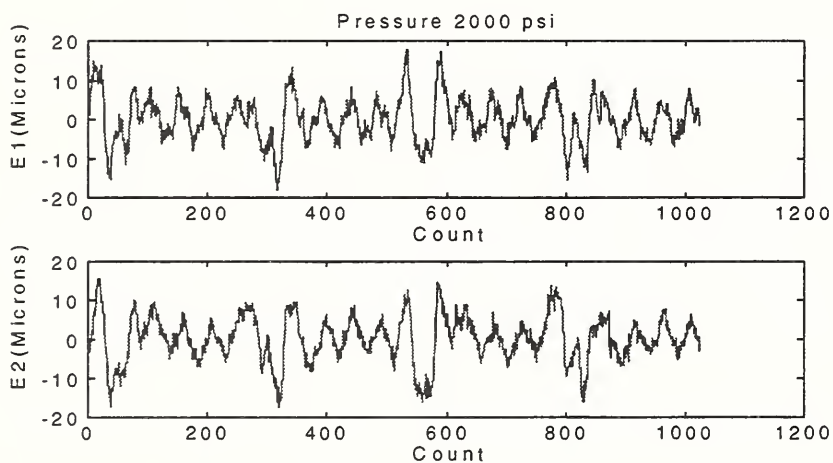
Fig. 9 Errors of Non-adaptive (top) and adaptive (bottom) repetitive controller (2000psi)
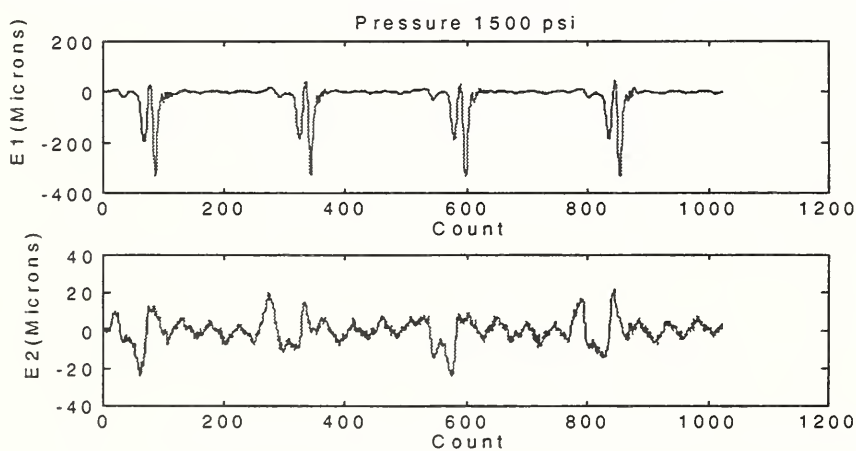
Fig. 10 Errors of Non-adaptive (top) and adaptive (bottom) repetitive controller (1500psi)

Voice Coil Linear Motor Experimental Results on the voice coil linear motor at 600 and 900 rpm show similar tracking performance. Figure 11 and 12 shows typical cam lobe tracking performance. Machining tests have not been completed . Work is underway to test whether it will have enough force and stiffness capabilities to maintain similar performance in machining.
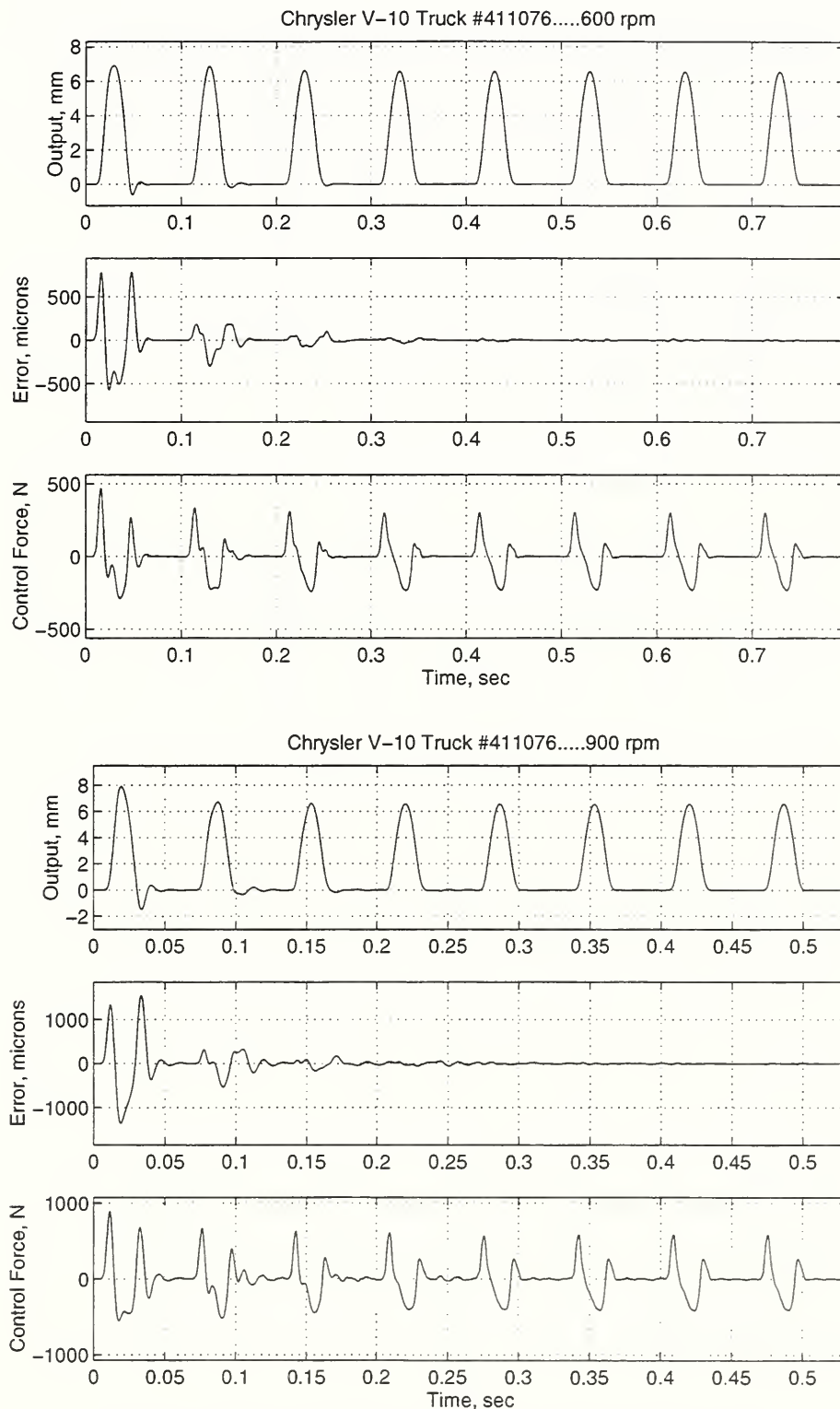
Figure 11 Voice Coil Motor Tracking Performance

P/N E-12497-.070866 Low Volume Prototype Cam.....600 rpm
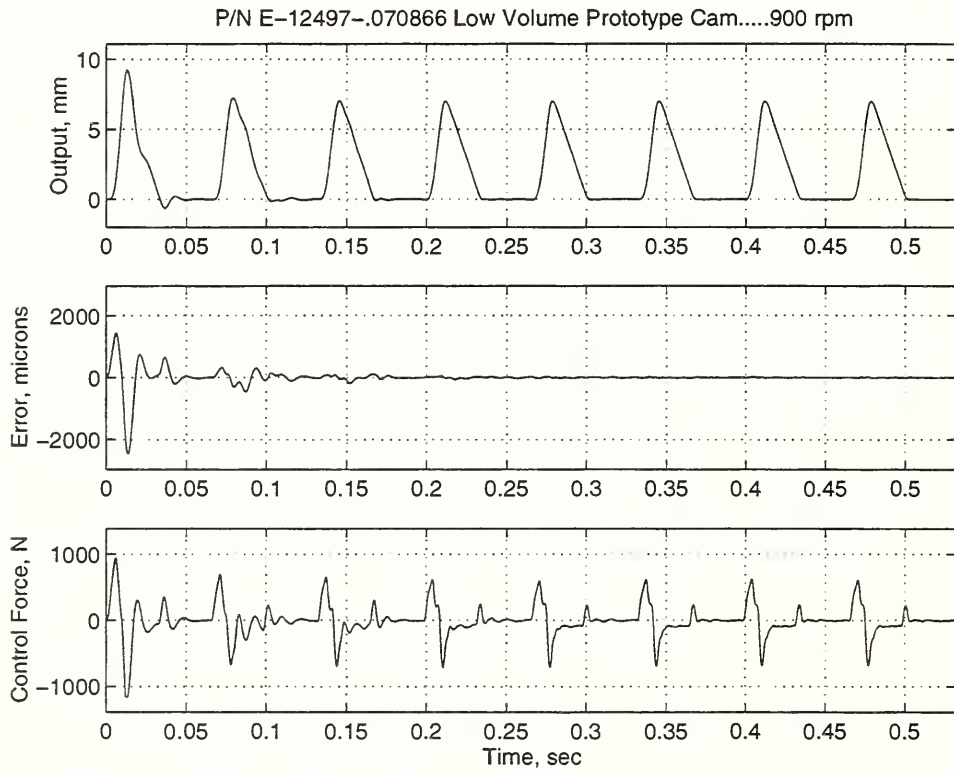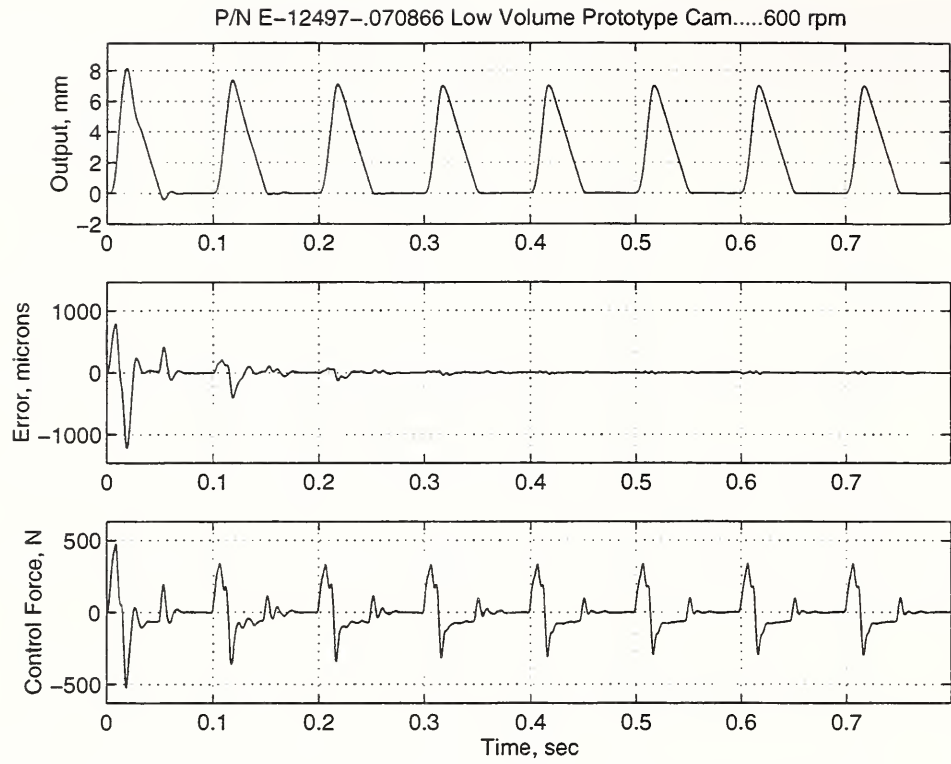
Figure 12 Voice Coil Motor Tracking Performance

Conclusion

Cam profile tracking control algorithms have been developed and implemented on electrohydraulic and voice coil motor actuators. The tracking errors for a spindle speed of 600 rpm have been consistently lower than 25 microns. Cam lobe machining of steel has been successfully carried out at 600 rpm using the electrohydraulic actuator. Present and future research effort focus on pushing the motion control performance envelope to higher spindle speeds. This project's industrial investigator, Saginaw Machine Systems is incorporating the motion control technology described herein for a prototype industrial camshaft turning machine for production testing at the end user, Camshaft Machine Company.

Acknowledgment

Development of a *Machining* *Diagnostic* System

William A. Kline
Vice President and Chief Technical Officer

R. Sriram
Research and Development Engineer

Montronix Inc.
Ann Arbor, MI

Richard E. DeVor
Professor and Director of NSF MT/AMRI

University of Illinois at Urbana-Champaign
Urbana, IL

1. Introduction: Overview and Goals of the Project.

In automotive machining applications, significant time and cost is consumed during periods of runoff of a new process or change-over of an existing production process. During these periods, the tool materials, tool geometries, and cutting conditions are adjusted to achieve the required levels of productivity and quality. Often, these adjustments are made by trial and error methods. These productivity and cost losses become more critical as the industry moves to smaller lot sizes and shorter development cycles for new model lines. A *machining diagnostic system* combining the elements of retrofittable sensors, portable data acquisition, process model based fault modeling and analysis is proposed as a solution to identify and diagnose these problems.

Traditional tool condition monitoring systems typically measure force, acoustic emission and other signals, compare the signals to pre-determined thresholds or patterns, and shut down the machine when failure is detected, to avoid a catastrophic event. While they provide significant economic and productivity benefits in many applications, a major shortcoming of these systems is that they contain no diagnostic element. The proposed machining diagnostic system contains a strong diagnostic element and is able to detect and isolate specific faults in rotating-tool machining processes. This paper provides a status report on the *'MD-System'*, a portable, integrated system for machining visualization and fault diagnosis.

The project is funded by the NIST-ATP under the Motor Vehicle Manufacturing Technology Program. The project began in September 1995 with a proposed four year time-frame, out of which the first two years is funded by NIST. The budget for the project is $1.2 million. The development of the system was organized into 13 different task areas grouped into three categories. Significant progress has been achieved in all task groups with exciting results in the following areas : ability to characterize fault condition in the process models, close correlation between retrofit and laboratory sensor data, and development of generic fault diagnosis indices and strategies.

The system is being developed in a collaborative effort between Montronix, Inc. and the University of Illinois at Urbana-Champaign (UIUC). The focus of Montronix has been in the development of the system hardware & software applications, system integration, and industrial beta-testing of prototype system concepts. The UIUC has focused on developing fault prediction and analysis capability for milling and drilling mechanistic process models, development & testing of indices via simulation and design of experiment, and the development of diagnostic strategies such as genetic algorithms and neural networks. From the initial concepts and proposal, our 'vision' of the project and system continues to evolve based on technical progress and feedback from the marketplace.

2.   A Prototype System for Process Monitoring and Fault Diagnosis.

A prototype system, the *MD-System*, is under development by Montronix, Inc. and the University of Illinois at Urbana-Champaign as the outcome of the project development effort. The system incorporates high-speed data sampling and signal processing, data visualization and recall, and computation of fault diagnosis information. The initial configuration of the system will be a portable, 'suitcase' system for diagnosis of shop floor problems.

2.1.   System Architecture.
The architecture of the **MD-System** is described in Figure (a) in Appendix A. A digital signal processor unit termed the Rotating Tool Pre-Processor (RTPP), and a Pentium PC are the main hardware elements of the system. The RTPP is equipped with two analog signal input channels and a synchronization input. The RTPP acquires data synchronously with tool revolution, performs on-line data analysis, and sends packets of data and diagnostic information to the PC at regular intervals via a serial link. An application running on the PC enables visualization of sensor data and logging of data and diagnostic information. In addition, the PC also acquires analog data from up to four channels directly. This approach of distributing functions between a DSP unit and the PC enables the system to combine the high-speed data acquisition and analysis features of the DSP, and the graphical capability of the PC.

As shown in Figure (b) in Appendix A, the features of the diagnostic system are designed to operate in a modular, layered architecture. Depending on the application, available inputs and process information, these modules are deployed in an on-line or off-line mode. At the lowest level, the system is a comprehensive, high-speed machining data visualization and recall system. Using high-speed DSP technology, the system performs signal processing and diagnostic index computations on-line. On top of these basic modules are features such as visual fault detection, empirical rule-based fault detection, and process model based fault detection. These features are available both for on-line fault detection, as well as comprehensive offline analysis that would yield valuable insights into the process. The modules can be employed for both short term process monitoring and failure detection as well as more long term trending and SPC based monitoring.

The application of the system is described in Table 1 in Appendix (a), starting with a particular machining process on the left, the various signal analysis methods used, to a defined cost, quality, or productivity benefit on the right. For different processes and types of problems,

different sensors and signal analysis techniques may be used. For example, the faults and analysis techniques for a drilling process may be completely different compared to a milling process.

The ability to relate machining fault conditions to tangible economic cost is critical to defining the customer benefit of the system. Our view is that 'faults' in the process amplify the peak forces or vibrations and thus have a direct impact on rate or tool wear, risk of tool breakage, deflection, and overall process variability. Preliminary results support this view, and further research is planned.

2.2. Data Acquisition and Signal Processing Techniques.

Synchronous Sampling and Averaging

Data acquisition in the **MD-System** is synchronous with the revolution of the cutting tool thus focusing on rotating tool applications such as milling and drilling. This synchronization is achieved by means of a proximity sensor or encoder signal that provides a number of pulses per tool revolution. Synchronization renders the data acquisition robust to changes in rotational speed. The acquisition and analysis can then be performed in a time frame normalized to tool rotational speed, i.e., a per-revolution time frame. This scheme also allows data averaging over multiple revolutions, attenuating asynchronous noise while preserving synchronous signal content.

FFT and Index Computation

The **MD-System** computes sets of diagnostic *indices* from the sampled data. Indices are statistical measures that characterize the cutting-process signal and show significant variations in the presence of faults in the process. These indices are based on the underlying structure of the metal-cutting process, and are useful in various cutting operations involving rotating tools. The indices are based both on the time as well as the frequency domains, and utilize factors such as the signal power content at various harmonics, signal peak information, etc. Based on field results, new indices are proposed and developed to capture and characterize specific fault conditions.

Fault Diagnosis

In addition to basic data and index display, an automated fault diagnosis strategy is required. During machining testing, vast quantities of data are rapidly collected. An automated scheme for fault screening and diagnosis is needed to sift through the data and highlight potential fault conditions. Various diagnostic strategies including such as genetic algorithms and neural network based methods are under development. These strategies are discussed in Sections 3-7.

Monitoring

In addition to diagnosing and locating faults, our customers have requested dedicated monitoring capability in the system. The monitoring capabilities envisioned include both short-term monitoring on the sensor, index, and fault data for detecting transient conditions in the process, and trend monitoring over days and weeks to identify longer term process variations.

2.3. Visualization Features.

A software application running on the laptop PC provides the **MD-System** with comprehensive and sophisticated data visualization and logging capabilities. The various features of the application are shown in Figure (c) in Appendix A and are discussed below.

One-Revolution Plot – Polar Plot

This window plots data corresponding to the latest tool revolution in a polar fashion. The polar plot emphasizes the action of individual teeth, and enables easy location of chipped or missing teeth. The amount of runout in the process is also easily judged from the eccentricity of the plot. The table at the bottom of the plot shows index values calculated from the displayed sensor data.

One Cycle Plot

The One Cycle Plot shows sensor or index values plotted versus cutting time for a single machining window. This plot enables observation of the trend of the sensor data and indices over the cycle duration.

For tracking the process, the system can sample and plot four 'low speed' and two 'high speed' sensor channels. The low speed channels are simply sampled and plotted without signal processing. This is useful for monitoring a simple motor power signal to 'time-stamp' the start, end, and trends of the process. In addition to sensor data, the calculated indices may be plotted to further study their behavior over the cutting cycle.

FFT Plot

The FFT of the data corresponding to the latest tool revolution is graphed with the frequency scale normalized to the tool rotational speed. This normalization makes it easy to pick out power content at important frequencies such as the spindle frequency or tooth-passing frequency.

Data Logging and Replay

Data acquired during a machining cycle is saved to disk along with all computed indices and FFT information. This data can then be played back in an off-line mode.

Fault Diagnosis and Monitoring

The fault diagnosis and monitoring features will be added as tools to be invoked and overlaid on the basic data displays. These features are presently under development.

Configuration Utility

The PC application also serves as a configuration front-end to the RTPP. Settings such as the diagnostic index list, specific tool information etc. can be set by the user and are communicated to the RTPP at the beginning of a machining cycle.

2.4. Case Study.

Test Setup

The prototype **MD-System** was used to conduct a series of tests involving monitoring of face-milling operations on an automotive cylinder block line. Three tools were used in the operation, and the tool details are shown in Table 2.

Table 2: Tooling Details

Tool Number	Location	Geometry	Speed (RPM)
1	T16	8" diameter; 14 teeth	1950
2	T17	4.82" diameter; 5+5 teeth	2021
3	T18	3" diameter; 12 teeth	2500

T16 was used to rough mill the cylinder head faces, T17 to rough mill the bearing cap seat and width, and T18 to rough mill various mounting bosses. A standard Montronix vibration sensor (VS-100) installed on the side of the spindle was used to collect data for all three tools. The operation cycles fell into two categories. In the first case, one insert was deliberately removed from T16, while the other tools were not altered. The second category of cycles were regular production runs. Example plots from the tests are shown in Figures (a) through (g) included in Appendix B.

Missing insert detection for Tool 1

Figure (a) shows the sensor data for one revolution of Tool 1, for a normal cycle. Spikes can be seen in the one-rev plot corresponding to the cutting action of each of the 14 inserts. The FFT plot shows that most of the signal energy is concentrated at the tooth-passing frequency and its first harmonic (14 & 28).

Figure (b) shows a revolution for the same tool, in a cycle where one insert had been deliberately removed from the tool. The missing insert is clearly seen in the one-rev plot, around the 9 o'clock position. The missing insert is shown by a drop in the Effective Tooth Number (ETN) index, from 64 to 53. The ETN, measured as a percentage, is designed to indicate the cutting efficiency of the tool. A "perfect" tool, with no faults, would result in an ETN approaching 100. Process faults such as runout and tool imperfections such as setting errors or chipping reduce the ETN value.

Breakage detection for Tool 1

Figure (c) shows a revolution immediately after three inserts of Tool 1 were chipped or broken upon re-entry of the tool into cut. The shape of the one-rev plot is remarkably different. There is no longer the periodic spike corresponding to the cutting action of an insert. The signal energy is instead shifted to the lower frequencies, as illustrated by the FFT plot and the Power Ratio index. The ETN has fallen sharply to 42, and the power ratio PR has fallen to the low value of 54. The changes in indices clearly indicate the insert breakage, which is not shown by the average signal level.

Figure (d) shows a revolution two cycles after the insert breakage. By this time, the average signal level has risen considerably.

Observations for Tool 2

Figure (e)[1] shows a sample revolution for Tool 2. It can be seen that most of the signal energy is concentrated at the tooth-passing frequency at 5. The peak vibration levels are relatively even, and there is no significant runout.

Observations for Tool 3

A revolution for Tool 3 is shown in Figure (f). It can be seen from Figure (f) that in the cutting action of Tool 3, there is significant energy content at the revolution frequency. This is also seen from the one-rev plot, which is slightly "off center", shifted downward towards 6 o'clock. This shift is characteristic of runout in the tool.

It was also noted that of the three cutters, tool 3 (T18) experienced the highest mean and peak vibration levels of the three cutters. In addition, due to the process, T18 experiences more periods of high vibration during one cycle than the other two cutters. A revolution with such high signal levels is shown in Figure (g). Figure 2 shows 'typical' mean and maximum vibration levels for the three cutters taken over several cycles. Cutter T18 experiences signal levels almost twice as high as T16 and T17.

Summary of Findings

Tool T16	No significant runout. Significant tooth to tooth variation in vibration level. May be due to insert setting.
Tool T17	No significant runout. Even levels of peak vibration.
Tool T18	Moderate runout. Significant tooth to tooth variation in vibration levels. High mean and peak vibration levels.

This case study was performed with an early version of the system and served to verify the performance of the data collection modules and contained no automated fault analysis. Additional case studies have been and will be conducted with the system.

---

[1] Figures (e), (f), and (g) incorrectly show 14 teeth on the one-rev plot. This is a limitation in the display module of the current software, and does not affect computation of indices.
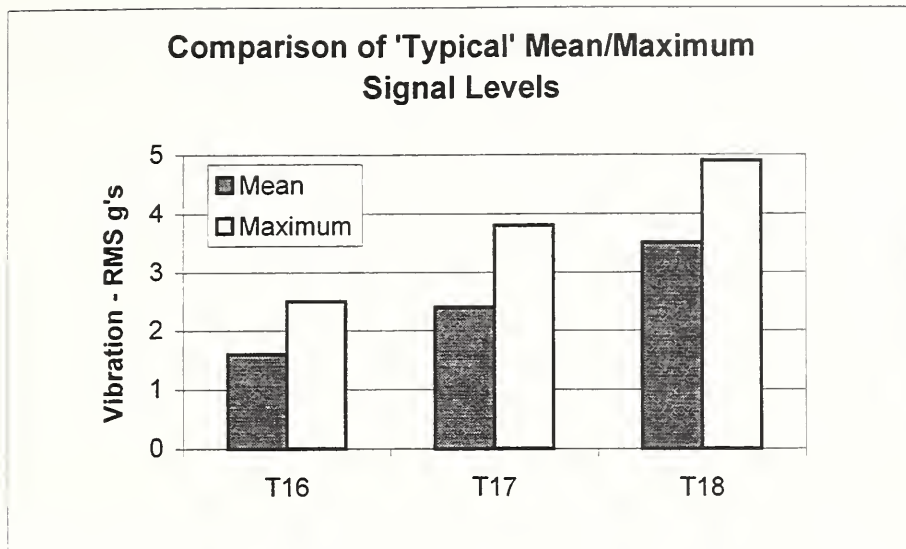
Comparison of 'Typical' Mean/Maximum Signal Levels

Figure 4: Signal Levels

3. A Model-Based Approach to Fault Detection and Diagnosis in End Milling.

Significant research has been published in the area of mechanistic process modeling in the last 15 years. This research has demonstrated the viability of the approach for prediction of cutting forces and a variety of related process measures such as tool or part deflection and finish. In this project, the mechanistic modeling approach has been extended to model the effects of typical tool and geometry faults with the goals of predicting the resulting cutting force behavior and developing fault diagnosis strategies. Particular focus has been placed on end milling, face milling, and drilling.

The cutting forces in end milling need to be estimated in order to plan the process with regards to choosing the cutting parameters and the fixturing scheme. Often, the cutting force in end milling differs from the nominal predicted value because of the presence of process faults in the system, such as, parallel axis offset runout, spindle tilt, chipped or broken flutes and flute deviation. The presence of such faults in the system cause uneven cutting and increase the peak cutting force. Therefore it is necessary to identify the presence of these faults and take corrective action. To understand the effect of these faults a comprehensive force model that can predict cutting force in the presence of faults is needed. This model can then be used to monitor the experimental cutting force signal and identify the process faults.

3.1. Mechanistic Model and Fault Prediction.

EMSIM is an end milling force prediction software based on the mechanistic cutting force model. In this model the cutting force is assumed to be proportional to the chip area. The mechanistic coefficients kn and μ which are the constants of proportionality relating the cutting force to the chip area are determined by calibration experiments for a specific tool and workpiece combination.

As shown in Figure 5 below, the cutter radius along the flute varies depending on the
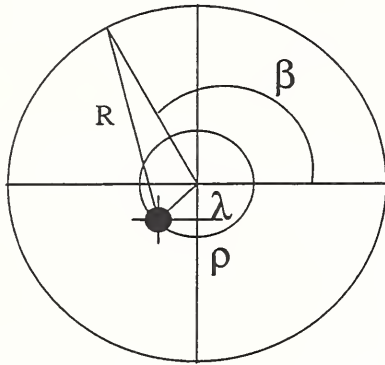
$$R(i,j,k) = RAD^2 + \rho^2 + 2\rho RAD \cos(\lambda - \beta)$$
$$\beta = (k-1)\pi/4 + z\tan(hx)/RAD$$
$$ct = \min(R(i,k) - R(i,k-1) + f_t \sin(\beta))$$

RAD

magnitude of the parallel axis offset runout (ρ) and the locating angle of this offset (λ). Because of this variation in the cutter radius the chip thickness (ct) fluctuates and causes uneven cutting forces. In the above equations *hx* refers to the helix angle and RAD is the nominal radius of the cutter. If a flute is chipped or broken it changes the chip load also, and a similar adjustment can be made for this as well. The elemental cutting force is obtained as:

$$dftq = kn(\cos(nr)\cos(hx) + \mu\cos(flow)\sin(nr)\cos(hx) + \sin(flow)\sin(hx))ctdz$$

$$dfrad = kn(-\sin(nr) + m\cos(flow)\cos(nr))ctdz$$

$$dflon = kn(-\cos(nr)\sin(hx) + m(-\sin(nr)\sin(hx)\cos(flow) + \cos(hx)\sin(flow))ctdz$$

Where, *nr* is the normal rake angle and *flow* is the chip flow angle.

3.2. Problem of Multiple Fault Detection.

In end milling, process faults can occur either singly or in concert. The following fault combinations are considered here: (1) only runout, (2) only chipping, and (3) both chipping and runout. Each fault condition changes the cutting force in a particular fashion. Suitable fault indices that can characterize the cutting force signal are needed to specifically identify the fault. Indices such as the harmonic power ratio, the effective tooth number, peak multipliers and the crest factor have been studied for their discriminating ability. For a no fault condition, the harmonic power is concentrated at the spindle frequency. In the presence of pure runout, power shifts to the spindle frequency. For pure chipping, harmonic power shifts to twice the spindle frequency. For a combination of chipping and runout, the trends in the fault indices are more complicated, and need either regression analysis, neural network based training or search algorithms.

A model based system for fault diagnosis utilizing EMSIM and LabView is being developed at the University of Illinois at Urbana-Champaign. This system takes in collected experimental data and generates fault indices for this data. Through an interactive interface, users can run EMSIM for specified cutting conditions. Fault parameters can be varied so that the experimental and EMSIM generated signals match both in a visual sense as well as by matching of the fault indices. This system is shown in Figures (h) and (i) in Appendix A. Figure (h) shows a fault condition of spindle runout. The value of 0.0004in for the magnitude

of the runout and 268 degrees for the locating angle match up the experimental and EMSIM generated signal well. The harmonic power ratio at H4 (tooth passing) starts to drop from 1 to 0.98 for the experimental data and from 1 to 0.95 for EMSIM simulated data. H1 (spindle) increases to 0.01 for the experiment and 0.03 for EMSIM. The ETN matches very well with a value of 3.66 for the experiment and 3.65 for EMSIM. Figure (i) shows the flute breakage fault condition. In this case the cutting force is close to zero for one of the flutes. The H4 valued drops down to 0.59 for the experiment and 0.44 for EMSIM. ETN is 2.45 for the experiment and 2.21 for EMSIM.

Simulation studies are ongoing for the automatic detection of process faults facilitated by observing trends in faults indices and fitting models to predict faults. Development of new fault indices such as conic section fits to the peaks in the cutting force signal is also underway.

3.3. Case Study.

Prediction of the fault pattern and the magnitude of the fault constitute complete fault detection. This requires identifying trends in the fault indices. Table 3 shows the variation in fault indices with increasing values of runout for a 4 fluted end mill. No chipping is present. LAN is the locating angle and Ro is the magnitude of the runout. The cutting conditions are axial depth = 0.3in, Radial depth = 0.2in, RPM = 1000 and feedrate = 0.004 in/flute. With an increase in the magnitude of the runout, the peak cutting forces increase, the value of ETN drops, the CF value increases, indicating uneven cutting. The power at the fourth harmonic drops and the power at the first harmonics starts to rise. These values can be used to predict the value of the runout present in the system.

Table 3. Trends in Fault Indices for Progressive Runout

Ro	LAN	Max(lb)	ETN	CF	PM1	PM2	PM3	PM4	H1	H2	H3	H4
0.0004in	0	170.7	3.57	2.27	0.879	0.926	1.12	1.075	4.74	0.01	1.72	89.51
0.0006in	0	179.54	3.39	2.41	0.818	0.889	1.18	1.113	9.2	0.02	3.66	83.32
0.0008in	0	188.24	3.23	2.55	0.755	0.852	1.24	1.152	13.98	0.05	6.05	76.27
0.001in	0	196.85	3.07	2.69	0.692	0.815	1.301	1.192	18.61	0.08	8.72	69.01

4. Retrofit Sensors and Fault Detection Indices: An Experimental Study.

In order to investigate the ability of different indices to diagnose faults in the end milling process, a set of experiments was designed and run. These experiments served several purposes.

- They created an experimental database that could be used by researchers to validate their own work.
- The use of a variety of different sensors including laboratory dynamometers and retrofit force and vibration sensors for fault diagnosis was investigated.
- The sensitivity and variability of various fault diagnosis indices were studied.

4.1. Design of Experiment.

The experimental design that was used for this study was a five variable central composite design. The faults of parallel axis offset runout and flute chipping/breakage were investigated under varying cutting conditions. Each variable in the design could take on five discrete levels as shown in Table 4. The variables used are normalized parameters developed to filter out the effects of cutter size. TW is the engaged tooth width, W is the cutting window width, and L is the length of a quarter circumference of the end mill cutter. These three parameters are calculated from the key cutting condition parameters including axial depth (AD), radial depth (RD), cutter radius (R), and the number of flutes on the cutter (NT) according to the following relations.

$$TW = AD \times \tan\alpha_{hx}$$
$$W = R \times \cos^{-1}(1 - \frac{RD}{R})$$
$$L = \frac{2\pi R}{NT}$$

Chipping was normalized with respect to TW/L and runout was normalized with the feedrate. The coded values of the cutting conditions for each of the 27 tests in the central composite design are shown in Table 5. All tests were run with a 0.75 inch diameter uncoated four flute end mill with a 30° helix angle and a 10° radial rake angle. Down milling at a cutting speed of 1000 RPM was used for all tests. The material being machined was 2024-T351 aluminum.

Table 4: Coded values of independent variables.

	-2	-1	0	1	2
TW/L	0.196	0.245	0.294	0.343	0.392
W/L	0.476	0.59	0.696	0.784	0.872
Feedrate (in. per flute)	0.002	0.003	0.004	0.005	0.006
Chipping/breakage (in.)	0	0.68*R*T W/L	1.36*R*T W/L	2.04*R*T W/L	2.72*R*TW /L
Runout (in.)	0	0.1*FT	0.2*FT	0.3*FT	0.4*FT

Table 5: Cutting conditions for central composite design.

Test	TW/L	W/L	Feed	CH	RO	Test	TW/L	W/L	Feed	CH	RO
1	-1	-1	-1	-1	1	15	-1	1	1	1	-1
2	1	-1	-1	-1	-1	16	1	1	1	1	1
3	-1	1	-1	-1	-1	17	-2	0	0	0	0
4	1	1	-1	-1	1	18	2	0	0	0	0
5	-1	-1	1	-1	-1	19	0	-2	0	0	0
6	1	-1	1	-1	1	20	0	2	0	0	0
7	-1	1	1	-1	1	21	0	0	-2	0	0
8	1	1	1	-1	-1	22	0	0	2	0	0
9	-1	-1	-1	1	-1	23	0	0	0	-2	0
10	1	-1	-1	1	1	24	0	0	0	2	0
11	-1	1	-1	1	1	25	0	0	0	0	-2
12	1	1	-1	1	-1	26	0	0	0	0	2
13	-1	-1	1	1	1	27	0	0	0	0	0
14	1	-1	1	1	-1						

Chipping was introduced to the end mill by grinding one flute on the cutter to the proper dimension. Runout was introduced by placing shims between the cutter and the tool holder. The tests were then run on a Bridgeport Discovery 300 machining center and data was collected using a National Instrument data acquisition board. The encoder signal from the spindle was used to time the data acquisition and resulted in 500 samples being taken each revolution. One hundred revolutions of data were collected for each test shown in Table 2.

Several different sensors were installed on the machine and all signals were sampled at the same rate. These sensors included a Kistler 3-axis force dynamometer, a Montronix Retrobolt sonsor located on the spindle cap, a Retrobolt sensor on a fixturing component, a power sensor on the spindle motor, a vibration sensor on the spindle, and an acoustic emission sensor on the workpiece.

The **RetroBolt** sensor (patent pending) is a single axis force sensor developed by Montronix providing the ability to measure resultant cutting forces. The sensor is easily installed and retrofit to existing machines by simply removing a load bearing bolt, such as a spindle cap bolt on a machining center or a turret hold down bolt on a lathe, and installing the Retrobolt sensor as a washer under a slightly longer bolt and re-installing. Installation normally takes a matter of minutes, requires no significant machine modifications, and has little or no effect on machine stiffness or dynamics.

4.2. Sensitivity Analysis.
So far, six different indices have been studied to compare the force data from the dynamometer and Retrobolts. These indices are the crest factor, the effective tooth number (ETN), and the percentage of power contained at each of the first four harmonic powers where the first harmonic power corresponds to the spindle frequency. If A is the data array for one revolution, p is the number of peaks/flutes, P is the array of force values at the peaks, and H_i is

the magnitude of the power spectrum at the ith harmonic then these indices are calculated as follows:

$$\text{Crest Factor} = \frac{\max(A) - \min(A)}{\text{mean}(A)}$$

$$\text{Effective Tooth Number} = \frac{100}{p} \times \sum_{i=1}^{p} \frac{P_i}{\max(P)}$$

$$\text{Harmonic Power}_i = \frac{H_i}{\sum_{i=1}^{4} H_i} .$$

In order to clearly see how sensitive the indices were to the faults and cutting conditions, the first sixteen tests from the design were analyzed. These sixteen tests correspond to a fractional factorial design. Effect estimates for each index were calculated for both the dynamometer force data and the data from the Retrobolt on the spindle cap. Normal probability plots of the effect estimates were made to determine which effects were significant. For the dynamometer force data, the following trends were observed. The crest factor increased as chipping or runout increased, but an increase in W/L caused a decrease. The tooth efficiency decreased as chipping or runout increased. The power at the first harmonic increased as chipping increased, but an increase in W/L also caused an increase. The second harmonic power increased as chipping increased. The third harmonic power increased as chipping or runout increased. The fourth harmonic power decreased as chipping or runout increased, but an increase in W/L caused a decrease also. The Retrobolt force data from the spindle cap corresponded well with the dynamometer force signal while the Retrobolt signal from the machine table was not as good. For this reason, only the sensitivity of the Retrobolt indices on the spindle cap were investigated. The effect estimates for the Retrobolt data corresponded very well with the effect estimates obtained from the dynamometer cutting forces for the frequency domain indices, but the time domain indices behaved differently for the two signals. This can be explained by drift seen in the Retrobolt force data.

4.3. Conclusions.
The experimental work that has been done has yielded much information pertaining to the detection of faults in the end milling process. Some specific conclusions are listed below.

1. The indices that are currently being used are very effective at detecting chipping of a cutter flute, especially the second harmonic power. The indices were also affected by parallel axis offset runout, but no index was affected exclusively by runout.
2. The data has shown that it may be possible to use a Retrobolt force signal for fault detection rather than a much larger, more expensive dynamometer. This is especially true when indices from the frequency domain are being used.
3. The data has shown that the indices have very little variation both within a single pass and between consecutive passes. For this reason, if the true variability of the indices is to be estimated, true replicates should be used rather than just separate cutting passes. As a result of this, further experiments are being performed in this stage of the research.

Other indices will need to be investigated to see if they are more robust at detecting process faults. These indices could come from the cutting force signature or from the other sensor signals that have been collected. Specifically, using wavelets as a diagnostic tool is being investigated along with indices pertaining to the vibration sensor signal.

5. Use of Neural Networks and Genetic Algorithms for Fault Detection and Diagnosis.

5.1. Use of Neural Networks.

A neural network (NN) is defined by an architecture of highly interconnected computational units (neurons) to simulate the relationship between the feature inputs and response outputs. The selection of feature inputs and the design of training vectors are essential to the success of neural networks, which use learning algorithms to adjust the values of weights and biases in the neural network until certain criteria are met (such as the least mean square error in backpropagation network). The use of neural networks has been reported for single fault detection such as tool breakage (Tansel et al., 1995) and tool wear (Hong et al., 1996; Li and Elbestawi, 1996). In these works, dedicated experimental force/vibration signals were needed to train the neural network. However, for the general purpose of fault diagnosis, it is prohibitively difficult to obtain a large experimental database under all possible combinations of faults.

The strength of the mechanistic force models developed by UIUC lies in their capability to predict force signals accurately under the conditions of common process faults. With the help of process models, the training for neural networks can be done off-line using the simulation data alone, as shown in Fig. 6. The trained neural network can then be used on-line or off-line to do the fault diagnosis based on the input of the monitored force signal.
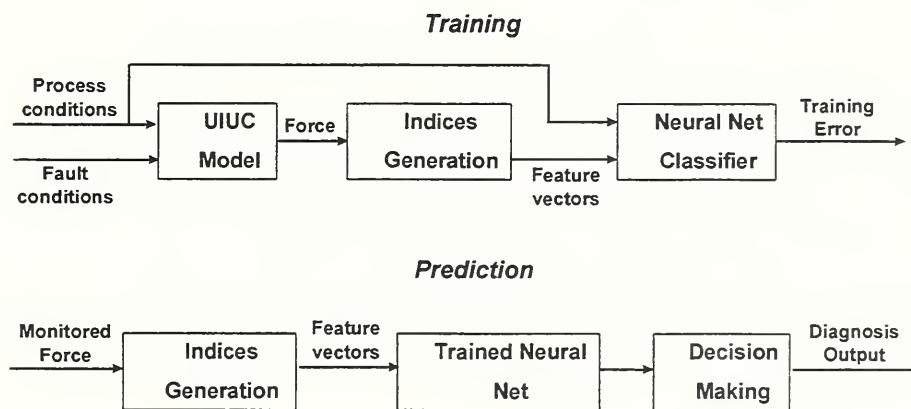
Fig. 6. Fault Diagnosis Using Neural Network

5.2. Use of Genetic Algorithms.

The genetic algorithm (GA) is an adaptive search algorithm that transforms a set (population) of mathematical objects (typically fixed-length binary character strings), each with an associated fitness value, into a new set (new generation of the population) of offspring objects, using operations based in the mechanics of natural genetics and natural selection. A model-

based GA diagnostic system is illustrated in Fig. 7. The GA takes a monitored force signal and evolves a population of strings containing fault estimates using a process model (such as EMSIM) until it converges on a matching set of conditions (i.e., the difference between the simulation force signal and the monitored force signal has been minimized). In the recent work (Zhu et al., 1997), a three-operator simple genetic algorithm (Goldberg, 1989) has been used to implement this diagnostic system. The validation results using the EMSIM-based simulation data have demonstrated the capability and robustness of GA to the recognition of chipping, breakage, and parallel axis offset runout.
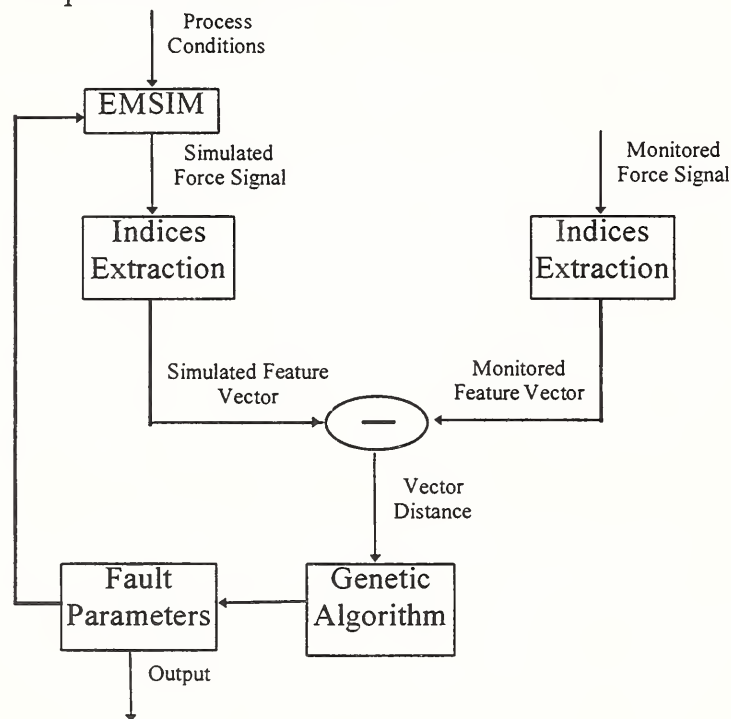
Fig. 7. Fault Diagnosis Using Genetic Algorithm

5.3. Case Study for the End Milling Process.

Two case studies are examined here to demonstrate the performance of the model-based GA diagnostic system. The experimental force signals measured by the dynamometer are from (Kim, 1996). In the tests, a coated HSS 4-flute end mill with 0.75 inches in diameter, 30 degrees in helix angle and 11 degrees in radial rake angle, was used to down mill the 7075 aluminum workpiece. The cutting speed was 1000rpm and a total number of 240 data samples were collected for every spindle revolution. The faults like chipping/breakage and parallel axis offset runout were introduced artificially by removing some material from the cutting tooth and inserting shims in the tool adapter.

The micro-genetic algorithm (MGA) was used to test the diagnostic system due to its small population size (usually 4-5) and the ability to reach quick results with much fewer function evaluations than simple genetic algorithms (Krishnakumar, 1989). One revolution of force data was compressed by wavelet transform four times and the resulting 15 wavelet coefficients were formed into the feature vector. The wavelet transform was chosen because it has a good resolution in both time and frequency domains and it is more compact and typically faster than Fourier transform (Daubechies, 1990).

Case Study 1: RD = 0.3", AD = 0.2", FT = 0.006", Tooth#3 Chipped by 0.05"
In this case, the GA diagnosed exactly that the tooth#3 was chipped with the magnitude of 0.05947". At the same time, the GA also detected a very small parallel axis offset runout with the magnitude of 0.00014" and the locating angle of 260.7°. In Fig. 8, the actual force data, simulated force data using designed fault values, and GA-searched force data are plotted together. It can be seen that although some mismatch exists between the actual force signal and the simulated force signal, the GA did find the optimum solution by minimizing the difference between the actual force signal and GA-searched force signal.

Case Study 2: RD = 0.3", AD = 0.2", FT = 0.006", Parallel Axis Offset Runout by 0.001"
In this case, the GA diagnosed a parallel axis offset runout with the magnitude of 0.0009" and the locating angle of 298.3°. At the same time, the GA also detected a small chipping on the tooth#2 with the magnitude of 0.02051". It is shown in Fig. 9 that the GA-searched force signal also matches well with the actual force signal in this case.
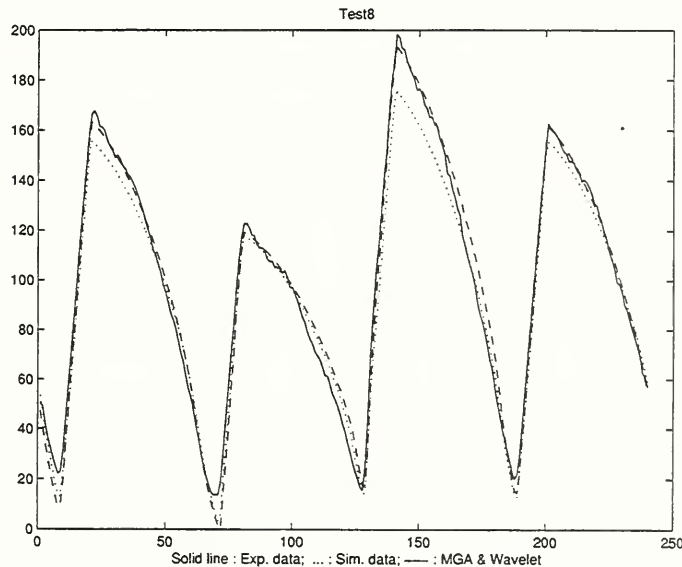

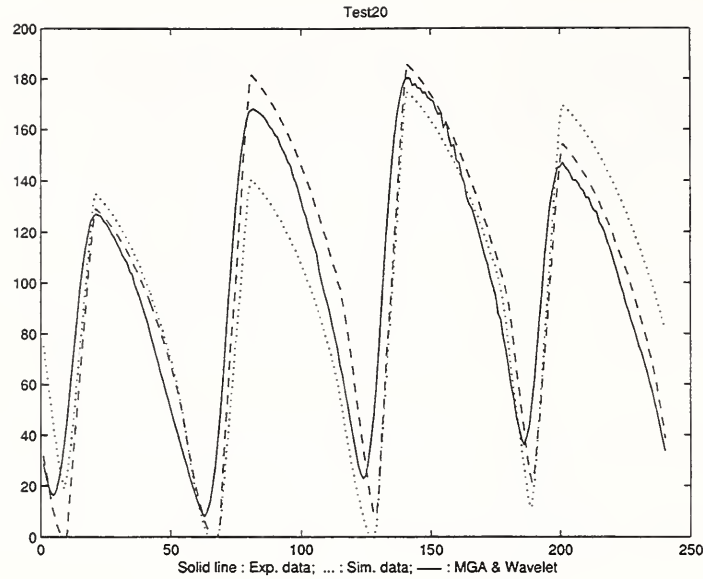**Fig. 8. Case Study 1 - One Tooth Chipping**

Fig. 9. Case Study 2 - Runout

6. Model-Based Fault Detection and Diagnosis in Face Milling.

Run-out is a process fault that is a characteristic of multi-point cutting operations such as face milling. Excessive run-out can cause some inserts not to cut at all and this can lead to a very uneven pattern in the chip load faced by the cutter. The use of high-speeds in milling operations has made the problems associated with run-out very significant. An increase in tool rotational speed increases the centrifugal forces on the inserts and causes dynamic variations in the run-out values. Patterns in the run-out variations can identify problems occurring due to spindle eccentricity and offset.

A model based run-out estimation algorithm has been developed by Jayaram (1997). The algorithm predicts the run-out values when machining workpieces having complicated surface discontinuities such as slots. However, the algorithm can only handle workpieces whose geometry does not change during the cut, i.e., the geometry of the workpiece faced by an insert during one cutter revolution remains the same throughout the cutter path. This is a major limitation because in practice surface interruptions have complicated shapes and the cutter faces changing workpiece geometry during the cut. A run-out estimation algorithm for the face milling process has been recently developed by Akshay (1997) . This new algorithm can take into account the effect of complex workpieces that have varying surface discontinuities such as holes and slots. It also models the effect of cutting speed, the cutter geometry that includes the lead angle, the radial rake angle and the axial rake angle, and the nose radius of the cutting inserts. The validity of the algorithm has been verified with the help of simulations as well as experimental results.

6.1. Estimation of Run-out.

The runout estimation algorithm developed by Akshay (1997) is presented here. The number of unknowns (runout values) is equal to the number of inserts, N. The radial runout values can be obtained from the following equation.

80

$$
\begin{Bmatrix}
1 & 1 & \cdots & 1 \\
R_1^y(j\omega_s) & R_2^y(j\omega_s) & \cdots & R_N^y(j\omega_s) \\
\vdots & \vdots & \vdots & \vdots \\
R_1^y((N-1)j\omega_s) & R_2^y((N-1)j\omega_s) & & R_N^y((N-1)j\omega_s)
\end{Bmatrix}
\times
\begin{Bmatrix}
\delta\varepsilon_1 \\
\delta\varepsilon_2 \\
\vdots \\
\delta\varepsilon_N
\end{Bmatrix}
=
\begin{Bmatrix}
0 \\
F_y^{total}(j\omega_s) - N_y(j\omega_s) \\
\vdots \\
F_y^{total}((N-1)j\omega_s) - N_y((N-1)j\omega_s)
\end{Bmatrix}
$$

Where R_i^y (jw_s) is the amplitude high harmonic of Y cutting force component due to run-out at the spindle frequency in the ith tooth of the cutter. N_y (jw_s) is the amplitude high harmonic, at the spindle frequency, of Y cutting force component acting on the cutter in the absence of any run-out. F_y^{total} is the actual total cutting force acting on the cutter in the y-direction in the presence of the faults.

The above equation has an important physical significance. Since N_y (t) is the total force on the cutter in the absence of radial run-out faults and F_y^{total} (t) is the actual total force on the cutter in the y direction in the presence of faults, the right hand side of the equation represents the subtraction of the nominal value from the true/actual value. What is left is the residual signal that contains the information about the run-out faults.

It should be also noticed that the difference between the constant workpiece geometry case and the varying geometry of the surface discontinuities is that N_y in the former case has peaks only at the tooth passing frequencies whereas the N_y in the latter case has values at the spindle harmonics also.

6.2. Optimized Run-out Estimation Algorithm.

Since the energy content in the first $(N-1)$ spindle harmonics of the signal is too small to be able to represent the signal properly, a method needs to be developed to be able to account for the rest of the energy content present in the rest of the spindle harmonics. Since there are only N unknowns, the only way to be able to account for the rest of the harmonics is to use a least square optimization routine. A constrained least square optimization as given below is used

$$
\underset{\delta\varepsilon_i}{Min}\frac{1}{2}\left[Norm(A*\{\delta\varepsilon_i\}-b)^2\right]
$$

$$
subject \quad to \quad \sum_i \delta\varepsilon_i = 0 \quad and \quad -f_t \le \delta\varepsilon_i \le f_t
$$

Where

$$
A = \begin{Bmatrix}
R_1^y(j\omega_s) & R_2^y(j\omega_s) & \cdots & R_N^y(j\omega_s) \\
\vdots & \vdots & \vdots & \vdots \\
R_1^y(Nj\omega_s) & R_2^y(Nj\omega_s) & \cdots & R_N^y(Nj\omega_s) \\
\vdots & \vdots & \vdots & \vdots
\end{Bmatrix}
\qquad
b = \begin{Bmatrix}
F_y^{total}(j\omega_s) - N_y(j\omega_s) \\
\vdots \\
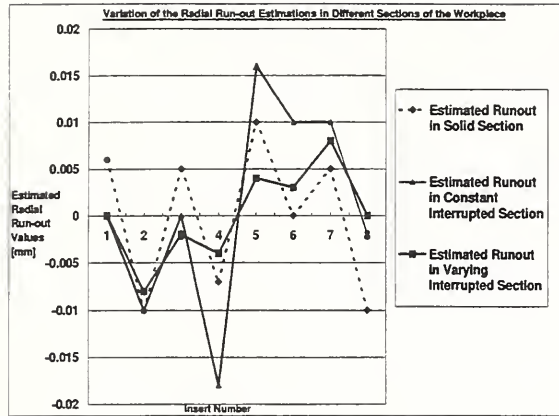F_y^{total}(Nj\omega_s) - N_y(Nj\omega_s) \\
\vdots
\end{Bmatrix}
$$

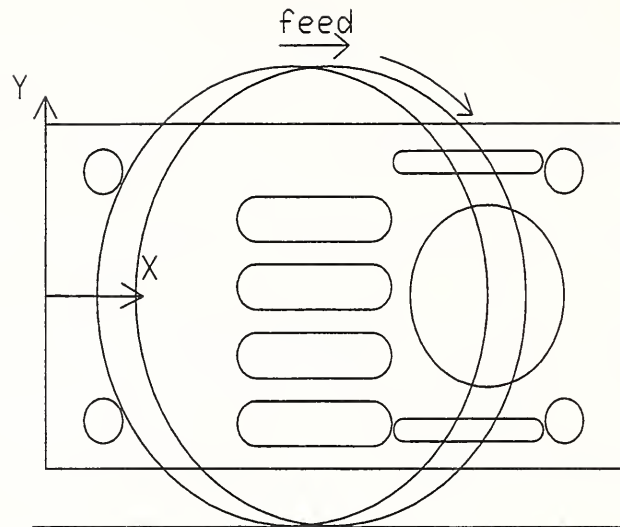| Figure 10. Illustration of cutter engagement with varying interrupted workpiece geometry | Figure 11. Illustration of estimated run-out values in different sections of the workpiece. |

Ideally every row of the objective function should be zero. However, that would represent an over determined system of equations. The least square routine is used to obtain the run-out values that would give a best fit in the solution space. Here, the number of spindle harmonics that can be included is limited by the Nyquist frequency. It was observed that 32 harmonics gives as good as using all harmonics up to Nyquist frequency. A set of typical run-out estimation results are illustrated in Figure 10 and Figure 11.

7. Summary and Conclusions.

Although still in development and prototype stages, the MD-System holds the promise to introduce a new concept in production machining, *machining diagnostics*, and to provide widespread improvements in productivity and quality through the use of the techniques and methods. Several key technical accomplishments have been achieved to date including:

- Development and verification of high speed, synchronous sampling techniques.
- High speed signal analysis including fault indices.
- Strong correlation of retrofit to laboratory sensor data to enable shop floor use.
- Development of sets of diagnostic indices covering a range of machining operations and fault conditions.
- Enhancement of mechanistic process models for end milling, face milling, and drilling to model various fault conditions and predict the resulting cutting forces.

- Prediction of cutting forces for single and multiple fault conditions.
- Application of process models to develop and study diagnostic power of various fault indices.
- Application of process models to develop fault diagnosis techniques including estimation of fault levels for multiple fault conditions through application of advanced techniques such as genetic algorithms.

References

Akshay, 1997, "Process Fault Diagnosis in Face Milling", M.S. Thesis, University of Illinois at Urbana-Champaign.

Daubechies, I., 1990, "The wavelet transform, time-frequency localization and signal analysis,"

IEEE Transactions on Information Theory, Vol. 36, No. 5, pp. 961-1005. Goldberg, D. E., 1989,

Genetic Algorithms in Search, Optimization, and Machine Learning, Reading, MA: Addison-Wesley.

Hong, G. S., Rahman, M., and Zhou, Q., 1996, "Using Neural Network for Tool Condition Monitoring Based on Wavelet Decomposition," Int. J. Mach. Tools Manufact., Vol. 36, No. 5, pp. 551-566.

Jayaram, S., S.G. Kapoor and R.E. DeVor, 1997, "A Model Based Approach for Detection of Process Faults in the Face Milling Process", Trans. of 25th NAMRC, pp.117-122, 1997.

Kim, D. 1996, Model-Based Fault Detection in Machine Tool Diagnostic System, M.S. Thesis, University of Illinois at Urbana-Champaign.

Krishnakumar, K., 1989, "Micro-Genetic Algorithms for Stationary and Non-Stationary Function Optimization," SPIE: Intelligent Control and Adaptive Systems, Vol. 1196, pp. 289-296.

Li, S. and Elbestawi, M. A., 1996, "Tool Condition Monitoring in Machining by Fuzzy Neural Networks," Journal of Dynamic Systems, Measurement, and Control, Vol. 118, pp. 665-672.

Tansel, I. N., Mekdeci, C., and McLaughlin, C., 1995, "Detection of Tool Failure in End Milling with Wavelet Transformation and Neural Networks (WT-NN)," Int. J. Mach. Tools Manufact., Vol. 35, No. 8, pp. 1137-1147.

Zhu, R., Skerlos, S. J., DeVor, R. E., and Kapoor, S. G., 1997, "Application of Genetic Algorithm to Machining Process Diagnostics with a DOE-Based GA Validation Scheme," In Koza, J. R. (editor), Late Breaking Papers at the Genetic Programming 1997 Conference, Stanford, University, CA, pp. 273-279.

Appendix A

Montronix - Machining Diagnostic (MD) System

Machine Type	Operations	Sensors	Signal Processing	Process Faults	Fault Detection Algorithm	Customer Cost or Quality Issue
Machining Center	Milling - All Types	Force - RetroBolt	Time Domain	Runout	Rule Based	Rate of Wear
Transfer Station	Drilling - Twist	Vibration	Frequency Domain	Throw	Expert System	Risk of Breakage
Special Purpose	Drilling - Inserted	AE	Averaging/Filtering	Alignment	Neural Networks	Dimension Error
Lathe	Gundrilling	Sound	Diagnostic Indices	Tilt	Genetic Algorithm	Surface Finish
	Tapping	Displacement	Effective Tooth No.	Breakage		Surface Damage
	Balancing		Peak Indices	Chipping		
	Gear Testing		Harmonic Ratios	Regrind Errors		Process Variability
	Special Applications		Signal to Noise	Wear		Process Improvement
				Vibration		Machine Runoff
						Machine Capability
			Mechanistic Model			Tooling Studies
			Based			Comparative Studies
Applications Engineering			NIST Project Focus			Marketing
		Montronix Machining Diagnostics (MD) System				

Table 1 - Seven Application Elements of Machining Diagnostic System
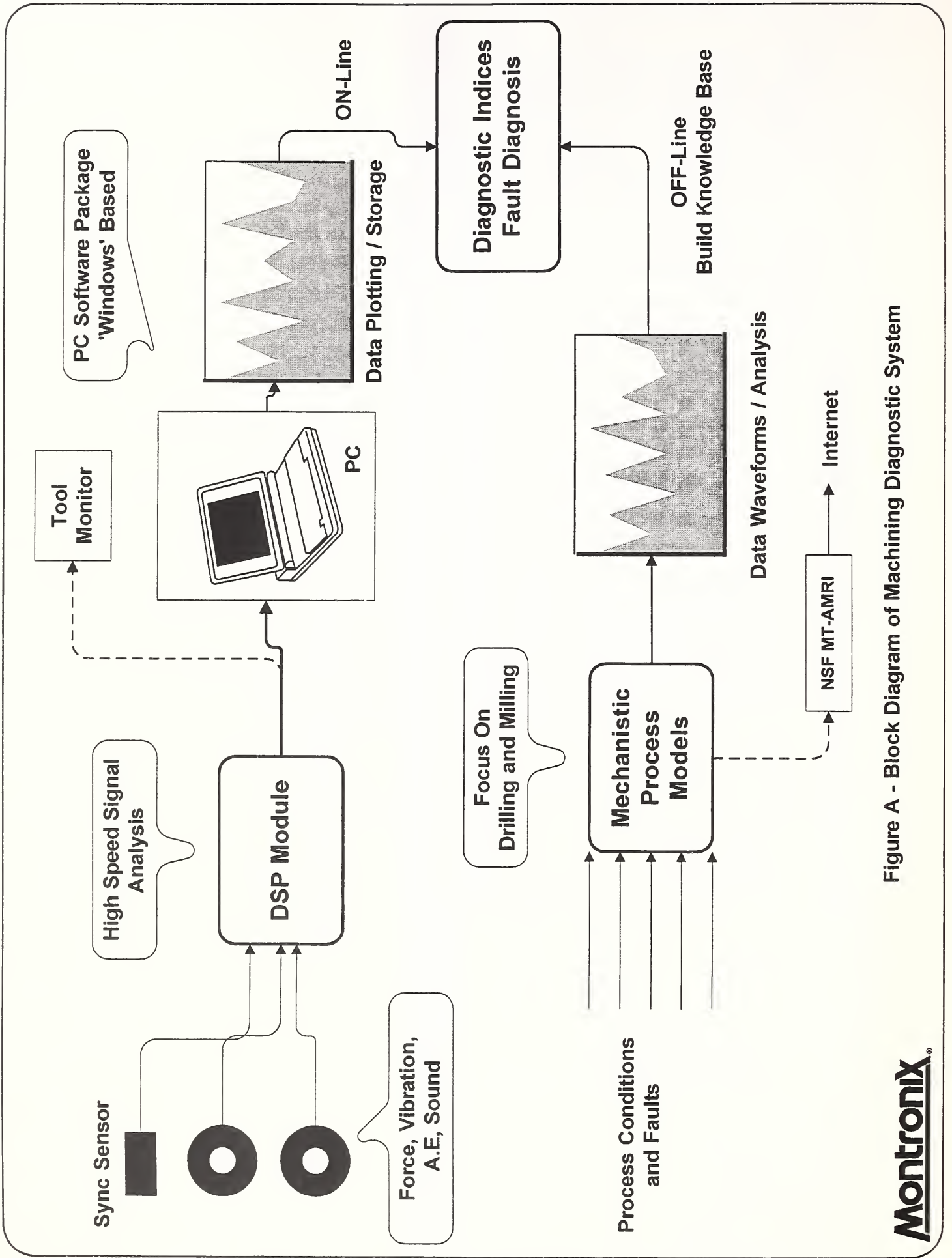
Figure A - Block Diagram of Machining Diagnostic System

Process Model Based Fault Detection On-Line Analysis	Short Term Process Monitoring on Signal and Index Values	Long Term Process Monitoring Based on Trending and SPC
Process Model Based Fault Detection Off-Line Analysis		
Empirical Rule-Based Fault Detection Off-Line or On-Line		
Visual Fault Detection and Analysis		

Signal Processing and 'Index' Calculation Using DSP Technology

Machining Data Display, Storage, and Recall in Windows 95 Application

Figure B - Architecture for Fault Diagnosis and Monitoring Features

MD-System Window Layout

One-cycle plot

Indices being plotted currently

FFT plot for current revolution

Toolbar options for mode selection, scrolling, etc.

Current rev. cursor

Tool and signal information

One-Rev Window

Time and rev. num. in current cycle

Diagnostic indices for current rev

Timing and debugging info

Test date and time

Estimated RPM

Figure (c)

90
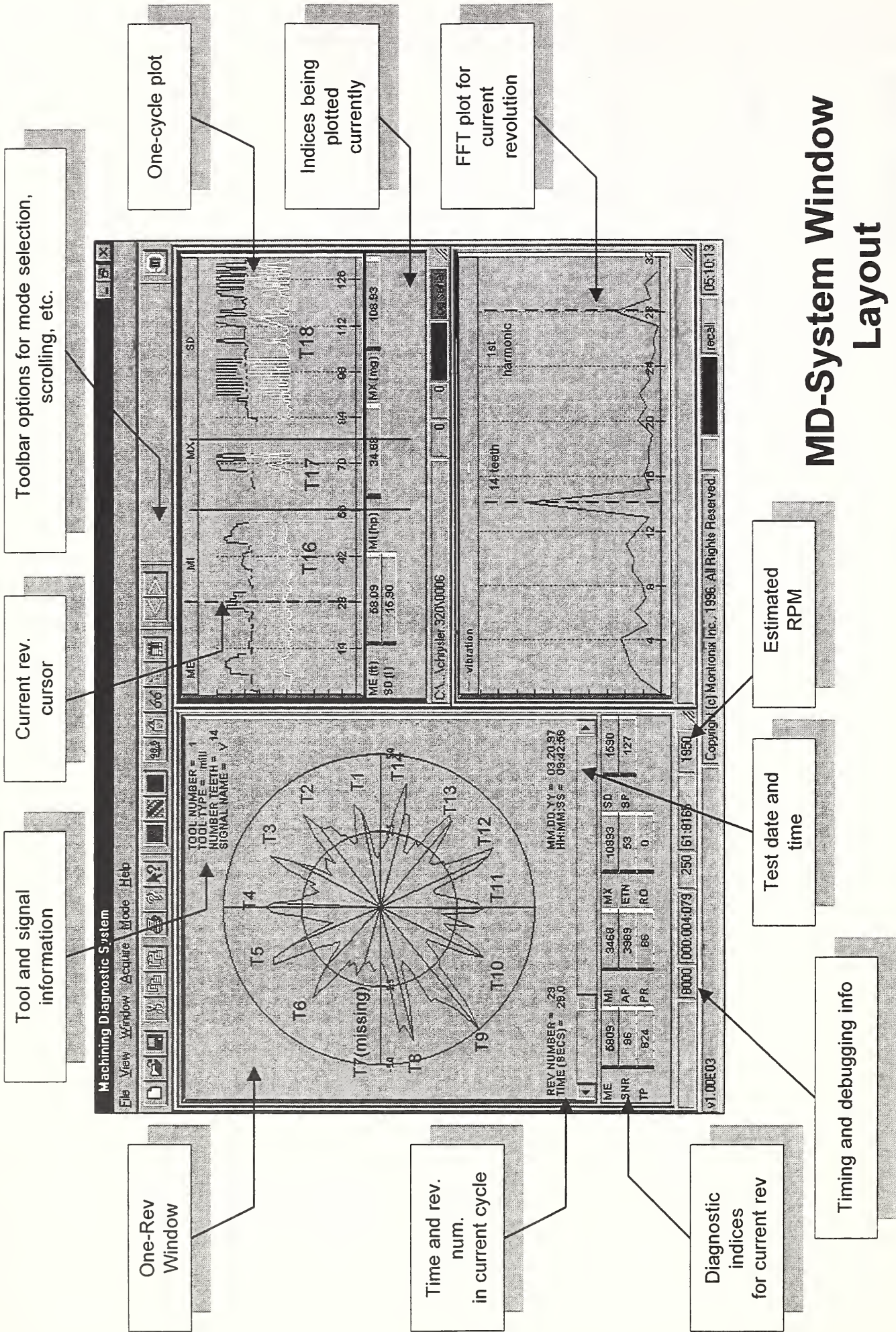
Appendix B

Figure (a): Normal cycle for Tool 1

Figure (b): Tool 1 cycle with missing insert

Figure (c): Tool 1 cycle immediately following insert breakage
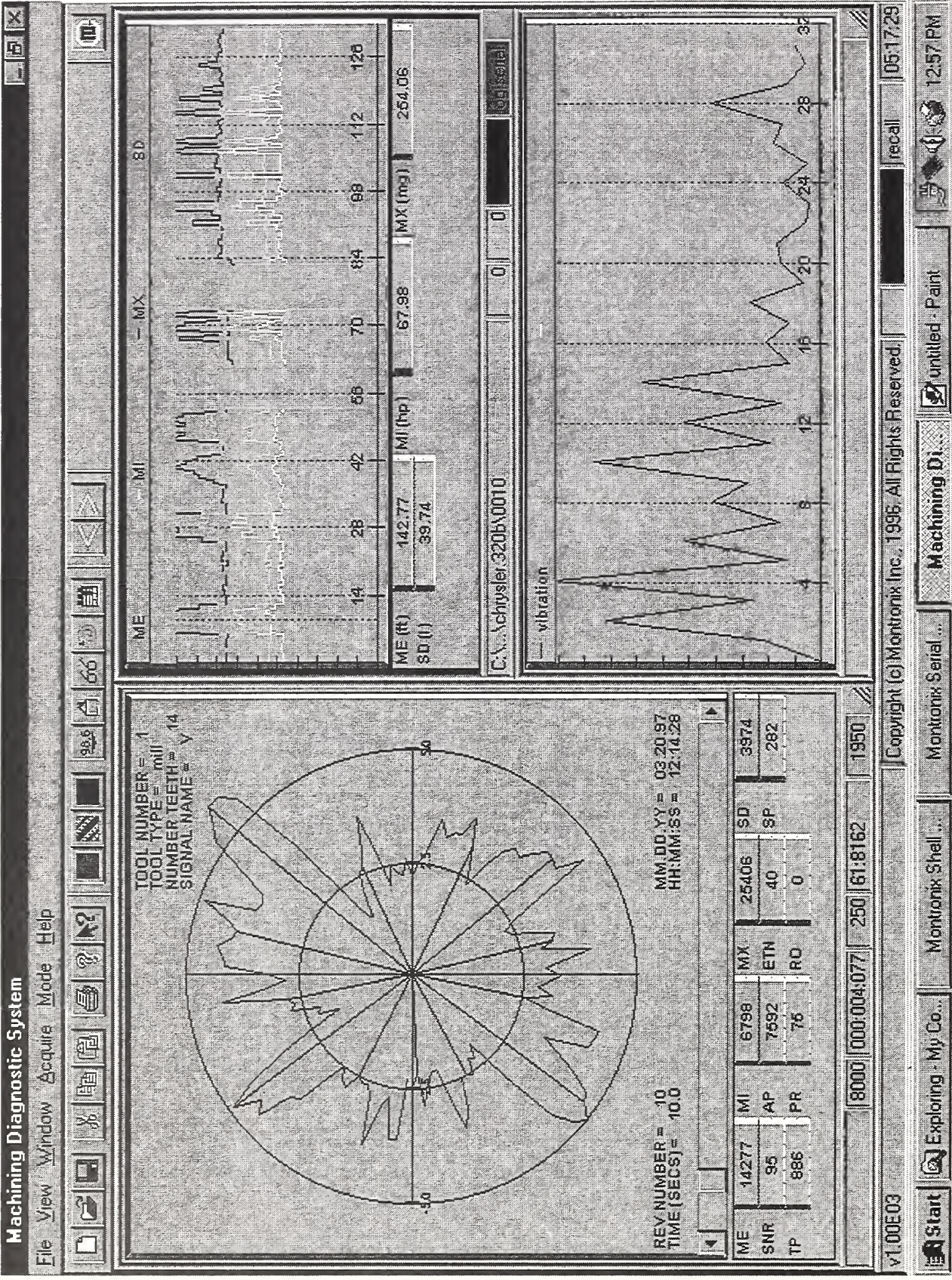
Figure (d): Tool 1 revolution, 2 cycles after insert breakage
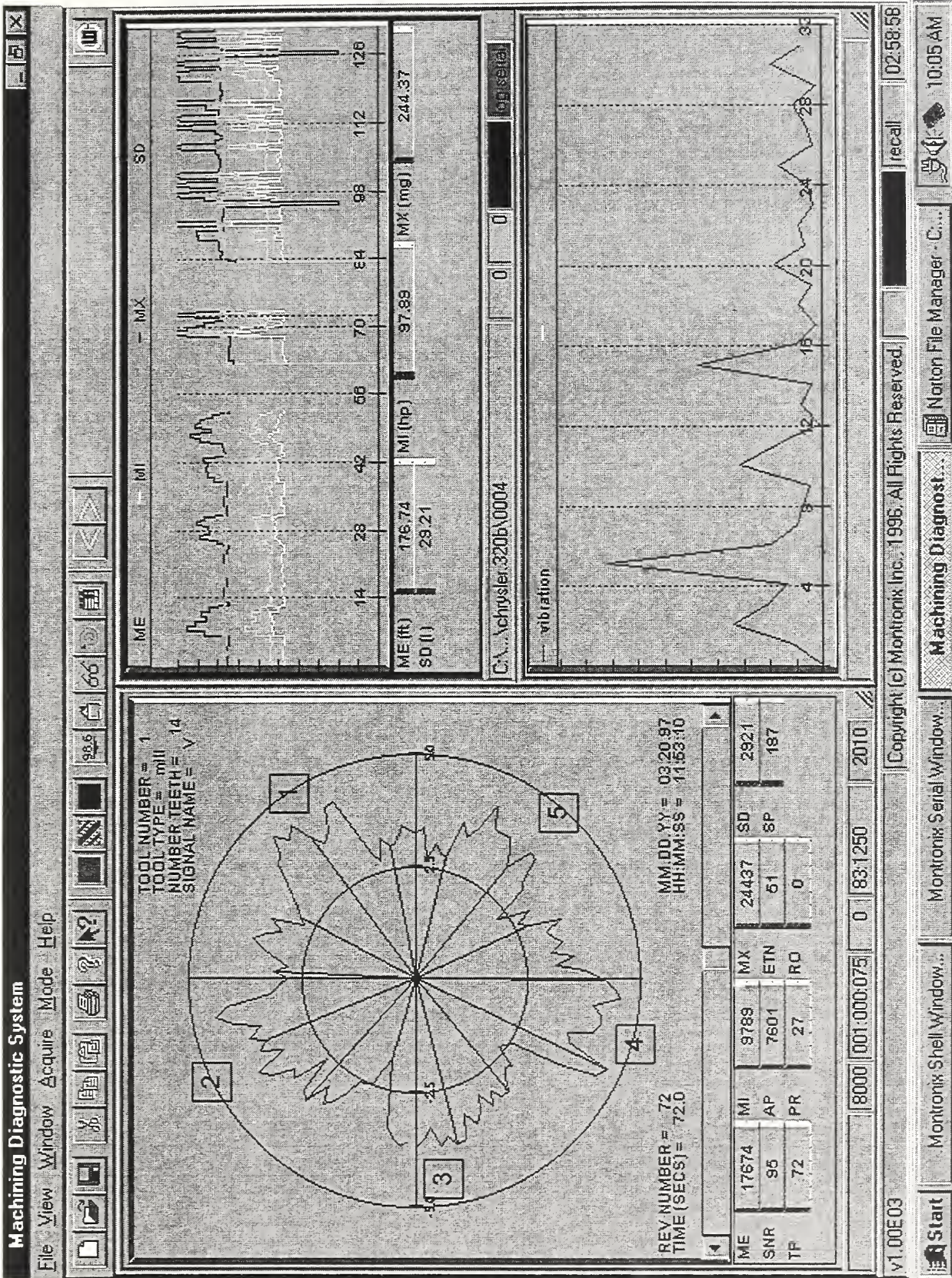
Figure (e): Normal revolution for Tool 2
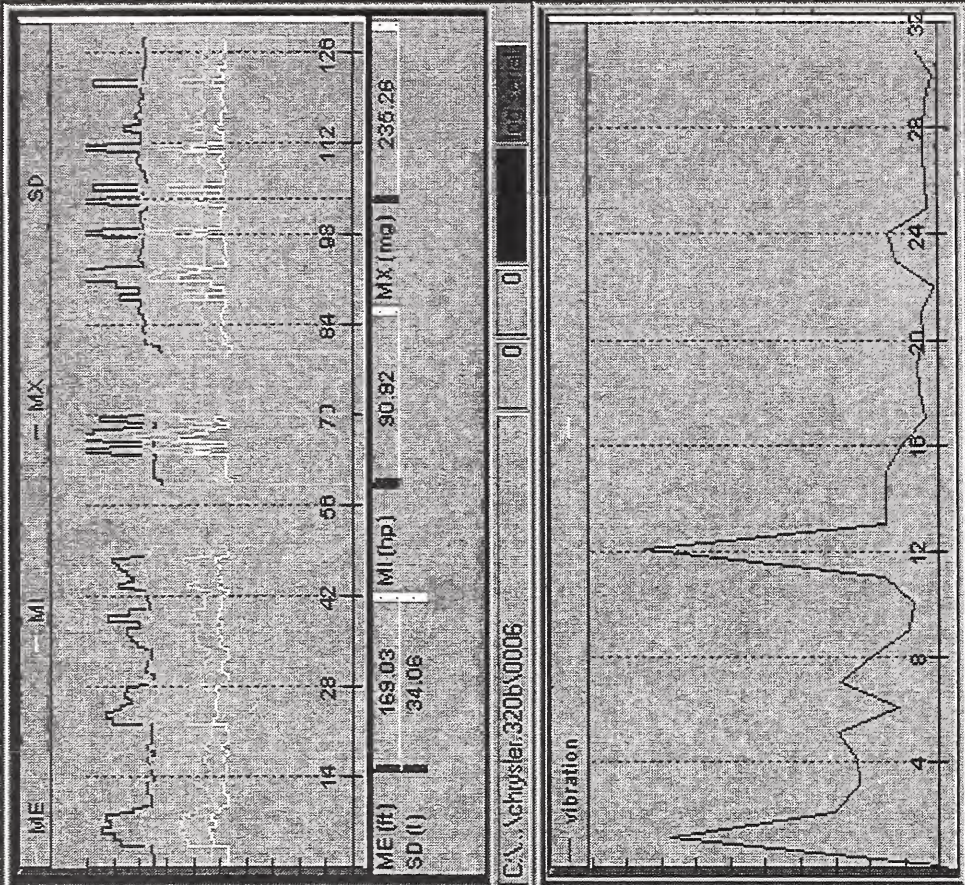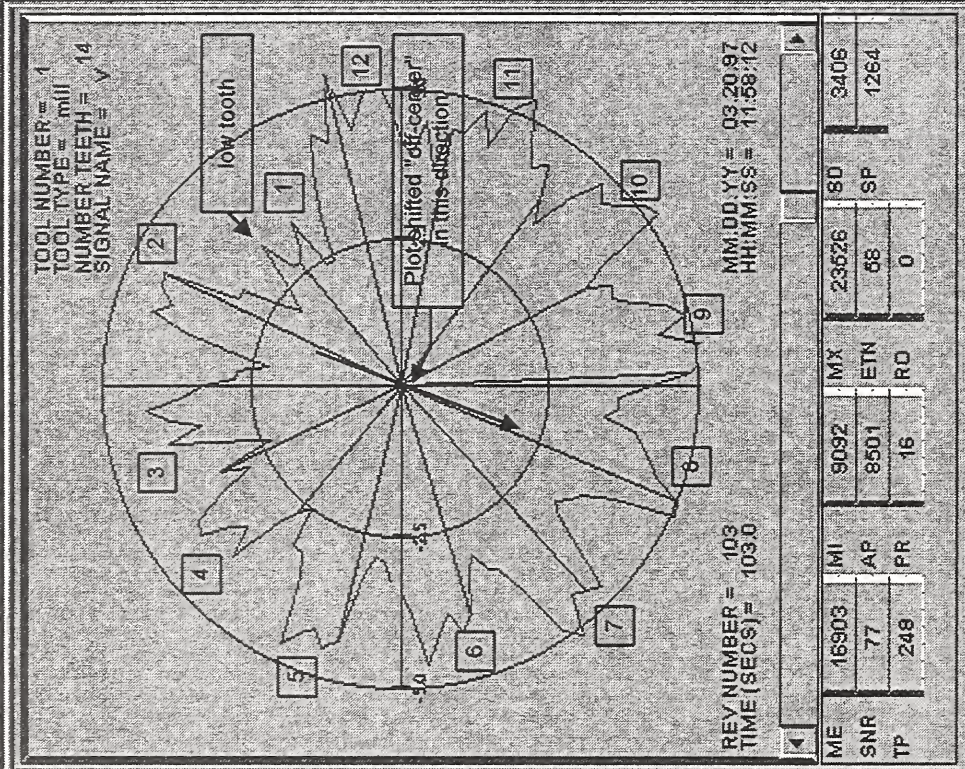
Figure (f): Tool 3 revolution
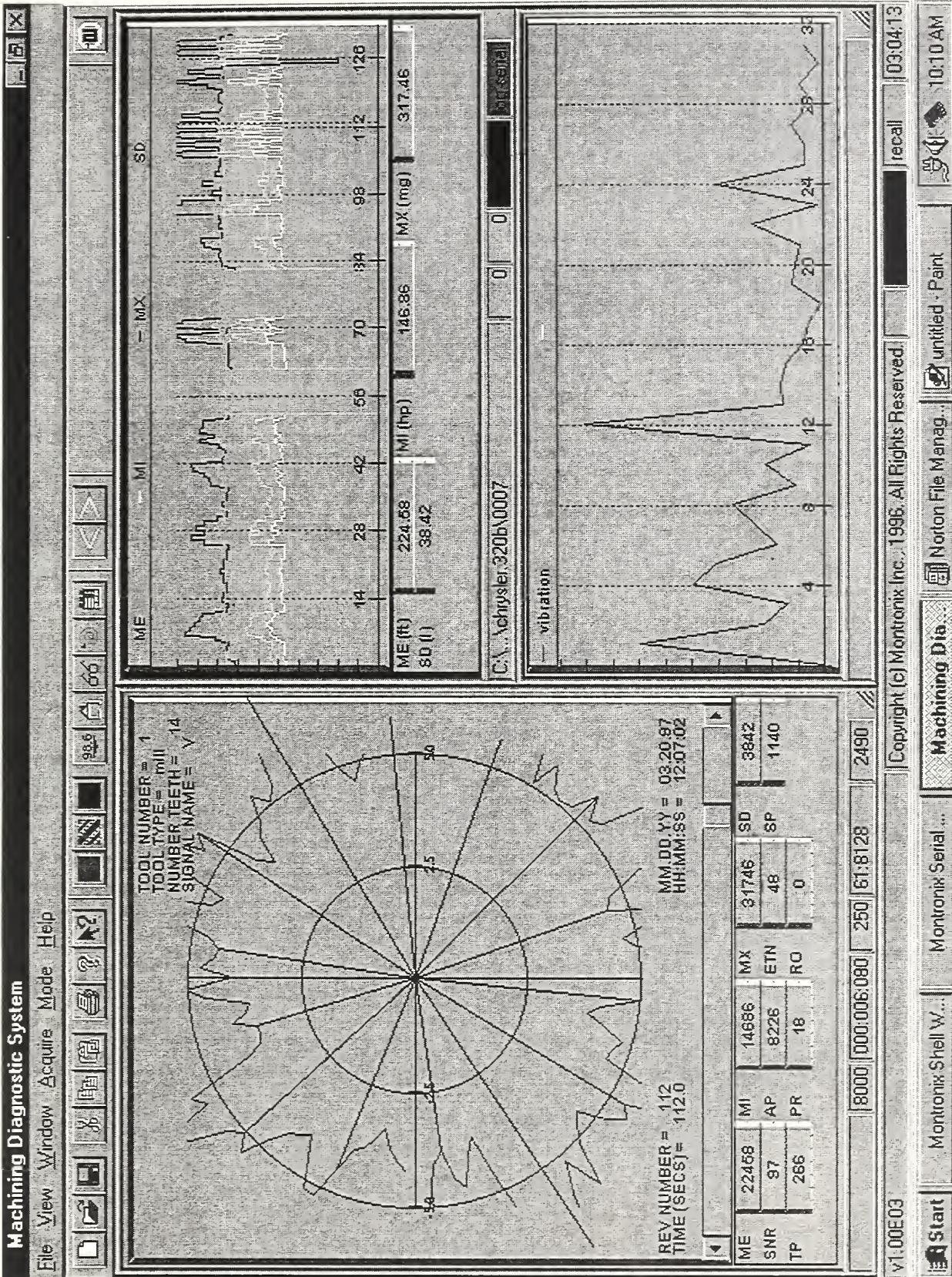
Figure (g): Tool 3 revolution with high signal levels

onit.vi
ast modified on 6/18/97 at 5:47 PM
rinted on 8/4/97 at 12:07 PM

Figure (h)

MODEL BASED FAULT DIAGNOSIS OF THE END MILLING PROCESS

Experimental data   Plot 0

Filtered Experimental Data   Plot 0

Peak Array
0   107.39

Peak Mutiplier
2   1.09

Model Prediction   Plot 0

Peak Array
0   96.93

Peak Multiplier
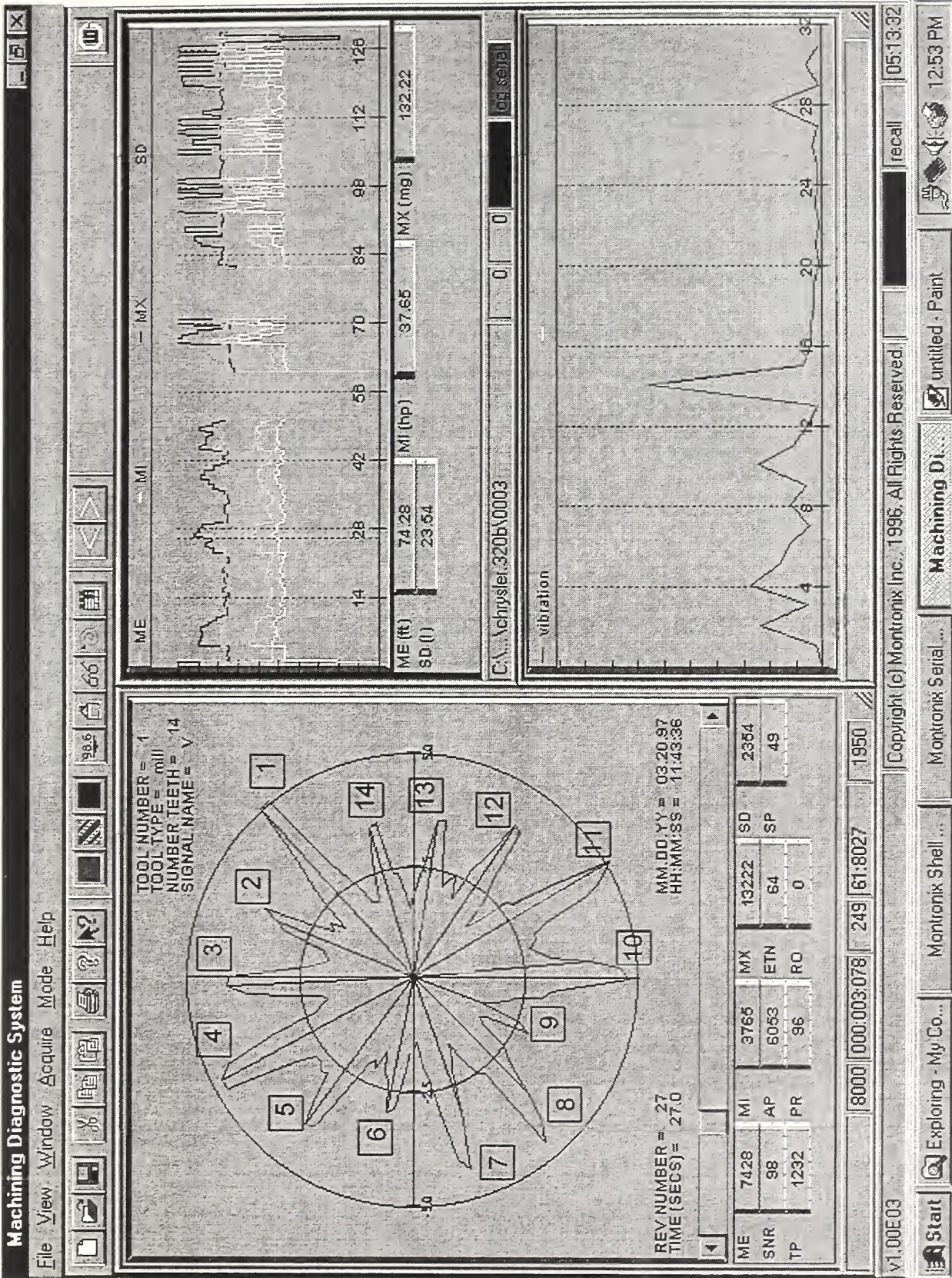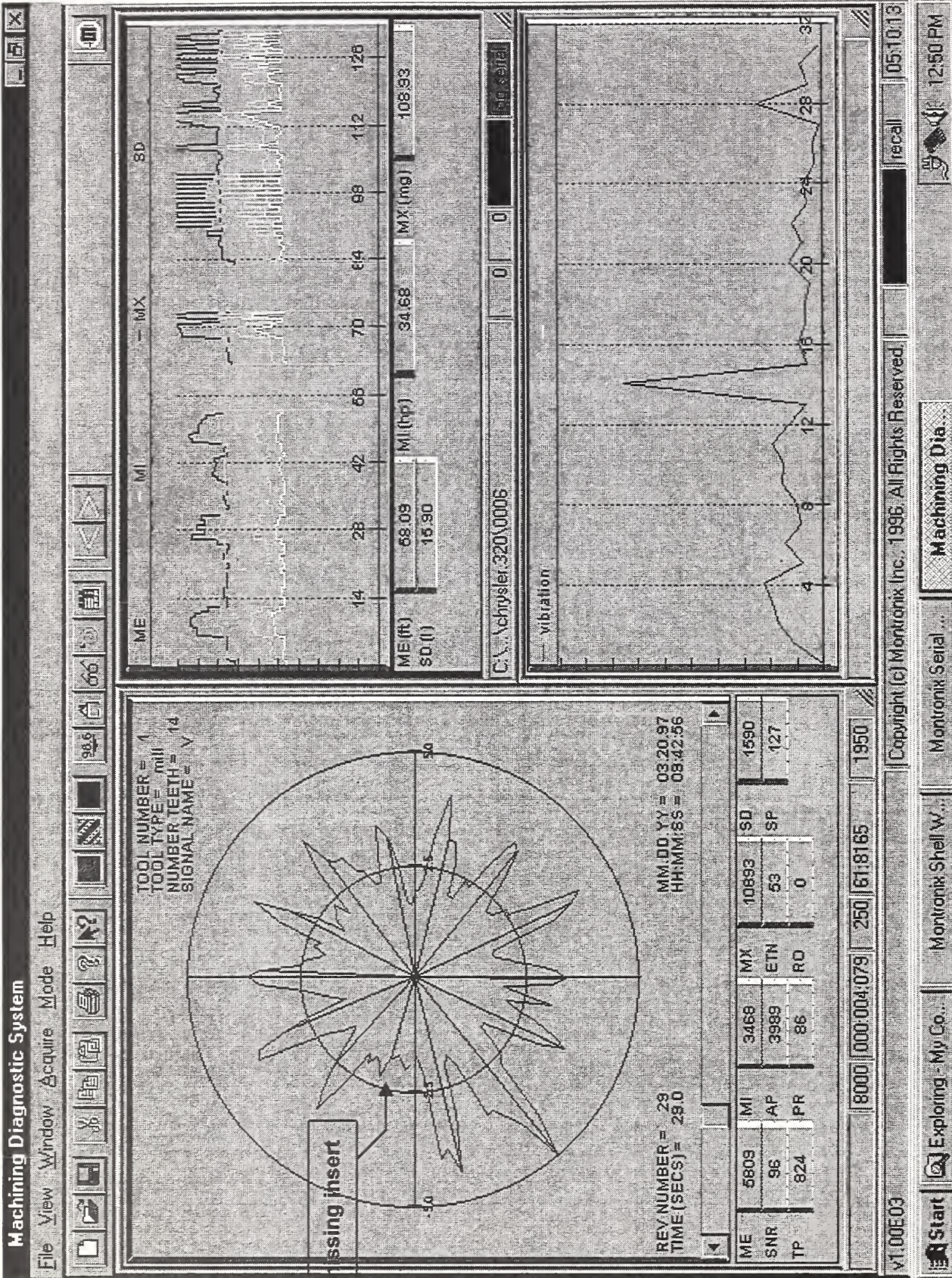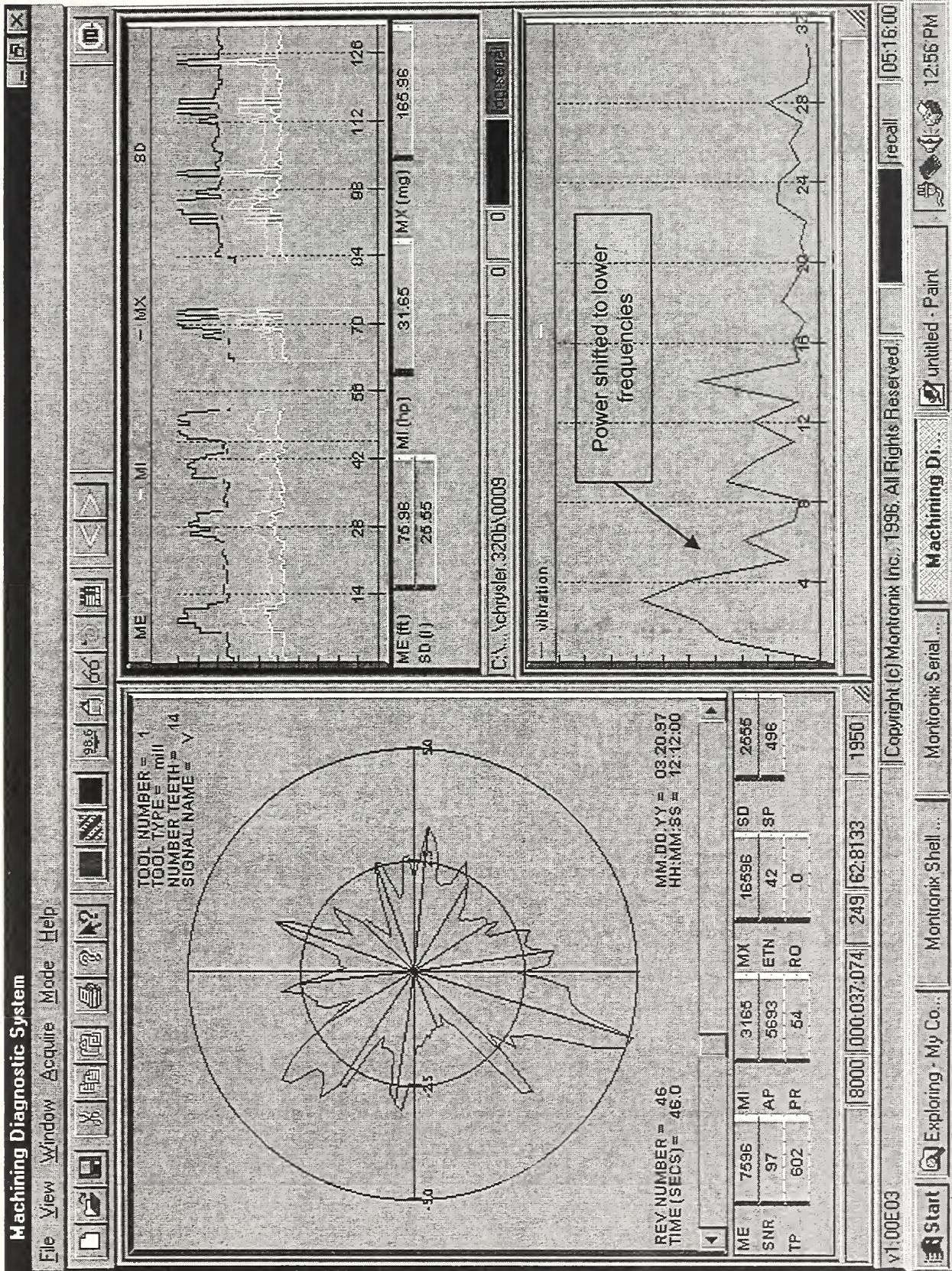1   1.10

M2	M4	Crest Factor	ETN
3.19E-3	3.80E+9	2.74	3.66

% H1	% H2	% H3	% H4	PAR
0.01	0.00	0.01	0.98	0.18

M2	M4	Crest Factor	ETN
2.17E+3	42.04E+6	2.29	3.65

% H1	% H2	% H3	% H4	PAR
0.03	0.00	0.01	0.95	0.19

Axial Depth of Cut
0.20
0.1 0.2 0.3 0.4 0.5

Radial Depth of Cut
0.200
0.100   0.200   0.300

Feed per Tooth
0.004
0.002   0.005   0.008

Spindle Speed
1000.00
0.0   1000.0   2000.0

Simulate

STOP

Runout (Offset)
0.0003
0.0000   0.0010   0.0020

Locating angle
268.2
0.0   200.0   360.0

Flute Deviation
OFF   ON

Flute Deviation
3   0.0000
0.0000   0.0005   0.0010
0.0000
0.0000   0.0005   0.0010
0.0000
0.0000   0.0005   0.0010
0.0000
0.0000   0.0005   0.0010

Flute Breakage
OFF   ON

Flute Breakage
0   0.00
0.0   0.5   1.0
0.00
0.0   0.5   1.0
0.00
0.0   0.5   1.0
0.00
0.0   0.5   1.0

Flute Chipping
OFF   ON

Flute Chipping
0   0.0510
0.0   0.2   0.5
0.0510
0.0   0.2   0.5
0.0510
0.0   0.2   0.5
0.0510
0.0   0.2   0.5

power spectrum graph (experimental data)

power spectrum graph (simulated data)

Plot 0

Plot 0

nit1.vi
st modified on 6/18/97 at 5:46 PM
inted on 8/4/97 at 12:13 PM

Figure (i)

MODEL BASED FAULT DIAGNOSIS OF THE END MILLING PROCESS

Session III. Advanced Machining (2)

Compact, High-Power Spindles
J.W. Grant, Ford Advanced Technology Development Center;
J.T. McCabe, NCMS

Advanced Hydrostatic Bearings for a New Class of Machining Spindles
R. Schaffa, GM Manufacturing Center; J.T. McCabe, NCMS;
and K. Wasson, Aesop

Ultrasonic Detection of Unstable Plastic Flow in Hard Metal Cutting
M.A. Davies, S.E. Fick, C.J. Evans, and G.V. Blessing,
NIST Manufacturing Engineering Laboratory

Compact, High-Power Spindles

John W. Grant
The Ford Motor Company

John T. McCabe
National Center for Manufacturing Sciences

Abstract: In support of manufacturing flexibility and high-velocity machining, the automotive industry seeks lightweight, small, high-powered spindles. A class of rugged spindles on rolling element bearings is described that enables flexibility by providing torque-speed characteristics suitable for machining steel, cast iron, and aluminum. New spindle motor technology addresses the requirements of high-velocity machining through permanent magnet motors controlled by drives with the capacity to deliver 180 amps rms per phase at frequencies over 1,000 Hz, and accelerate the spindle at a rate in excess of 30,000 rpm/sec. This paper describes the design concept and provides preliminary test results for these spindles.

Introduction

Rapid response to customer demands for diversity has directed the automotive industry to seek more flexible means for machining components. This has led to the current evolution of multi-axis, high-velocity machining centers for production-run quantities that cannot justify the cost of retooling transfer lines. Concurrently, the automotive industry is developing transfer lines that incorporate stations with programmable machining heads. In both cases, there is a need for a family of spindles that have two basic characteristics: minimum size and weight, and maximum torque over a broad speed range. Small size allows a spindle to be applied to a large variety of machine configurations. Low weight allows rapid repositioning of the cutter with minimum slide drive power. High torque over a broad speed range allows more flexibility in cutting a greater range of material hardness without suffering machine down time to change spindles.

On July 1, 1994, the team consisting of Ford Motor Company (Ford), General Motors Corporation (GM), the Torrington Company (Torrington), SETCO-Whitnon (SETCO), Giddings & Lewis (G&L), and Manufacturing Laboratories, Inc. (MLI), led by the National Center for Manufacturing Sciences (NCMS), began spindle development work co-funded by NIST under the Advanced Technology Program (ATP). The general purpose of the project was to develop families of machining spindles that would meet the needs imposed by global competition.

Two families of spindles that are distinguished by their bearings are under development. One family is based on rolling element bearings and the other on fluid film bearings. This paper deals with machining spindles supported by rolling element bearings that address the requirements of the spindle end users who initiated the project on which this paper is based. These requirements focus on a family of spindles for which size and weight are minimized for high-velocity machining, and for which torque and speed range are maximized to provide flexibility for a variety of workpiece materials. (A companion paper in these Proceedings, "Advanced

Hydrostatic Bearings for a New Class of Machining Spindles," deals with the other family of spindles, those supported by a new fluid film hydrostatic bearing.)

To demonstrate the enabling technologies, two spindles were built and are undergoing evaluation tests. One spindle, the 75-hp spindle, was designed as a hogging spindle with maximum power of 100 hp and a maximum speed of 12,000 rpm. The other spindle, the 35-hp spindle, was designed as a general-purpose machining spindle with maximum power of 50 hp and maximum speed of 20,000 rpm. Both spindles are distinguished by their low weight and the high power they deliver for the space they occupy.

Key Design Features

Motor Influence on Spindle Properties. The size, weight, and shaft stiffness of a integral motor machining spindle is dictated primarily by its motor. A trade-off study was made of AC motors, switched-reluctance DC motors, and permanent-magnet, brushless DC motors. Table 1 shows a self-consistent set of comparative results for the three types of motors.

Table 1. Comparison of AC Induction, Switched Reluctance DC (SRB DC) and Permanent Magnet Brushless DC (PMB DC) Motors

Parameter	AC Induction Motor	SRB DC Motor	PMB DC Motor
Power. kW (hp)	56.25 (75)	56.25 (75)	56.25 (75)
Rotor Inside Diameter, mm (in.)	121 (4.75)	121 (4.75)	121 (4.75)
Stator Outside Diameter, mm (in.)	373 (14.7)	318 (12.5)	208 (8.18)
Rotor Length, mm (in.)	381 (15.0)	188 (7.4)	140 (5.5)
Rotor Outside Diameter, mm (in.)	211 (8.3)	203 (8.0)	145 (5.7)
Weight, N (lb)	1557 (350)	712 (160)	356 (80)
Rotor Heating	High	Medium	Low

The motor study led to the selection of a water-cooled, three-phase, six-pole permanent magnet motor as the motor of choice for the spindle family. These motors are electrically commutated by an encoder wheel fitted to the shaft. For the sizes investigated, this motor type was significantly smaller and lighter than the other two types and, because rotor laminations were not required, the shaft diameter in the motor area could be made larger and therefore, stiffer. Table 2 presents the development goals for the two prototype spindles.

Table 2. Development Goals for 75-hp and 35-hp Spindles

Parameter	75-hp Spindle	35-hp Spindle
Weight, N (lb)	2,224 (500)	1,334 (300)
Maximum Outside Diameter,* mm (in.)	254 (10)	254 (10)
Static Stiffness 102 mm (4 in.) From Nose, kN/m (lb/in.)	96,305 (550,000)	61,285 (350,000)
Maximum Speed, rpm	12,000	20,000
First Modal Frequency, Hz	600	600
Stiffness of First Mode, kN/m (lb/in.)	131,325 (750,000)	65,662 (375,000)

* Includes stator cooling jacket.

Motor Influence on Speed Range. As the motor speed increases, the back electromotive force (EMF) generated by the motor also increases. When the back EMF equals the line voltage, the motor speed reaches its limit. To extend the speed limit, the flux in the air gap needed to be altered by a technique called "phase advance." For example, without phase advance, the maximum attainable speed for the 75-hp motorized spindle was approximately 8,000 rpm. With phase advance, the speed could be increased to 12,000 rpm. Consequently, phase advance was made part of the requirements for the motor drive electronics.

Motor Influence on Torque. A design study indicated that the optimum means for delivering the required torque over the full speed range was to provide one type of winding at low speed and a different type of winding at high speed. Each phase included at least two coils. The winding leads were configured so that the motor controller could switch from series connection of the coils at low speed and parallel connection at high speed. Winding switching capability was also made a requirement of the motor drive electronics.

Motor Drive. Early in the project, the motor drive was recognized as the critical limitation in meeting the torque-speed goals for all members of the spindle family. At the low-speed end, adequate torque to machine steel was desired; at the high-speed end, adequate high power was desired to cut aluminum. These objectives translated into a need for a motor drive that could deliver 630 V DC, 180 amps rms per phase at frequencies up to 1,000 Hz.

Contacts were made with 22 drive suppliers. Twenty-one could meet either the current requirement, or the frequency, but not both. Performance Controls, Inc., (PCI) was selected as the supplier because they were familiar with permanent-magnet brushless motors and could deliver the current and frequency. However, neither PCI, nor anyone else in the U.S., could provide off-the-shelf phase advance and winding switching. Subsequently, phase advance and winding switching became part of this development effort. Both these features have been developed and implemented successfully by the PCI digital signal processor (DSP) based, all-digital drive controller.

Bearings and Lubrication: Because the primary application for the 75-hp spindle was high-power hogging or roughing cuts, the principal requirements for the bearings were very high stiffness, very high load capacity, and the ability to survive accidental crashes without catastrophic bearing failure. Oil-jet lubricated Timken taper roller bearings were chosen; a 95-mm bore modi-

fied Hydra-Rib for the front bearing; and a 75-mm bore, modified, preload-controlled Hydra-Rib for the rear bearing.

The primary application for the 35-hp spindle was general purpose machining, where its torque-speed curve would be stretched (compared to the state of the art) to a minimum of 35 hp in the speed range between 2,500 rpm and 20,000 rpm. To achieve this goal, the motor drive was equipped with a PCI high-current, high-frequency digital drive with phase advance and winding switching capability. Oil-air lubricated, hybrid-ceramic ball bearings were chosen; a 70-mm bore Fafnir ceramic-ball, steel-race front bearing; and a 60-mm bore Fafnir ceramic-ball, steel-race rear bearing.

Toolholder System. The 75-hp spindle is designed for a hollow shank taper HSK 80-B toolholder retained by an Ott draw bar. The 35-hp spindle is designed for a hollow shank taper HSK 63-A toolholder also retained by an Ott draw bar. A proximity sensor manufactured by API, Inc., is used to detect the status of the draw bar.

Prototype Designs

75-hp Prototype. Figure 1 shows the design of the 75-hp prototype spindle (56.25 kW). The stator outside diameter (OD) is only slightly over 203 mm (8 in.), and the OD of the spindle housing is 245 mm (10 in.)—the difference being used by the water cooling jacket and structure. The outside dimension was selected to fit inside the ram of the G&L RAM 630 machine tool, which is being used to perform cutting tests with the spindle.
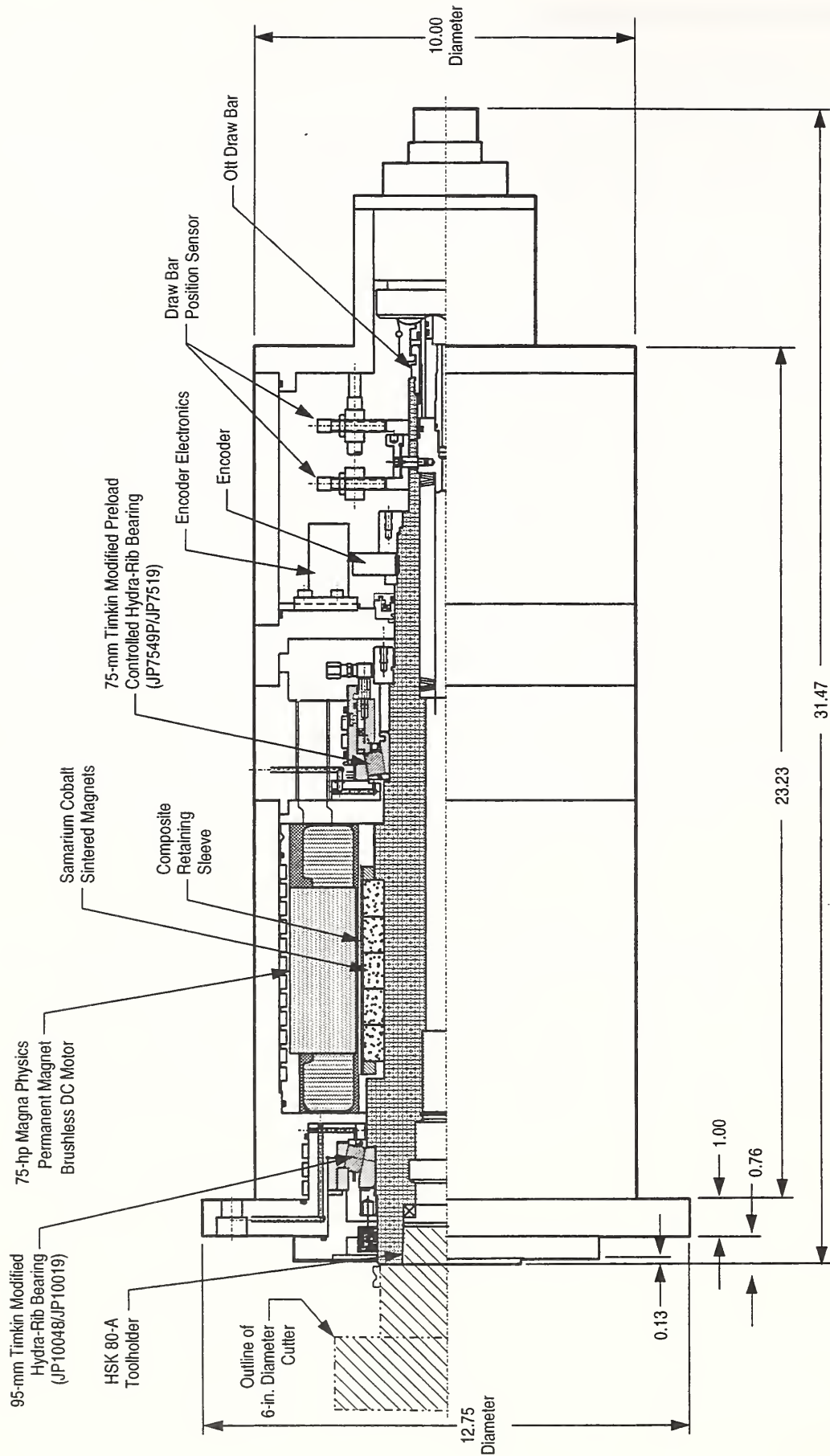
35-hp Prototype. Figure 2 shows the design of the 35-hp prototype spindle (26.25 kW). The OD of this stator is only 163 mm (6.4 in.) but the OD of the spindle housing is also 254 (10 in.), making the two spindles interchangeable in the RAM 630 test machine.

Test Results

Before the machining spindle prototypes were constructed, two test rigs were built. One test rig, built to simulate the 75-hp spindle, was instrumented to evaluate the modified Hydra-Rib bearings and their lubrication system. This rig was coupled to a variable-speed drive. The second test rig was built to evaluate the 75-hp spindle motor/motor drive system. To gain information regarding the performance of the bearing and lubrication system planned for the 35-hp spindle, the 75-hp motor test rig was fitted with hybrid-ceramic bearings and an ORSCO oil–air lubrication system. From these tests, the following conclusions were reached:

- Hydra-Rib Bearing System
    - The hot spot in the Hydra-Rib bearing is at the interface between the rib and rollers. By placing one thermocouple in the housing close to the rib/roller interface, the most critical temperature of the bearing can be monitored effectively.
    - Up to a speed of about 6,000 rpm, which corresponds to a DN of 570,000, the ORSCO oil–air controlled drip lubrication system performed satisfactorily. Above 6,000 rpm, heat generated at the rib-roller interface can no longer be transferred adequately, and jet lubrication must be used.

- Hybrid-Ceramic Bearing System

  – An ORSCO air-oil system lubricates this bearing system adequately to 20,000 rpm, a DN of 1.4 million.

  – Due to the small footprint of the ceramic balls, the stress on the steel raceways can be high during heavy load operations such as drilling, and the corresponding bearing life might be low. Under lighter load operations such as milling, bearing life is satisfactory.

- Permanent-Magnet Motor

  – The PM motor produced the expected torque; however, the motor generated more heat than was expected from the design analysis, limiting the amount of current (and therefore torque) that could be applied.

  – DC testing combined with heat transfer modeling indicated that the motor was developing more electrical heat than anticipated. Experimental testing of the stator showed a marked increase in stator electrical resistance as a function of frequency (motor speed). This increase is shown in Figure 3 for one phase of the 75-hp and 35-hp spindle motors.

  – Using a specialized finite element analysis, the unexpected motor loss was identified as a proximity effect. In this effect, the effective winding resistance is increased when current in the windings is crowded to one side of the winding wire by the mutual influence of multiple current-carrying conductors.

  – The magnitude of the proximity losses have been calculated and measured. These losses are sensitive to the way the motor is wound.

Figure 1. 75-hp Prototype Spindle

All dimensions are in inches.

Draw Bar
Position Sensor

Olt Draw Bar

Encoder Electronics

Encoder

75-mm Timkin Modified Preload
Controlled Hydra-Rib Bearing
(JP7549P/JP7519)

Samarium Cobalt
Sintered Magnets

Composite
Retaining
Sleeve

75-hp Magna Physics
Permanent Magnet
Brushless DC Motor

95-mm Timkin Modified
Hydra-Rib Bearing
(JP10048/JP10019)

HSK 80-A
Toolholder

Outline of
6-in. Diameter
Cutter

10.00
Diameter

31.47

23.23

1.00

0.76

0.13

12.75
Diameter

Figure 2. 35-hp Prototype Spindle

All dimensions are in inches.
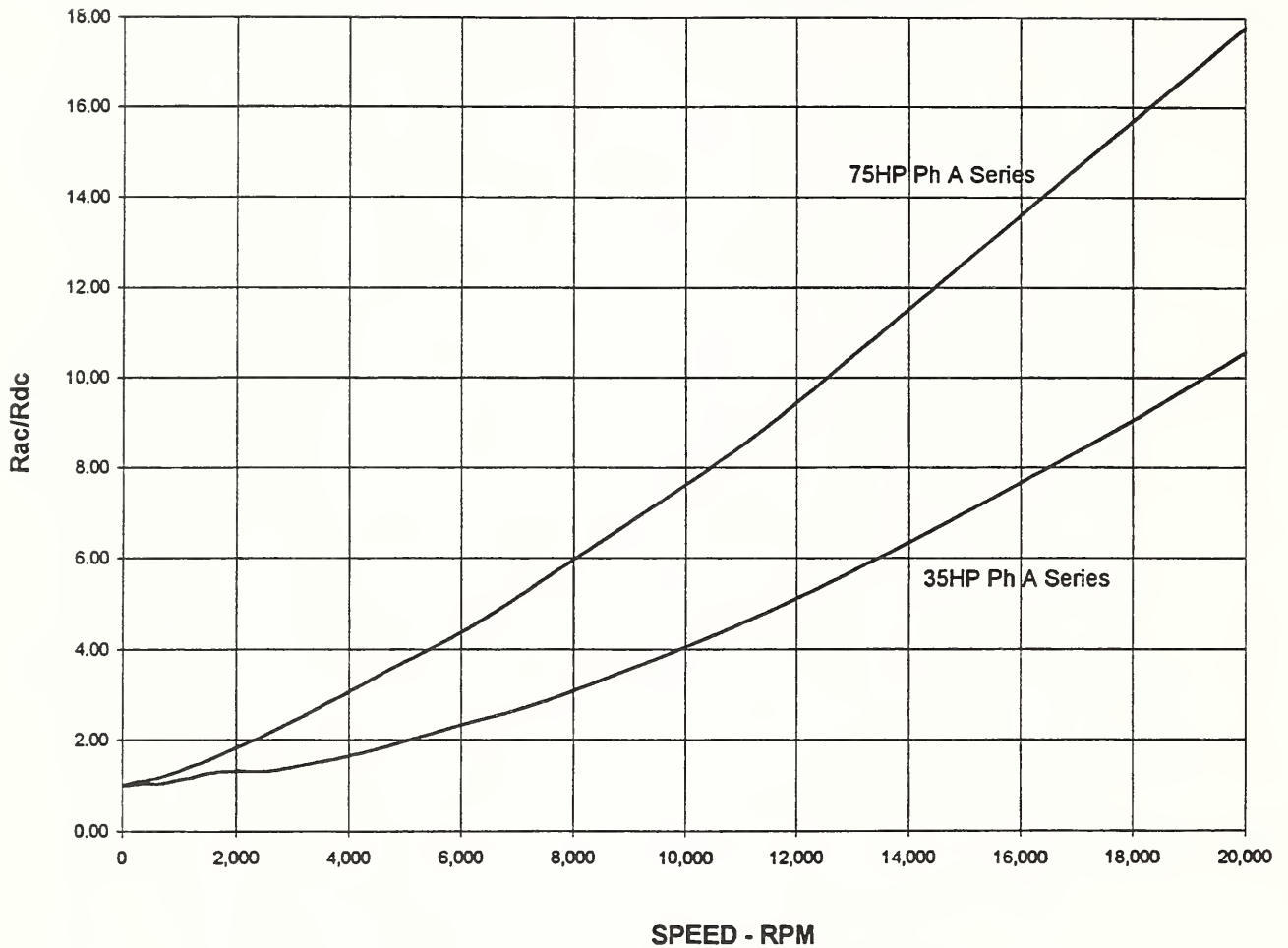
Rac/Rdc for 75HP-1 and 35HP-2 Motors

Figure 3. Ratio of AC to DC Winding Resistance Over Spindle Speed Range

- Permanent-Magnet Motor Drive
  - Tests showed that the motor drive system provided the power required to meet the torque-speed curve requirements for both spindles; however, the motor cooling system was not designed to dissipate the unanticipated losses caused by proximity effects.
  - The ratio of AC resistance to DC resistance over the spindle speed range indicates the severity of proximity effect losses. This ratio appears to be a reliable indication of the motor heat that must be dissipated due to this effect.
  - A quality control test, based on the AC/DC resistance ratio, is being developed to measure the efficiency of the motor winding proximity effect. This test will allow the stator to be evaluated before it is potted.

- Spindle Acceleration
  - As shown in Figure 4, in the speed range of 0 to 3,500 rpm, the acceleration of the 75-hp spindle was 52,000 rpm/sec. In the range of 3,500 to 5,000 rpm, the

acceleration was 22,000 rpm/sec. Figure 4 also shows the current and phase advance applied during this acceleration test.

– During the acceleration tests, the maximum current was 240 amps, which produced a torque of 3100 in.-lb.

Variables:
1. Speed: 1 krpm/div
2. Current command: 280 A rms/div
3. Phase advance: 90°/div (+ down)

Acceleration:
0–3.5 krpm: 52 krpm/sec
3.5–5 krpm: 23 krpm/sec

Maximum Torque:
259 lb-ft

Total Inertia:
Approximately 0.52 lb-in.-sec^2

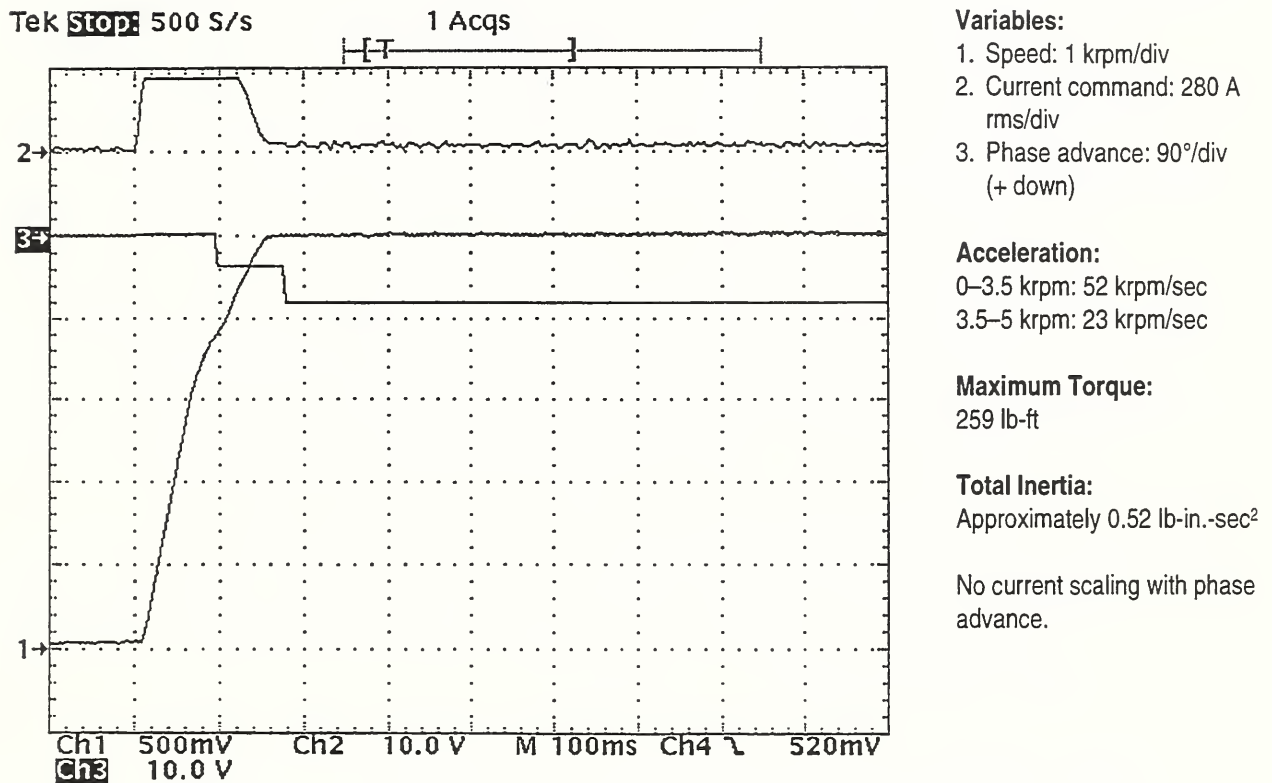No current scaling with phase advance.

Figure 4. 75-hp Motor Acceleration to 5,000 rpm Series Winding Connection (From Performance Controls, Inc.)

- Static and Modal Stiffness

  – Table 3 presents the measured dynamic and static stiffness taken from the 75-hp tapered roller bearing test spindle. These measurements were taken at various spindle speeds and bearing preloads.

Table 3. Stiffness Versus Preload and Spindle Speed 75 hp Bearing Test Rig

Bearing Preload, kN/m²/(psi)	Spindle Speed, rpm	Dynamic Modal Stiffness, kN/m (lb/in.)	Static Stiffness, kN/m (lb/in.)
0	0	160,041 (914,000)	109,438 (625,000)
345 (50)	0	200,314 (1,144,000)	127,648 (729,000)
345 (50)	400	187,357 (1,070,000)	—
345 (50)	6,000	191,384 (1,093,000)	—
690 (100)	0	174,400 (996,000)	145,858 (833,000)
390 (100)	400	188,233 (1,075,000)	—
390 (100)	10,000	196,287 (1,121,000)	—
390 (100)	12,000	173,524 (991,000)	—
1034 (150)	0	164,244 (938,000)	145,858 (833,000)
1034 (150)	600	180,003 (1,028,000)	—
1379 (200)	—	168,442 (962,000)	145,858 (833,000)

- Measured and Modeled Dynamic Response
  - Table 4 presents a comparison between the measured and modeled characteristics for the 75-hp tapered roller bearing test spindle. As shown by the table, the models provide accurate predictions.

Table 4. Comparison of Measured and Modeled Values for Test Spindle

Parameter	Measured		Modeled	
Test Spindle	Tapered Roller Bearing	Ball Bearing	Tapered Roller Bearing	Ball Bearing
Frequency, Hz	691	651	692	653
Minimum Modal Stiffness, kN/m (lb/in.)	~175,100 ~1,000,000	143,292 (818,342)	191,621 (1,094,350)	164,983 (942,221)
Minimum Dynamic Stiffness, kN/m (lb/in.)	11,471 (65,510)	4,745 (27,100)	Not Calculated	Not Calculated
Damping Ratio	0.025 – 0.035	0.015 – 0.025	Not Calculated	Not Calculated
Static Stiffness, kN/m (lb/in.)	145,858 (833,000)	Not Measured	188,320 (1,075,500)	137,716 (786,500)

Technical Achievements

Motor Design. The means for designing high-power, compact machining spindles with integral permanent-magnet motors have been developed. Built with yesterday's technology, the motors of these and other compact spindles are likely to overheat when pressed to their maximum torque, particularly at low speed. The cause of this problem has been discovered. The means to avoid the problem will soon be available and a quality control test has been developed to ensure that the problem has been minimized in a particular design.

Drive Design. A unique motor drive has been developed that is capable of driving a permanent-magnet motor with frequencies in excess of 1,000 Hz and peak currents in excess of 250 amps per phase. This drive enables high torque over a broad speed range through winding switching and phase advance.

High DN Tapered Roller Bearing System. Timken tapered roller bearings have been modified to allow their use at higher speeds than standard roller bearings. The means for removing hot lubrication oil from the high shear zones of the bearing have been incorporated, a jet lubrication system has been designed to lubricate and cool the bearing, and a cooling jacket has been integrated to control the temperature of the outer race.

Draw Bar Position Sensor. A single API, Inc., noncontacting sensor system has been designed to monitor the position of the draw bar that retains the tool holder. This sensor system replaces three contact switches and makes the spindle system more robust and reliable.

Conclusion

The motor and its drive are at the heart of advanced spindle design for high-velocity machining. Selecting high-energy, permanent-magnet brushless DC integral spindle motors and extending motor drive technology to include winding switching and phase advance has led to compact, high-power spindles with extended speed ranges. The capability for forecasting the performance of such designs is at the current limit of today's technology. However, as a result of the spindles development project, a battery of analytical tools is now available that have been tested and found capable of accurately characterizing motor performance. The project team improved spindle evaluation test equipment and test procedures. They developed a quality control test device for assessing frequency-related stator losses that can restrain the performance of these advanced motors. When design guidance from the analysis of cutting test performance is added to these tools, the spindles project will have made a significant contribution to the state of the art of building advanced machining spindles that perform as forecasted.

Advanced Hydrostatic Bearings for a New Class of Machining Spindles

Ronald Schaffa, Ph.D.
General Motors Corporation

John T. McCabe
National Center for Manufacturing Sciences

Kevin Wasson, Ph.D.
Aesop, Inc.

Abstract: For reasons of global competitiveness, the automotive industry demands new spindle technologies that will provide an edge in the quality, production rate, and cost of machined components. A variety of components that the automotive industry needs to manufacture requires different spindle characteristics. To help meet these spindle requirements, a new type of hydrostatic bearing system was developed and refined. The bearing system is versatile and can be used in spindles for operations such as milling, drilling, boring, and grinding. As an application example of the use of the new bearing system, the problem of improving the quality, production rate, and cost for manufacturing sets of closely spaced precision holes in a powertrain component is addressed. This example highlights the advantages offered by the new bearing system, which uses a water-based coolant as the lubricant, in a self-compensating design that simplifies its manufacture. This paper describes the spindle design concept and presents preliminary test results for a cluster head of four closely spaced spindles.

Introduction

In their drive to attain lower-cost six-sigma quality in manufacturing, General Motors Corporation (GM) and Ford Motor Company (Ford) proposed a project to develop advanced spindle technology. They chose spindles because spindles are the most critical mechanical component of a machine tool. Together with the National Center for Manufacturing Sciences (NCMS), GM and Ford assembled a development team consisting of the Torrington Company, a like-minded end user of spindles; two machine tool builders, Olofsson Machine Tool, Inc., and Giddings & Lewis (G&L); a spindle builder, SETCO; two spindle design and analytical companies, Aesop, Inc., and Manufacturing Laboratories, Inc. (MLI); and a lubrication system supplier, ORSCO, Inc.

The team submitted a spindle development proposal that was funded by the National Institute of Standards and Technology (NIST) under the Advanced Technology Program (ATP). The proposal offered to develop two families of spindles: one utilizing fluid film bearings; the other, rolling element bearings. This paper deals with machining spindles with fluid film bearings, the design of which was initiated at the Massachusetts Institute of Technology (MIT)

under a National Science Foundation (NSF) grant [1]. Through a joint industrial development effort with the aid of a Ph.D. thesis at MIT [2], the bearing design was refined [3] and made an integral part of a machining spindle. This spindle concept has features and characteristics that address the requirements of the end users who initiated the project on which this paper is based. (A companion paper in these Proceedings, "Compact, High-Power Spindles," deals with the spindles supported by rolling element bearings.).

End User Requirements

The end-user participants in the spindles development project reflect the needs and requirements of two industries: automotive and bearings. Of the many needs for improved spindle technology, one example application was selected because it has three key characteristics:

- The needs of a much broader class of applications was embodied in this application.

- This application would highlight the advantages of fluid film bearings for other spindle applications.

- This application itself addresses quality and cost issues that were important to the end users.

Bearing Options

Rolling element bearings are the bearings chosen by most builders of machining spindles. Rolling element bearings are attractive to spindle builders because they have a low first cost and are supported by a mature infrastructure that is well-established. However, hydrostatic bearings offer an alternative with advantages that are compelling and growing in popularity. Because hydrostatic bearing surfaces do not contact at any time, wear is negligible and their life is indefinite. In contrast, rolling element bearings start with rolling contact, and may experience skidding contact as the spindle is accelerated or decelerated rapidly during tool changes.

A rolling element bearing consists of a complement of balls, rollers, or needles that roll in raceways. To minimize spindle vibration, each element of the bearing must be manufactured to precise tolerances. With use, the bearing elements wear, and the spindle vibration induced by the bearings increases. Free from contact, hydrostatic bearings maintain low vibration levels indefinitely, which promotes sustainable machining accuracy. When the components of a bearing approach each other during impulse or vibratory loading, the resisting action of a fluid film bearing is substantially greater than a rolling element bearing. This damping effect, which promotes chatter suppression and enhances crash protection, was maximized by the bearing design developed under this project.

Another option is fluid film self-acting bearings. These are bearings that achieve load capacity and stiffness from the relative motion between rotating and stationary elements. At start-up and at low-speed operation, contact occurs between the bearing surface, resulting in wear.

Therefore, this type of bearing is seldom selected for machining spindles. However, a spindle-bearing option currently in use in high-performance machine tools is a hybrid of hydrostatic operation at low speed and self-acting effects at higher speeds. Compared to a hydrostatic bearing, a hybrid design would reduce first costs by avoiding the need for a high-pressure supply pump. On the negative side, hybrid bearings are not as rugged and forgiving as hydrostatic bearings.

Air bearings, another fluid film option, and magnetic bearings were not considered viable selections because neither bearing type has the ruggedness needed for the industrial environment of an automotive factory.

Hydrostatic Lubrication Issues

Advantages notwithstanding, hydrostatic bearings are not widely used in the machine tool industry for three historical reasons:

- High first cost and operating cost. High first cost is due to the cost of the fluid supply system. The fluid film spindle itself and is generally much less expensive to manufacture than a rolling element spindle. High operating cost is primarily due to the cost of running the lubricant supply pump and of maintaining the fluid supply system. However, these pump costs are no different than a rolling element spindle that requires a circulating lubrication system.

- Ergonomic problems. Such problems have been associated primarily with the lubrication system. Most hydrostatic spindles are oil lubricated, which requires seals to contain the oil within the spindle. Even without significant seal leakage, an oil lubricant eventually becomes contaminated and must be disposed of in an environmentally acceptable manner.

- High losses at high speeds. The issue of high shear losses is also related to oil lubrication. Even light oil has a high enough viscosity to cause substantial power consumption at today's state-of-the-art spindle speeds.

To a larger extent, these issues were eliminated in the developed design.

Advantages of a Water-Based Lubricant

To alleviate the problems arising from an oil lubrication system, the new family of hydrostatic spindles was designed to use water-based coolants as the bearing lubricant. The design provides several advantages for machining spindle applications:

- The problem of separating the lubricant from cutting fluid is eliminated.

- The low viscosity of water enables higher spindle speeds before excessive power becomes an issue.

- The high heat capacity of water helps control cutting tool position by minimizing thermal expansion of the rotor.

- A single fluid system design enables through-the-tool coolant to be implemented.

Advantages of Self-Compensation and a Hydrostatic Bearing Design

A typical hydrostatic bearing consists of a plane shaft fit into the bore of a housing that contains pockets surrounding the shaft. A pump delivers lubricant through restrictors and into the pockets. Flow of lubricant out of the pockets is further restricted by the gap between the lands surrounding the pockets and the shaft. The bearing compensates for an increase in load by increasing the pressure in one or more pockets in response to a decrease in the land gap surrounding the pocket. Orifices or capillary tubes are commonly used as restrictors, and these restrictors must be tuned normally to center the shaft in the bearing at zero load.

A new bearing design that does not require external compensators was developed [4] and used on this project. Since pressure compensation in the pocket areas is regulated by the bearing clearance distribution, the principle is called self-compensation. Figure 1 compares a conventional hydrostatic bearing with a self-compensated bearing. In addition to the absence of small-diameter restrictors, the self-compensated bearing offers other advantages:

- Machining is less expensive because it is performed on the shaft, not in a bore.

- By eliminating the restrictors, piping to restrictors is eliminated, the packaging envelope is reduced, and assembly is simplified.

- Shock load capacity is much larger because the squeeze film area is increased substantially.

- By eliminating pockets, the potential for turbulence and cavitation are decreased.

- When a load is applied, the self-compensation scheme channels high-pressure lubricant from the unloaded side to the loaded side, causing the pressure to rise rapidly, thereby increasing the bearing stiffness.
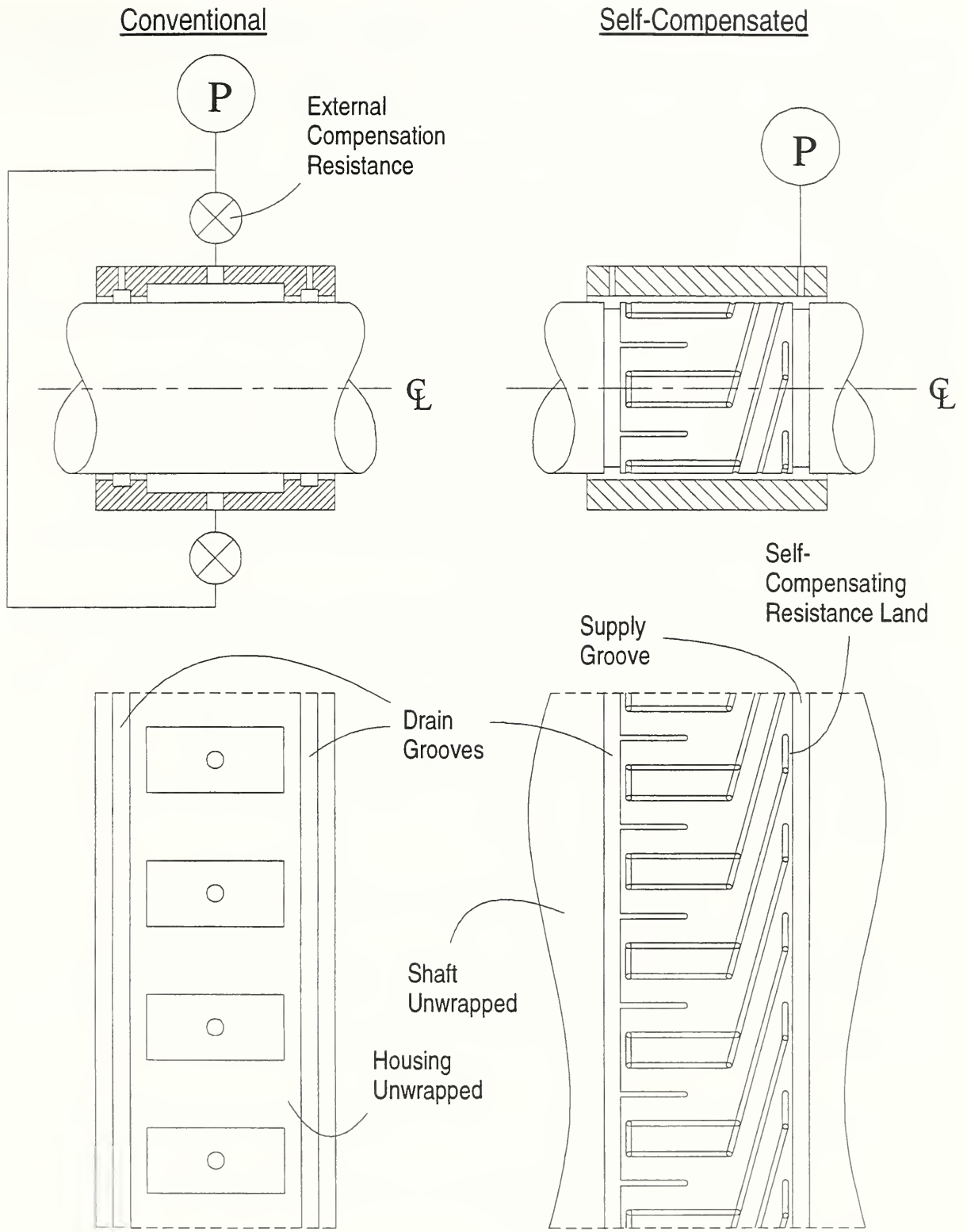
Figure 1. Comparison of a Conventional Pocketed Hydrostatic Bearing with a Grooved, Self-Compensated Bearing

Self-compensated bearings offer advantages for a number of spindle applications. The selected application for producing multiple precise holes simultaneously makes efficient use of the above advantages. Among these advantages, spindle stiffness is the parameter that has primary influence on the potential for reducing the number of process steps.

Application Objectives

The automotive industry requires precise sizing and placement of hole patterns in a variety of production parts. To meet productivity demands, heads containing assemblies of multiple spindles are used to perform drilling and reaming operations with single passes. The current spindles are not capable of producing quality holes with single passes; a series of progressively finer drilling and finishing operations must be performed.

An example of a part that is processed by a series of progressively finer operations is a planetary gear carrier. This part has an upper and lower deck spaced approximately 50 mm (2 in.) apart. Four gears are mounted on shafts with ends fixed to holes in the upper and lower decks. The center-to-center distance between adjacent holes is 61.02 mm. The meshing precision of these gears depends on the concentricity of the upper and lower holes and on the accuracy of their location.

Due to variations in material distribution in the part, the radial cutting force can be as high as 890 N (200 lb). With this cutting load and with the closeness of the holes, sizing the bearings and shaft to machine-finished holes in one pass is currently not possible. Consequently, bushing plates close to the workpiece guide long cutting tools. The project objective is to eliminate the bushing plates and reduce the number of operations from three to one. To reach this goal, the bearing and shaft stiffness must be maximized. Simultaneously, a tool is being developed to take advantage of the stiffness increase by drilling, reaming, and countersinking to produce a finished hole in a single pass.

Importance of Stiffness

Stiffness measured at the spindle nose, which relates directly to machining accuracy, depends primarily on two interactive components: the shaft and the bearings. How much they interact can depend on the type of bearing selected. Ball bearings are currently used in the selected application, whereas the new design proposes hydrostatic bearings.

Figure 2 compares the layout of the existing spindle and the new hydrostatic spindle. The existing spindle is supported by a pair of 35-mm (1.518 in.) bore duplex ball bearings near the spindle nose and a 30-mm (1.181 in.) duplex ball bearing at the rear. By substituting fluid film bearings for the ball bearings, the shaft diameter can be increased to 56 mm (2.205 in.), resulting in a substantial increase in bending stiffness.
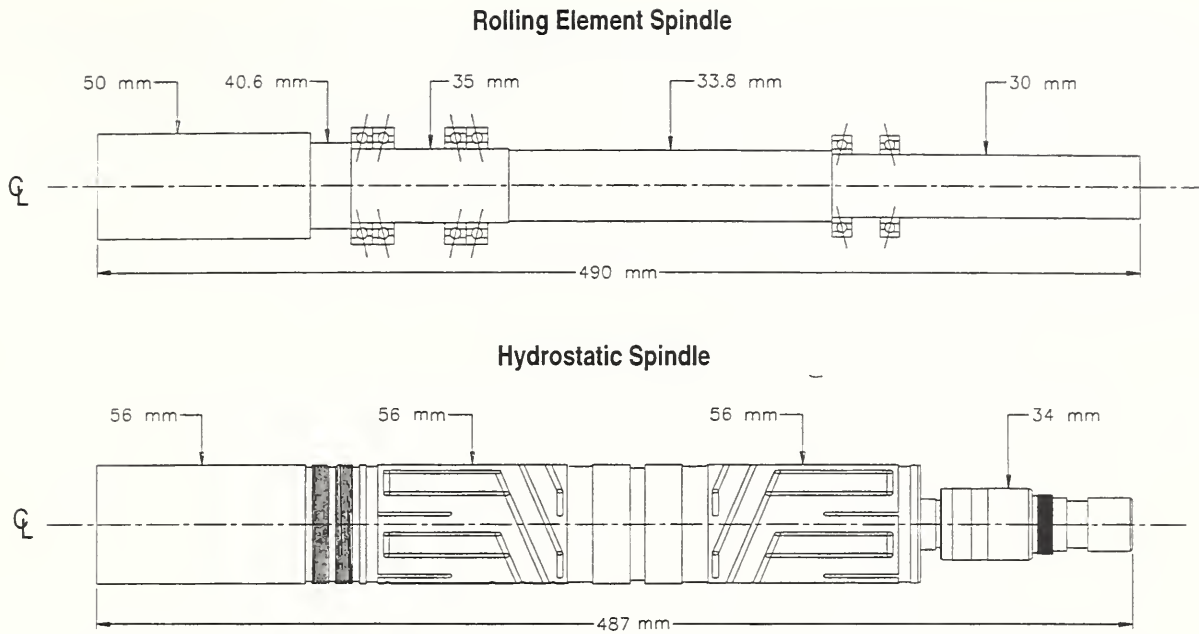
Rolling Element Spindle

Hydrostatic Spindle

Figure 2. Layout Comparison Between the Existing Rolling Element Spindle and the New Hydrostatic Spindle

Figure 3 compares the end view of a ball bearing design to a fluid film design. Both sets of spindles are spaced 61.02 mm on centers. However, the hydrostatic shaft diameter is 60% larger than the rolling element shaft.
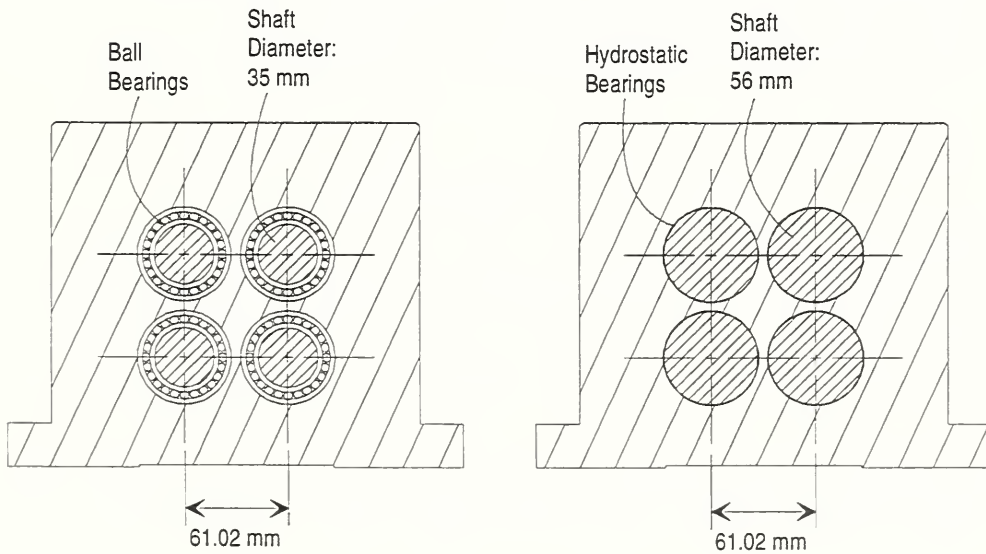

Figure 3. Cross-Sectional Comparison Between Ball Bearing and Hydrostatic Bearing Spindles

Figure 4 shows the relative deflection of the ball bearing and hydrostatic spindles 102 mm (4 in.) from the spindle nose where cutting occurs. To eliminate tool bending as a variable, the shaft extension of both spindles has a minimum diameter of 50 mm (2 in.). These deflection curves, which take into account the interaction of shaft bending on the bearing stiffness, show that the hydrostatic design will reduce tool deflection by a factor of two.
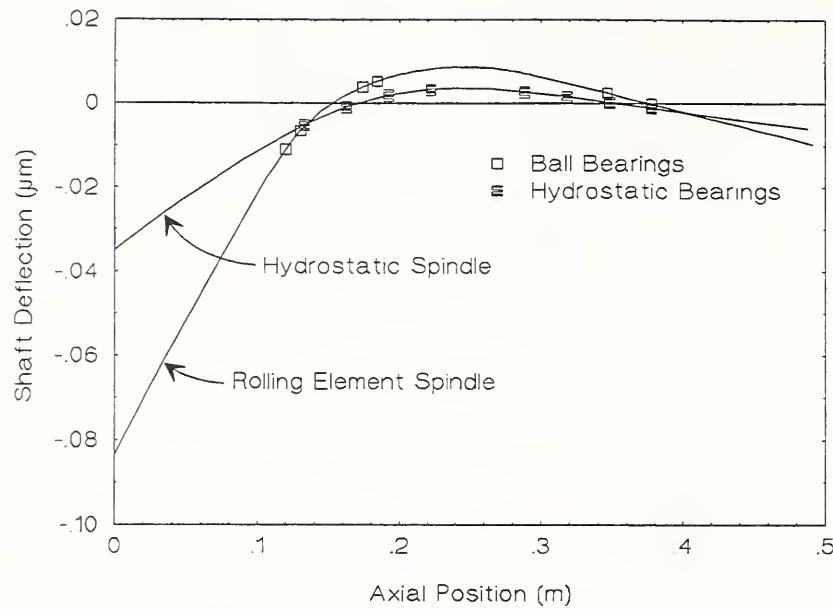
Figure 4. Deflection Comparison Between Ball Bearing and Hydrostatic Bearing Spindles

To compare the ball bearing spindles with the hydrostatic spindle, the bearing and shaft stiffness were computed. The stiffness of the front bearings of the ball bearing spindle are each 145 N/μm (827,805 lb/in.); the stiffness of the rear bearings are each 102 N/μm (582,318 lb/in.), and the spindle stiffness at the tool tip is 11,977 kN/m (68,400 lb/in.). The static stiffness of each hydrostatic bearing is 1,489 N/μm (8.5 million lb/in.). The stiffness of the equivalent hydrostatic spindle is 28,716 kN/m (164,000 lb/in.) using the designed supply pressure of 4,137 kN/m^2 (600 psi). The hydrostatic spindle is therefore 2.4 times more stiff than the rolling element spindle it will replace.

Design Validation Methodology

To validate the enabling technologies for spindle design, a series of prototype spindles were built and bench tested. These tests, which focused on component technology, are described in subsequent sections of this paper. Following the component tests, a fully integrated 4-spindle cluster assembly was built and mounted on an Olofsson machine tool. This unit is currently undergoing tests with various tools to characterize the process capabilities of the new cluster design. When these tests are completed, the unit will be installed in a powertrain plant and evaluated by plant personnel. This paper describes the bench tests of two prototype spindles and presents the current design of the fully integrated 4-spindle cluster.

Prototype Spindle No. 1

Prototype No. 1 Design Features

The first prototype fluid film test spindle, shown in Figure 5, has all of the bearing detail machined into the surface of the shaft with a ball nose end-mill. Virtually all details are loosely tolerances grooves. By encircling areas on the shaft surface with shallow grooves, the compensator areas and the high pressure load support areas of the bearing are defined. The mechanism in

this prototype for accomplishing a self-compensated bearing was by cross-drilled holes that connected the compensator areas with load support areas. The radial bearings were formed by the grooved shaft and a straight bore in the housing. The main thrust bearing was formed by a flange at the shaft nose and a rear housing. The rear housing contained a self-compensated hydrostatic radial bearing to support the drive pulley load. To minimize costs, machining elements, such as a toolholder interface were not included in the design. The spindle shaft was extended forward 102 mm (4 in.) beyond the spindle housing to facilitate deflection measurements.
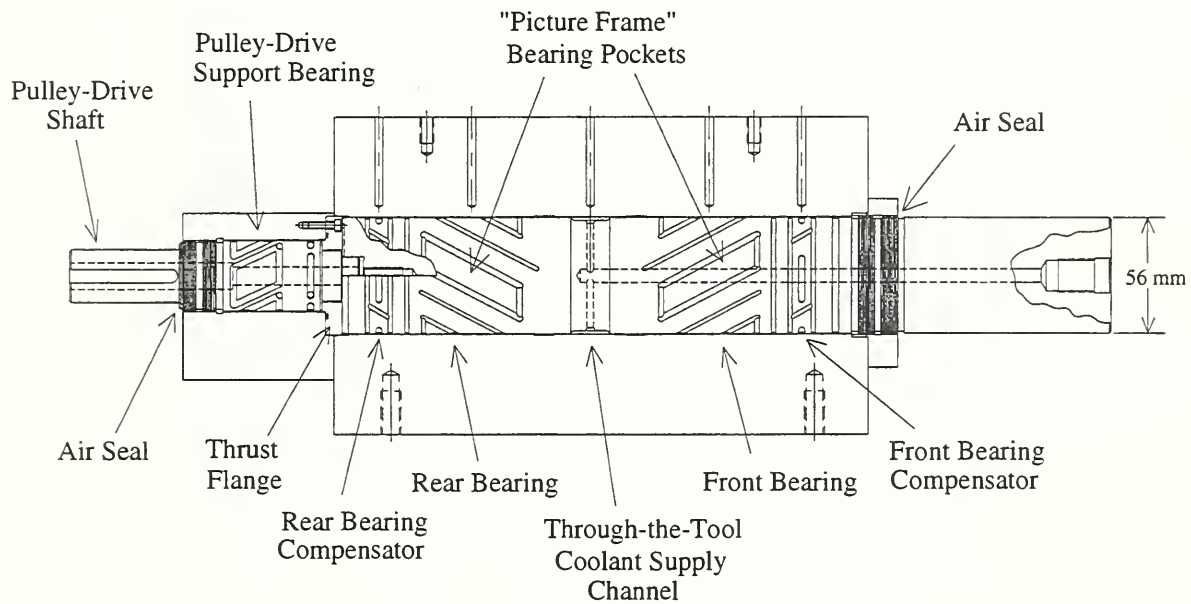

Figure 5. Prototype No. 1 Test Spindle With Cross-Drilled Self-Compensated Radial Bearings

Radial Bearings. Each radial bearing was designed to support 16.46 kN (3700 lb) of radial load when supplied with water at a pressure of 600 psi. With a radial clearance of 0.015 mm (0.0006 in.), the bearings each have a static stiffness of 1.489 kN/micron (8.5 million lb/in.). The pulley bearing was designed similarly to the main radial bearings. Its purpose is to support up to 445 N (100 lb) radial force from drive belt tension and provide vibration isolation between the belt drive system and the cutting tool.

The radial bearings, which were produced by milling grooves into the surface of the shaft, produce a "picture frame" pocket in which a central region of the pocket has the same clearance as the other bearing land regions [5]. This innovative technique was applied for two reasons: (1) to maximum the squeeze film damping capability of the bearings, and (2) to minimize the frictional shearing of the water in the bearing region. The flow of fluid in the land regions of the bearing remains in the laminar flow regime over the entire operating speed of the spindle (0 to 20,000 rpm). However, because the grooves are deep (0.4 mm), the flow of fluid through the grooves becomes turbulent at a speed near 2,000 rpm. By keeping most of the flow in the pocket

regions laminar, the frictional shearing of the fluid is minimized when the spindle is operating at high speeds.

Cross-Drilled Shaft. In the first prototype spindle (see Figure 6), the fluid was routed from the compensators to the pockets on the opposite side of the shaft through holes that were cross-drilled through the shaft. The holes did not have a significant effect on stiffness and they provided a convenient way to route the fluid. However, these holes did add cost and an addition process step to manufacturing the spindle so a better solution was evolved for the second prototype.
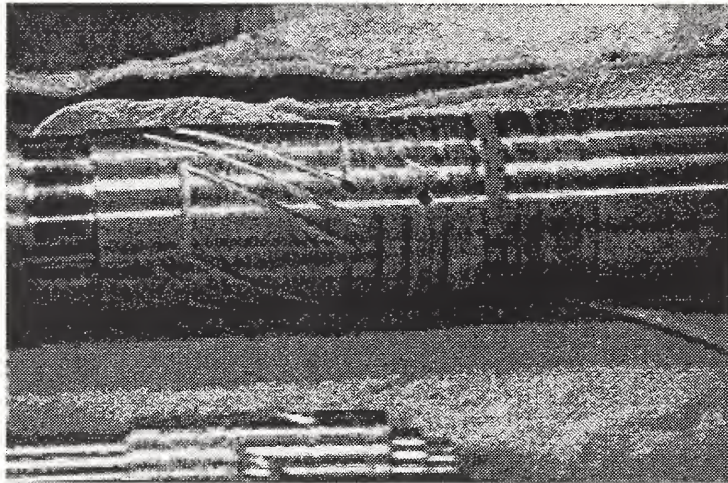

Figure 6. Prototype No. 1 Test Spindle Shaft Showing Cross-Drilled Holes Through the Shaft From the Compensators to the Load-Supporting Pockets

Thrust Bearing.[6] The thrust bearing was designed as a self-compensated bearing with a supply pressure of 4,137 kN/m^2 (600 psi). For drilling operations, the main thrust face can support axial loads up to 4.48 kN (1,000 lb). For tool pull-out, a smaller reverse thrust bearing provides a load capacity of 890 N (200 lb).

Air Seals. Non-contacting air seals in the front and rear of the spindle prevent lubricant leakage regardless of the spindle orientation. These seals, which are pressurized with shop air at 138-207 kN/m^2 (20-30 psi), prevent foreign matter from entering the bearing cavity.

Through-the-Tool Coolant Supply. Coolant is supplied to a groove in the center of the spindle and then routed through holes drilled radially to the center of the shaft and then to the front face of the spindle. This design eliminates coupling wear, a troublesome failure mode that accompanies the more common rotary coupling, which is used in today's state-of-the-art spindles.

126

Prototype No. 1 Bench Test Results

Static Radial Stiffness. Static stiffness was determined by loading the shaft at one point and measuring shaft deflections with a digital indicator. A dead weight of 111 N (25 lb) was placed 102 mm (4 in.) from the spindle nose. With the bearing supply pressure set at 4,000 N (580 psi), the static stiffness calculated from the deflection measurement at the applied load was 23,025 kN/m (131,500 lb/in.). Not counting the compliance of the tool itself, this stiffness indicates the radial deflection due to the bearings and shaft that can be expected at the tip of the tools used to machine the demonstration part.

Comparison with Theoretical Stiffness. An important goal of spindle testing was to verify the spindle design theory. With a reliable theory, the next generation spindle could be optimally designed with high confidence. The theory, which includes the effect of shear deformations and the nonlinear response of the bearings to bending shaft, was developed under the MIT doctoral thesis cited in Footnote 2. The theory predicted the radial stiffness of the static load test to within an error of less than 5%, providing a high confidence level in our ability to forecast this critical aspect of spindle design.

Dynamic Radial Stiffness. Dynamic radial stiffness tests were performed with and without the shaft rotating. For these tests, the spindle was excited with a load cell hammer and its resulting displacement was measured with a capacitance probe. Both the force and displacement measurements were made at a location on the shaft 102 mm (4 in.) from the spindle nose. The resulting frequency response spectrum is shown in Figure 7.
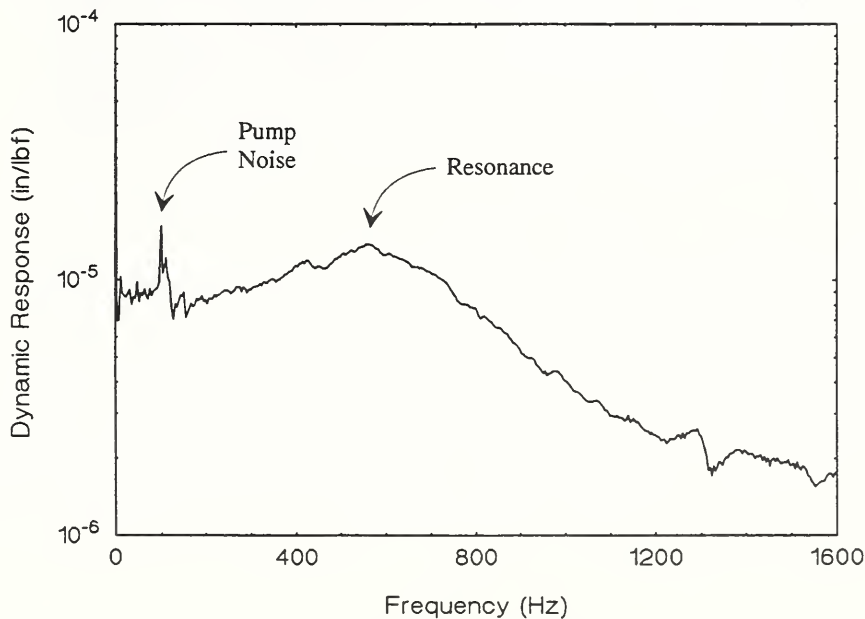

Figure 7. Dynamic Response of the Unrotating Spindle Measured at a Point 102 mm (4 in.) From the Spindle Nose

The natural frequency of the spindle is about 550 Hz. The dynamic stiffness at this peak is 12,782 kN/m (73,000 lb/in), which is high, but not uncommon with a well-damped hydrostatic

spindle. Figure 7 also shows a spike at 200 Hz caused by pump pulsation. In the subsequent prototypes, the problem was eliminated because it detrimentally impacts spindle error motion.

Spindle Error Motion. Figure 8 shows the setup used to measure spindle error motion. A small steel ball is mounted at the spindle nose and capacitance probes measure the motion of the ball as the spindle rotates [7]. Error motions were first measured with the spindle drive disabled (no rotation) and the pump supplying the bearings with a full pressure and flow. The radial and axial displacement were 0.76 μm (30.0 μin.). This contribution to the spindle error motion, which remains approximately constant with spindle speed, is the result of pressure pulsations in the bearing supply fluid caused by the piston pump used in these tests. A new screw pump was subsequently installed and the pulsations were eliminated.
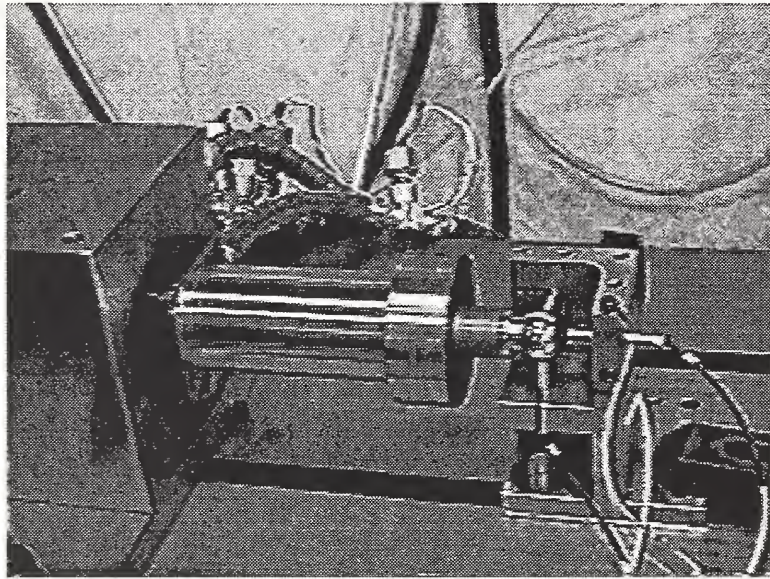

Figure 8. Experimental Setup to Measure the Dynamic Runout of the Spindle

The radial error motion was next measured at various spindle speeds. The spindle was externally driven with a ball bearing drive spindle connected in-line to the test spindle with flexible couplings. Table 1 shows that the maximum total radial runout of 1.94 μm (76.5 μin.) was experienced at a rotational speed of 7,000 rpm. The total runout was 1.70 μm (67.0 μin.) at 10,000 rpm, which is low compared to a comparably sized rolling element spindle. The low runout is achieved in a fluid film bearing because there is no mechanical contact, and minor perturbations are substantially dampened by squeeze-film effects in the bearings. With proper attention to the fluid system and bearing geometry, the error motion could be reduced to less than 0.25 μm (9.84 μin.).

Table 1. Measured Error Motion of the Test Spindle With an In-line Drive System and Flexible Couplings

Speed, (rpm)	Asynchronous Runout, μm (μin).	Average Runout, μm (μin.)	Total Runout, μm (μin.)
1,000	0.46 (18.3)	0.38 (14.9)	0.74 (29.0)
2,000	0.57 (22.5)	0.45 (17.6)	0.79 (31.2)
5,000	0.70 (27.5)	0.58 (23.0)	0.92 (36.2)
7,000	1.55 (61.2)	1.21 (47.7)	1.94 (76.5)
10,000	1.37 (54.0)	1.03 (40.6)	1.70 (67.0)

In the last error motion test, a belt-and-pulley drive system drove the spindle, as it would in the final design. The measured radial runout of the spindle using a belt tension of 69 N (15.5 lb) is shown in Table 2. A maximum total radial runout of 1.00 μm (39.3 μin.) was experienced at a spindle speed of 5,000 rpm. This is only marginally higher than the 0.92 μm (36.2 μin.) runout measured at this speed with the in-line drive system. This fact illustrates the effectiveness of the damping in rear pulley shaft bearings in isolating belt vibrations the rest of the spindle.

Table 2. Measured Error Motion of the Test Spindle Driven With a Belt-Drive System

Speed, rpm	Asynchronous Runout, μm (μin.)	Average Runout, μm (μin.)	Total Runout, μm (μin.)
920	0.71 (27.8)	0.42 (16.5)	0.83 (32.7)
1052	0.63 (24.8)	0.37 (14.4)	0.83 (32.6)
2729	0.62 (24.6)	0.59 (23.4)	0.95 (37.5)
2844	0.89 (35.2)	0.58 (22.8)	1.00 (39.2)
4995	0.63 (24.9)	0.57 (22.3)	1.00 (39.3)

Bearing Temperature Rise. This test compared the predicted and measured temperature rise of the flow between the bearing inlet and outlet ports. The bearing design analysis, which includes turbulent flow effects, forecasts the heat generated in the bearings. Figure 9 compares the predicted and measured temperature rise for spindle speeds up to 18,000 rpm. The temperature rise at low speeds is simply that generated from pumping 4,137 kN/m^2 (600 psi) fluid over the bearing lands to atmospheric pressure. The maximum temperature rise of the coolant due to viscous shearing of the fluid bearing clearance was 12.5°C at 18,000 rpm. This value is very close to the 12°C temperature rise predicted by the design analysis.
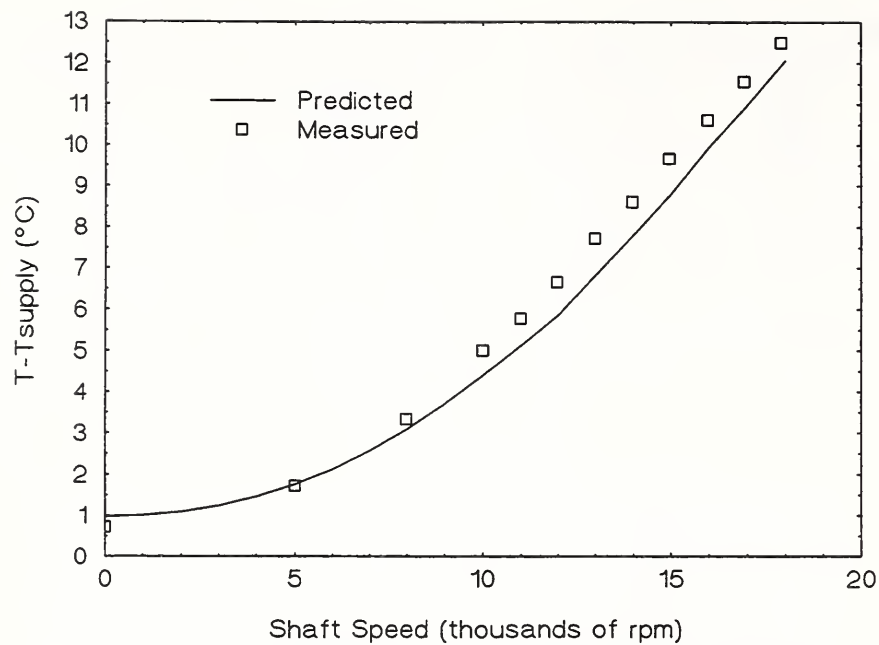
Figure 9. Measured Temperature Difference Across the Spindle at Speeds to 18,000 rpm

Through-the-Tool Coolant Supply. As shown in Figure 10, coolant is routed from a port in the housing, across a gap to a circular groove on the shaft. Four holes drilled in the groove allow the coolant to flow inward to a central hole drilled along the centerline of the shaft which terminates at the tool connection. Axial flow restrictor lands prevent excessive axial leakage. Although a small flow of the coolant leaks off over these lands, the majority flows through the radial holes to the center of the shaft.
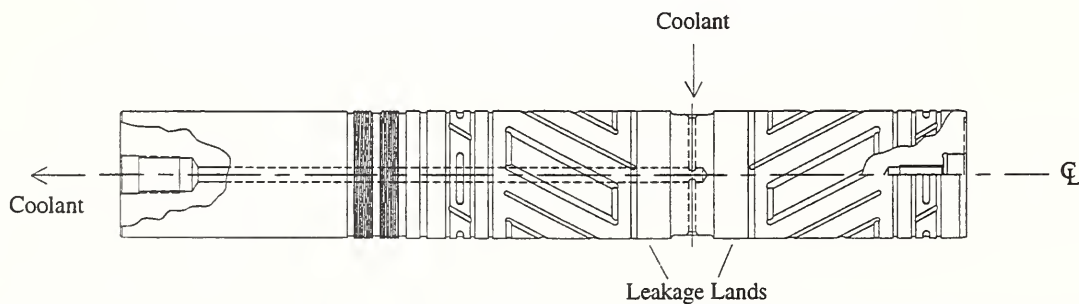

Figure 10. Scheme Used to Route Coolant to the Center of the Shaft for Through-the-Tool Coolant Applications

Since the radial holes rotate, they generate a centrifugally induced pressure. The coolant supply pressure must be high enough to overcome the centrifugal effect and provide the desired flow rate. However the maximum centrifugal pressure is proportional to the square of the rotational speed times the square of the shaft diameter at the hole. Consequently, this simple coolant supply method works best for low speeds and small shafts.

The pressure required to pump coolant to the center of the shaft was measured over a spindle speed range from 0 to 10,000 rpm. At 0 rpm, the supply pressure required to pump 5.75 l/min. (1.5 gpm) was 52.4 kN/m^2 (7.6 psi). This pressure drop was the frictional component required to pump the coolant through the radial holes, through the central supply hole, and

through the hose back to the coolant supply tank. As the shaft speed was increased, the supply pressure required to pump 5.75 l/min. (1.5 gpm) of coolant increased as shown in Figure 11. For the prototype spindle design and the test part, the maximum speed of the 4-spindle cluster head is 3,000 rpm. For through-the-shaft coolant test, the minimum supply pressure is slightly less than one atmosphere.
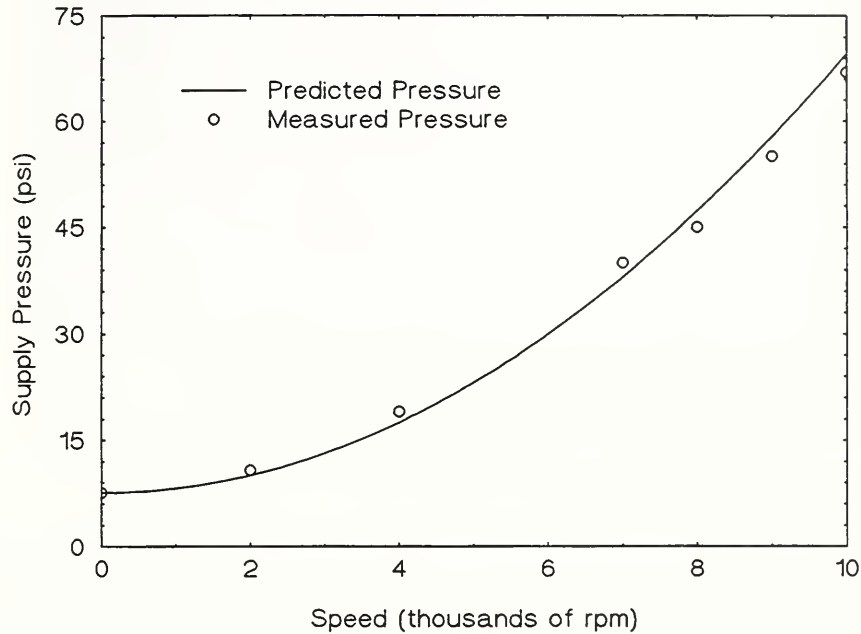

Figure 11. Measured and Predicted Supply Pressures Required to Pump Coolant to the Center of the Shaft, Through a Hose, and Back to the Coolant Supply Tank at 1.5 gpm

Prototype Spindle No. 2

Prototype No. 2 Design Features

The principal difference between the first and second prototypes is the method of implementing self-compensation. Prototype No. 1 used cross-drilled holes through the shaft; Prototype No. 2 used surface grooves. In Prototype No. 2 pressurized fluid flows out of a supply groove and across compensation lands into a set of grooves distributed uniformly around the shaft. The pressure in the groove depends on the pressure drop across a land which, in turn, depends of the bearing clearance at the crossover point. The grooves spiral around the surface of the shaft and encircle an area on the shaft, creating a pressure in that area approximately equal to the groove pressure. These areas of high pressure give the bearing its load-carrying capability. The fluid then flows axially across outlet lands to a drain groove which, in this prototype, is exterior to the housing. As a consequence of the grooving and land geometry, the load carrying areas are self-compensated by the restrictions at the crossover points. This bearing configuration may be seen more clearly by referring to the simplified sketch of a self-compensated bearing (Figure 1). Compared to Prototype No. 1, this design minimizes manufacturing costs and reduces the flow rate of fluid required to pressurize the bearings. (Refer to Footnote 2.)

Figure 12 shows a picture of the shaft, which has a diameter of 56 mm (2.205 in.). In this prototype, coolant is ported to the central supply channel at 4,137 kN/m^2 (600 psi). It then flows

axially across the compensating lands, enters the grooves which route it part-way around the shaft to the load-supporting pockets, then is discharge to a collecting pan. No provisions were made to seal the fluid at the ends of the spindle and no thrust bearing was provided. The primary purpose of this prototype was to evaluate the new compensation concept.
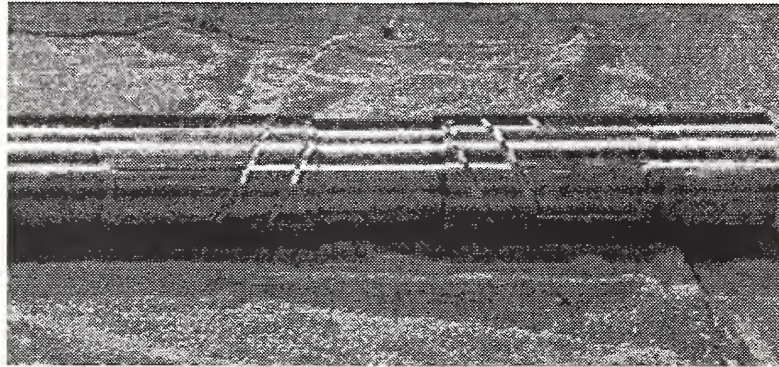

Figure 12. Prototype No. 2 Four-Pocket Bench Test Spindle Shaft

Prototype Spindle No. 2 Bench Test Results

Static Radial Spindle Stiffness. The static stiffness of the spindle was measured by hanging a 111 N (25 lb) weight from the front of the spindle and measuring the displacement with a digital indicator. With a supply pressure of 4,137 kN/m^2 (600 psi), the static stiffness was measured to be 39,800 kN/m (227,300 lb/in.). The calculated stiffness was 40,100 kN/m (229,000 lb/in.).

Shearing Power Consumption. To obtain shear loss data, Prototype No.2 was run up to 20,000 rpm. The torque required to drive the spindle up to 20,000 rpm was measured using a dynamic torque sensor flexibly coupled in-line between the drive motor and the spindle. The measured drive power consumed by the spindle is shown in Figure 13 . The predicted shearing power calculated using turbulent fluid flow analysis (included in Figure 13) is in good agreement with the measured shearing power at all speeds. Although a power loss of 9 kW at 20,000 rpm is high, the bearing system was designed for 3,000 rpm where the power loss is 0.15 kW. If the application had actually required a speed and shaft diameter corresponding to 1 million DN or greater (diameter in mm times spindle speed in rpm), then the spindle would have been modified to yield lower viscous power losses.
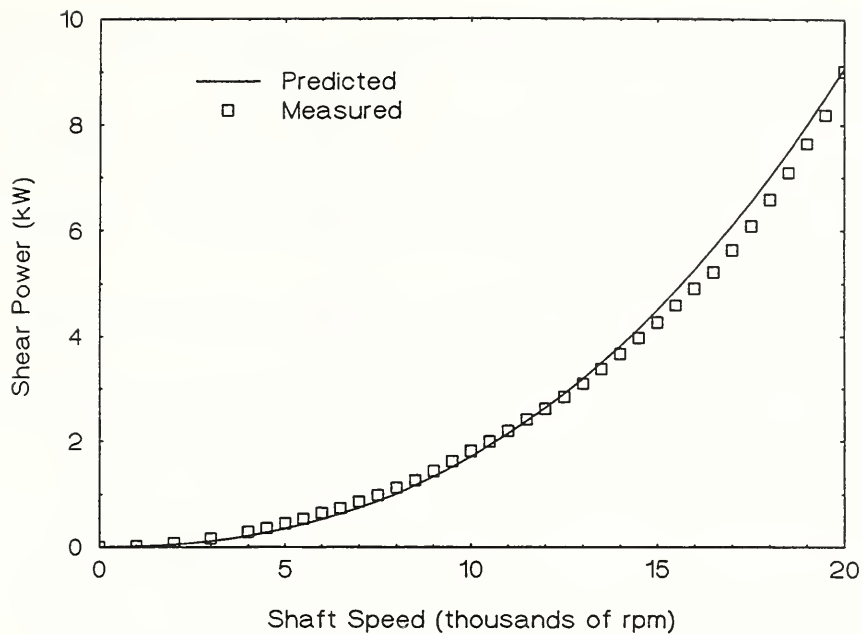
Figure 13. Measured and Predicted Shearing Power Required to Turn the Four-Pocket Surface-Routed Spindle Up to 20,000 rpm

Prototype Cluster Spindle Assembly

Figure 14 shows the cluster spindle assembly developed under this project. The four spindles are supported by hydrostatic bearings that use the same surface groove self-compensating method tested on Prototype No. 2. The design consists of a cluster of four spindles that are belt driven by an AC motor. Each spindle is fitted with a pulley and the pulleys are driven in pairs by two belts. Analysis indicated that the pulley load support bearing could be eliminated which simplified manufacturing and assembly, so these bearings were eliminated from the cluster prototype. Other features include:

- The grooves are oriented to pump more fluid through the bearings with increasing shaft speed, thereby reducing the temperature rise of the spindles.

- The thrust bearing is of the self-compensating type. This thrust bearing is designed to support a 4,448 N (1,000 lb) axial force in the pushing direction and 890 N (200 lb) force in the pulling direction when supplied with a water-based coolant at 4,137 kN/m^2 (600 psi).

- To measure the pressure of the fluid in the thrust bearings, a pressure transducer is incorporated to provide a direct measurement of the axial forces exerted on the spindles.

- Through-the-tool cooling is provided with non-contact coupling.

- The spindle can use both Futura and Hydra-Lock tool clamps.

- The four spindles are driven by a 7.5 kW (10 hp) variable speed motor.

- The spindle speed can be varied between 1,000 and 4,000 rpm.

- Non-contact, 138 kN/m^2 (20 psi) air seals are used to isolate the front of the spindles.

25 HP
900 rpm
Motor

Belt
Drive
System

Thrust
Bearing

Rear
Radial
Bearing

Front
Radial
Bearing

Hydra-Lock
Tool-Holder
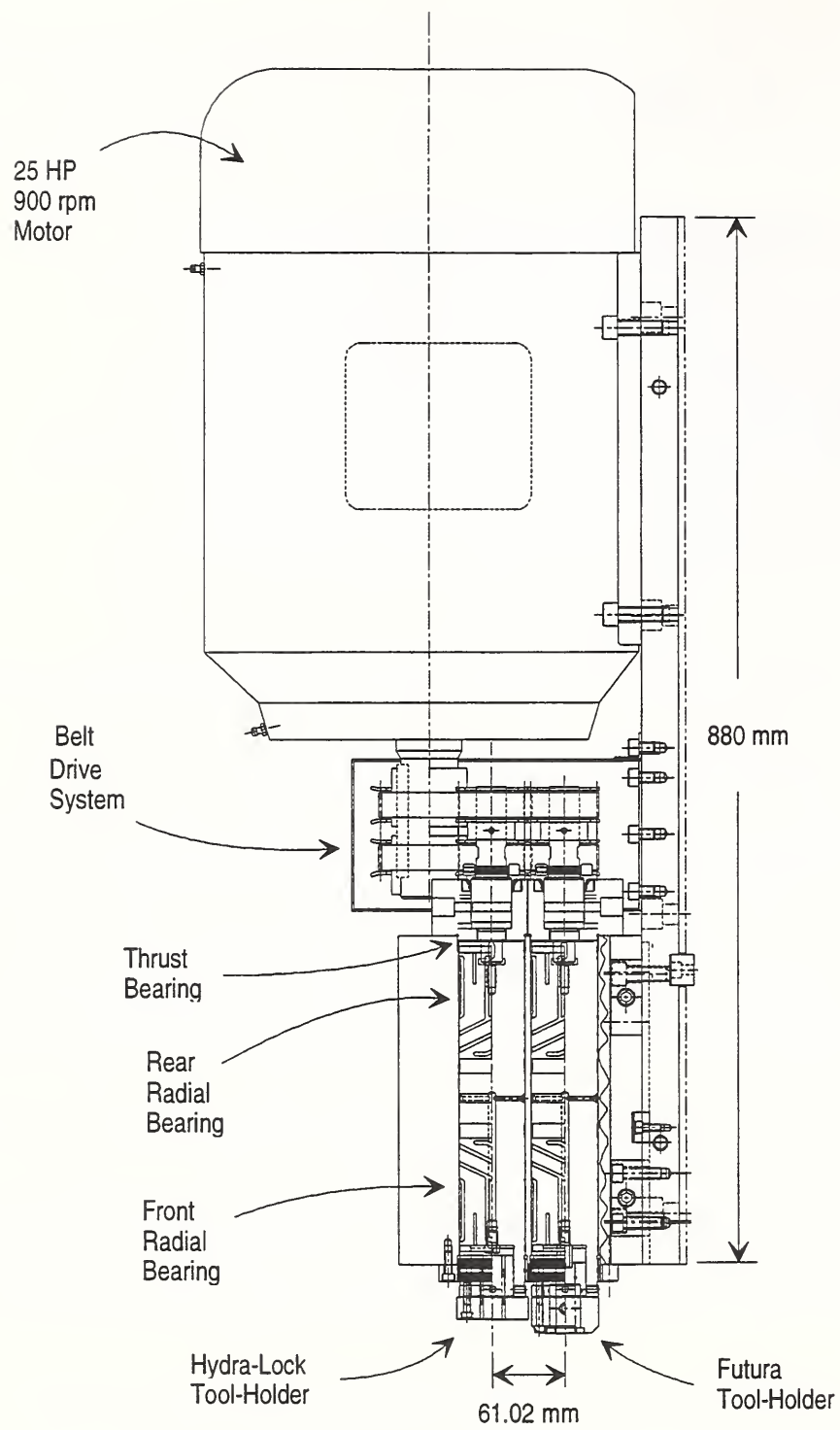
Futura
Tool-Holder

880 mm

61.02 mm

Figure 14. Four-Spindle Cluster Prototype Assembly

Table 3 lists the radial stiffness, thrust capacity, temperature rise, and flow rate as a function of supply pressure. This table indicates the following:

- Increasing the supply pressure from 2,758 kN/m^2 (400 psi) to 4,137 kN/m^2 (600 psi) increases the radial spindle stiffness by only about 15%. However, 4,448 N (1,000 lb) of thrust capacity is needed for this application, so a supply pressure of 4,137 kN/m^2 is required.

- The temperature rise is less than 3°C at 3,000 rpm, the maximum operating speed of the spindle.

- The total lubricant flow rate at this mean bearing clearance is 18.25 l/min (4.76 gpm) at 4,137 kN/m^2 which requires only 1.25 kW of pumping power.

Table 3. Effect of Supply Pressure on the Performance of the 4-Spindle Cluster

Supply Pressure, kN/m² (psi)	Spindle Stiffness, at 4 in. ext. kN/m (lb/in.)	Spindle Stiffness at Tool Holder, kN/m (lb/in.)	Pushing Load Capacity, N (lb)	Pulling Load Capacity, N (lb)	Temperature Rise at 3,000 rpm, (°C)	Four-Spindle Flow Rate, l/min (gpm)
2068 (300)	24,164 (138,000)	69,865 (399,000)	2580 (580)	534 (120)	2.9	9.35 (2.44)
2758 (400)	26,090 (149,000)	78,795 (450,000)	3425 (770)	712 (160)	2.5	12.27 (3.20)
3448 (500)	27,666 (158,000)	85,974 (491,000)	4270 (960)	890 (200)	2.3	15.18 (3.96)
4137 (600)	28,716 (164,000)	91,402 (522,000)	5116 (1150)	1066 (240)	2.3	18.25 (4.76)

Technical Achievements

New Radial Hydrostatic Bearings. Two new self-compensating hydrostatic radial bearings were evaluated. These bearings have several superior qualities over previous designs including increased bearing stiffness, reduced manufacturing cost, reduced sensitivity to fluid supply system contaminants, reduced fluid turbulence, and reduced possibility for cavitation at high speeds. The bearings were proven to be both mechanically and thermally stable up to a rotational speed of 1.1 million DN.

Water Lubrication.. Water-based coolants used as the lubricant was shown to have ergonomic advantages while reducing effects to the environment and providing superior frictional and thermal characteristics compared to an oil lubricant.

"Picture Frame" Bearings Pockets. A picture frame bearings pocket design, in which the center of the pocket remains at a clearance equal to the bearing land clearance, was shown to offer lower bearing friction, lower temperature rise, and higher damping compared to a conventional hydrostatic bearing with deep pockets.

Reduced Manufacturing Cost. The picture frame bearing geometry, with all its machined features placed on the shaft, combines with a simple straight bore in the housing and results in a spindle that is less expensive to manufacture than any previous hydrostatic spindle.

Increased Radial Spindle Stiffness. A 4-spindle cluster was produced in which the radial stiffness of each spindle at the tool tip is 2.4 times the radial stiffness of the rolling element spindles it is intended to replace.

Increased Shock Loading Capability. The prototype spindle was shown to be capable of sustaining repeated dynamic impacts of 667 N (150 lb) 102 μm (4 in.) from the spindle nose while rotating at 10,000 rpm. This impact force resistance, which would fail the rolling element spindles it replaces, provides the robustness needed for production applications.

New Bearing and Spindle Design Software. Computer programs were developed to design and analyze self-compensating hydrostatic bearings for spindles, including the effect of shaft deflections on bearing performance. These programs, which were verified by comparing their predictions with experimentally measured characteristics (as shown in Figure 15), allow spindle designers to quickly and efficiently design hydrostatic spindles for new applications.
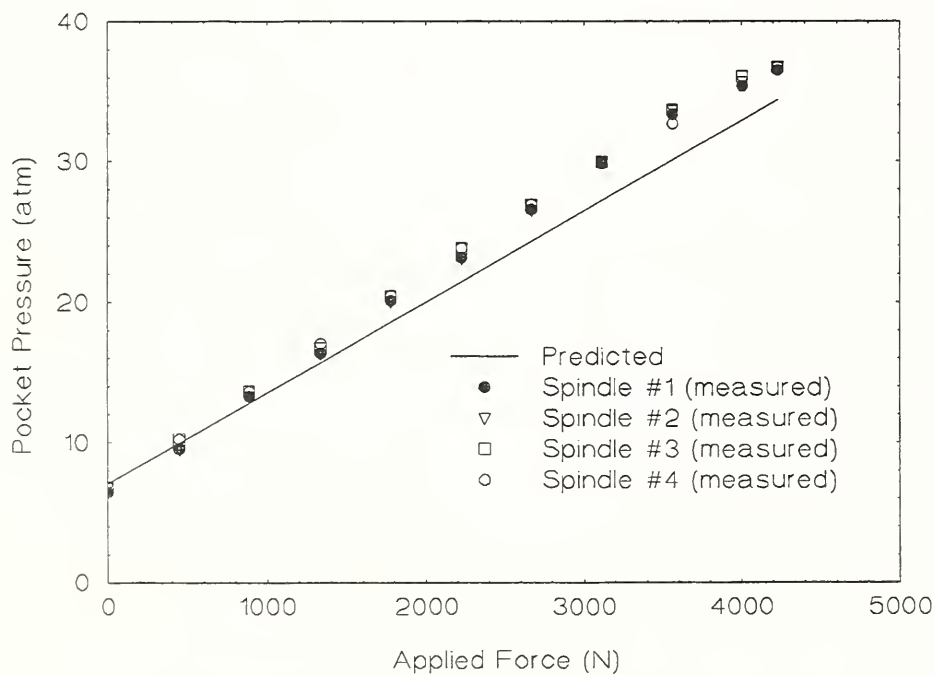

Figure 15. Predicted and Measured Pocket Pressure Versus Applied Force

New Hydrostatic Thrust Bearing. A self-compensated hydrostatic thrust bearing was proven effective at resisting the high axial forces that will be placed on the spindles during drilling operations.

Direct Force Measurement. By integrating a pressure transducer into the spindle assembly to measure the pocket pressure of a thrust bearing, the axial drilling forces during machining can be measured directly.

Air Seals. Air seals were integrated into the cluster spindle assembly and proven to prevent spindle leakage even when the assembly was mounted in the difficult to seal vertical nose-down configuration.

New Through-the-Tool Coolant Method. A method was developed whereby coolant can be supplied through a non-contact fluid coupling to the center of the spindle. Unlike rotary couplings, the new coupling does not wear out with extended use.

Summary and Conclusion

A new bearing system for machining spindles now exists that began with National Science Foundation support in a university. The technology was found creditable and became the subject of a Ph.D. thesis. With a strong basis in theory and laboratory experiment, the bearing concept was selected by a collaborating team for further development as a spindle bearing. The development, which was funded under the NIST ATP program, advanced to a point where the technology can be used with confidence in many applications. An example application was chosen that is important to automotive manufacturing—drilling and reaming four precision holes simultaneously. The criteria for success was to produce these holes faster and with better quality than current practice.

A key conclusion of this project is that the rapid movement from a theoretical bearing concept to a marketable commercial spindle requires synergistic collaboration. Rapid evolution requires the interaction among different experts—those who understand the underlying theory; how to build a machining spindle; how to interface a spindle to a machine tool; the part and its use; and the constraints and requirements of a factory. All these capabilities were present on this project. As a consequence, new bearing and spindle technology is available that is understood, well-documented, and will very likely be applied in an array of expanding fields.

Endnotes

[1] NSF Grant DDM-9201906. "Low Temperature Rise Water-Hydrostatic Bearing for Precision Machine Tools." Alexander Slocum and John Lienhard V.

[2] Wasson, Kevin L. "Hydrostatic Machine Tool Spindles." Doctoral Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., May 1996.

[3] U.S. Patents 5,104,237; 5,281,032; 5,466, 071; other U.S. and foreign patents pending.

[4] U.S. Patent 5,281,032.

[5] U.S. Patent 5,466,071; other patents pending.

[6] U.S. Patent 5,533,814.

Ultrasonic Detection of Unstable Plastic Flow in Hard Metal Cutting

M. A. Davies, S. E. Fick, C. J. Evans, and G.V. Blessing
Manufacturing Engineering Laboratory
National Institute of Standards and Technology

Introduction

Recently developed metal-cutting tools, made from such materials as ceramics and polycrystalline cubic boron nitride (PCBN), allow turning and milling operations to be applied to heat-treated steels in the hardened (50-65 Rockwell C) state. Understanding the linkage between controllable machining parameters and the associated mechanical phenomena is difficult because of the extreme complexity of the environment of the tool and workpiece. For most reasonable machining conditions, the flow of material is unstable, leading to oscillations in tool-workpiece stresses, local chip velocities, and temperature. In precision hard turning, with chip dimensions on the order of tens of μm, the oscillation frequencies may approach a few hundred kHz. The goal of this work is to investigate ultrasonic measurement techniques suitable for characterization of material flow during machining, as an aid to understanding the effect of material flow on important process variables and product quality. ·

Chip segmentation due to adiabatic shear localization[1], a material flow instability due to the local heating and subsequent softening of material during deformation, has been observed in many materials[2-4]. Adiabatic shear band formation is accompanied by a rapid decrease in the local shear stresses required to cause further deformation[5-7]. The formation of discontinuous chips by adiabatic shear localization during machining has been found to cause measurable kHz-frequency oscillations in the cutting forces. In this paper we report experimental measurements of shear band formation forces at much higher frequencies.

Experimental Procedure

Hard turning experiments were conducted using a two-axis diamond turning machine with air-bearing spindle to cut 52100 bearing steel through-hardened to approximately 62 Rockwell C. To establish a simplified orthogonal cutting geometry, a special workpiece and tools were used. The workpiece was a ring 94 mm in diameter (at the beginning of the cut) and 0.5 mm wide. At 15° intervals around the ring, radial slots 125 μm deep were ground into the side of the ring to allow measurement of the changes in length of the chips formed during cutting. Lapped custom-ground flat-nosed PCBN tools were used with rake angle -27°. For the results reported in this paper, the depth of cut was 31 μm, total tool motion during a cut was less than 2 mm, and variations in geometry due to tool motion were negligible. The cutting speed V was varied from 0.35 m/s to 4.3 m/s. High-frequency variations in tool loading were sensed by a mylar-backed polymer ($PVDF_2$) transducer assembly clamped between the insert and tool holder. A comparison of chip morphology with and without the sensor clamped in place revealed no noticeable effects on chip segmentation.

Chips were collected from each of the cutting experiments. Each chip was relatively flat in cross section; the surface which had originally been the free surface of the workpiece was always much rougher than the chip surface which had flowed over the rake face of the tool. Segment spacing, d_s, was estimated by microscopic examination of the rough surface of each chip. Since each chip exhibited notches readily attributable to the radial slots in the workpiece, the ratio of its internotch spacing to the circumferential slot spacing at the appropriate workpiece radius was easily calculated to determine r_c, the chip compression factor. Results of earlier experiments performed with both slotted and unslotted workpieces had shown that the presence of slots has little effect on observed chip morphology.

Dimensional analysis invites the definition of a simple parameter, f_s, the segmentation frequency, or number of segments formed per unit time, as the result of dividing the cutting velocity V by the average segment length *during formation*, with the latter determined by dividing d_s by r_c, so that:

$$f_s = V(r_c/d_s)$$

Results

The chips from these experiments showed a clear trend - from closely spaced but disordered segmentation for lower velocities, to wider and more uniform segments for higher velocities. Shown in Figure 1 are scanning electron photomicrographs of three chips obtained with V ranging from 3.1 m/s to 1.0 m/s. Clearly both the magnitude and the uniformity of d_s decreases with decreasing V.

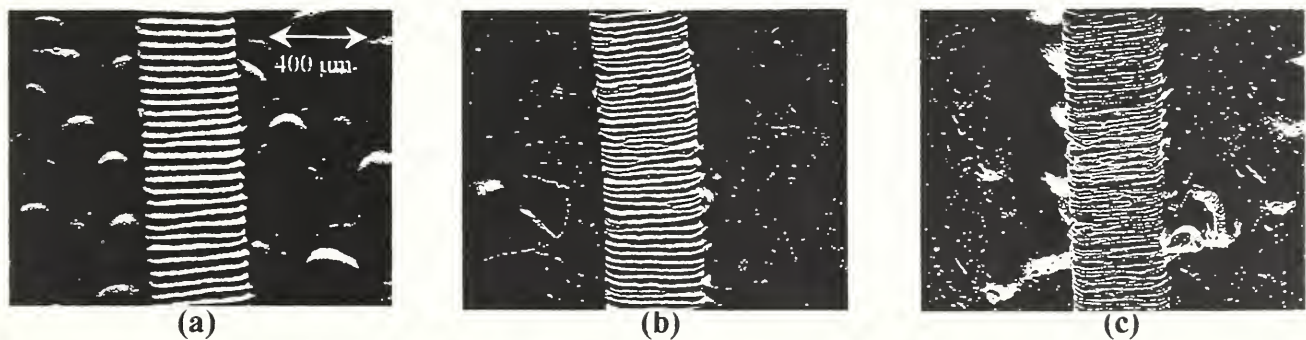**(a)**   **(b)**   **(c)**

Figure 1. SEM photographs of chips obtained from orthogonal cutting with uncut chip thickness 31 μm, rake angle -27 degrees, and cutting speeds: (a) 3.1 m/s, (b) 1.5 m/s, and (c) 1.0 m/s

The sensor output voltage was large enough to obviate preamplification and was measured directly using a digital oscilloscope with sampling rates ranging from 2 to 20 MHz. Based on independent knowledge of the frequency response of the sensor, the sampling rates exceeded the highest anticipated signal frequency by at least a factor of 10. Output voltage waveforms and corresponding power spectra appear in Figures 2,3, and 4.

In Figure 2, a strong component is evident at approximately 120 kHz. This frequency corresponds closely to the mean of the range (103 kHz-136 kHz) of f_s which takes into account the uncertainties of the velocity (4.3 m/s ± 0.1 m/s) for this experiment and the characteristics (r_c =1.6 ± 0.1, and d_s =58 μm ± 3μm) of the chips it produced. The (negatively) saturated part of the signal in Figure 2(a) is the result of the cutting tool traversing one of the radial notches in the workpiece. The periodicity of these notches is the cause of the dominant low frequency peaks in each of the power spectra. These low-frequency peaks were not seen in the data for other experiments using workpieces having no notches. From Figure 3 it is clear that a weaker signal can have an equally dominant spectral peak. Its frequency, about 100 kHz, corresponds to the mean of the range (84 kHz-124 kHz) of f_s which takes into account the uncertainties of the velocity (3.1 m/s ± 0.1 m/s) for this experiment and the characteristics (r_c =1.6 ± 0.1, and d_s =47 μm ± 3μm) of the chips it produced.
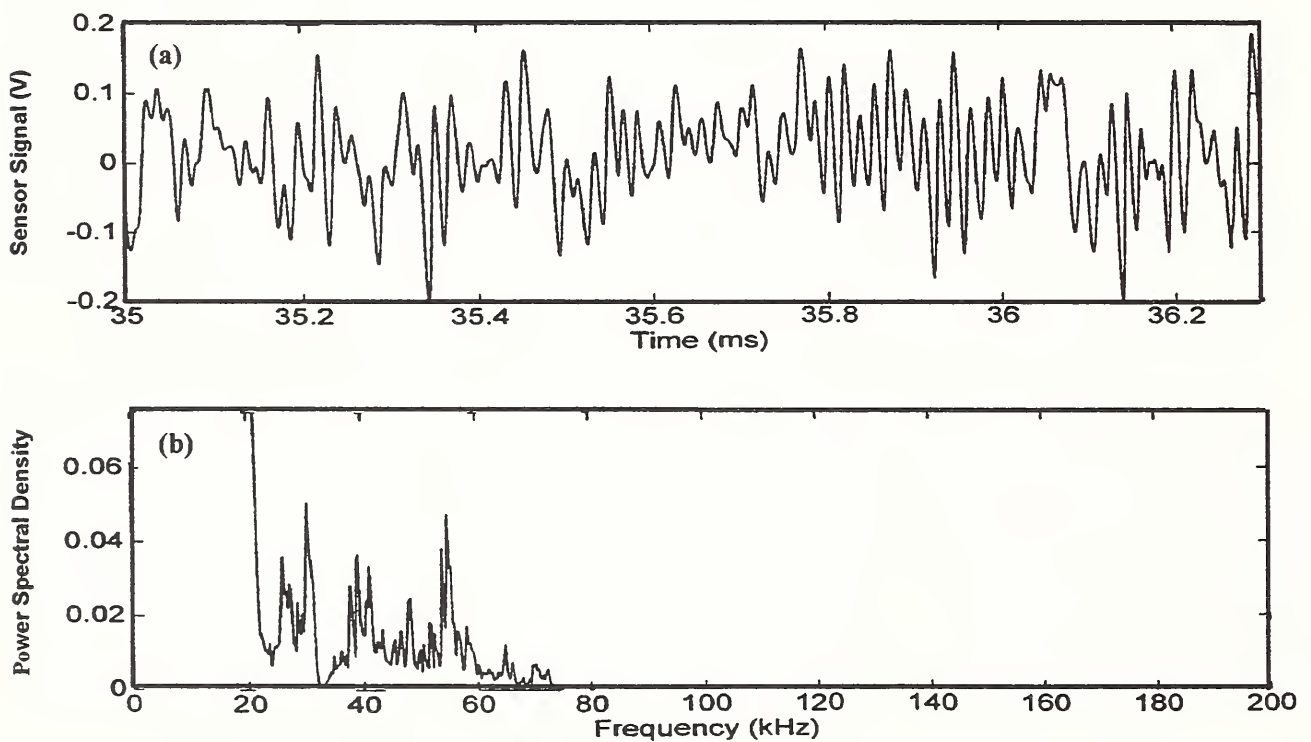

Figure 4. (a) Time trace and (b) power spectrum of sensor signal V = 1.5 m/s.

Results for a lower cutting velocity are shown in Figure 4. The general shape of the power spectrum is consistent with the range (38 kHz-66 kHz) of f_s which takes into account the uncertainties of the velocity (1.5 m/s ± 0.1 m/s) for this experiment and the characteristics (r_c =1.0 ± 0.1 m/s, and d_s =30 μm ± 3μm) of the chips it produced. Examination of the voltage waveform shows numerous sequences of oscillations near 55 kHz. The spectral broadening is consistent with the increased disorder observed in the chips cut at lower velocities.
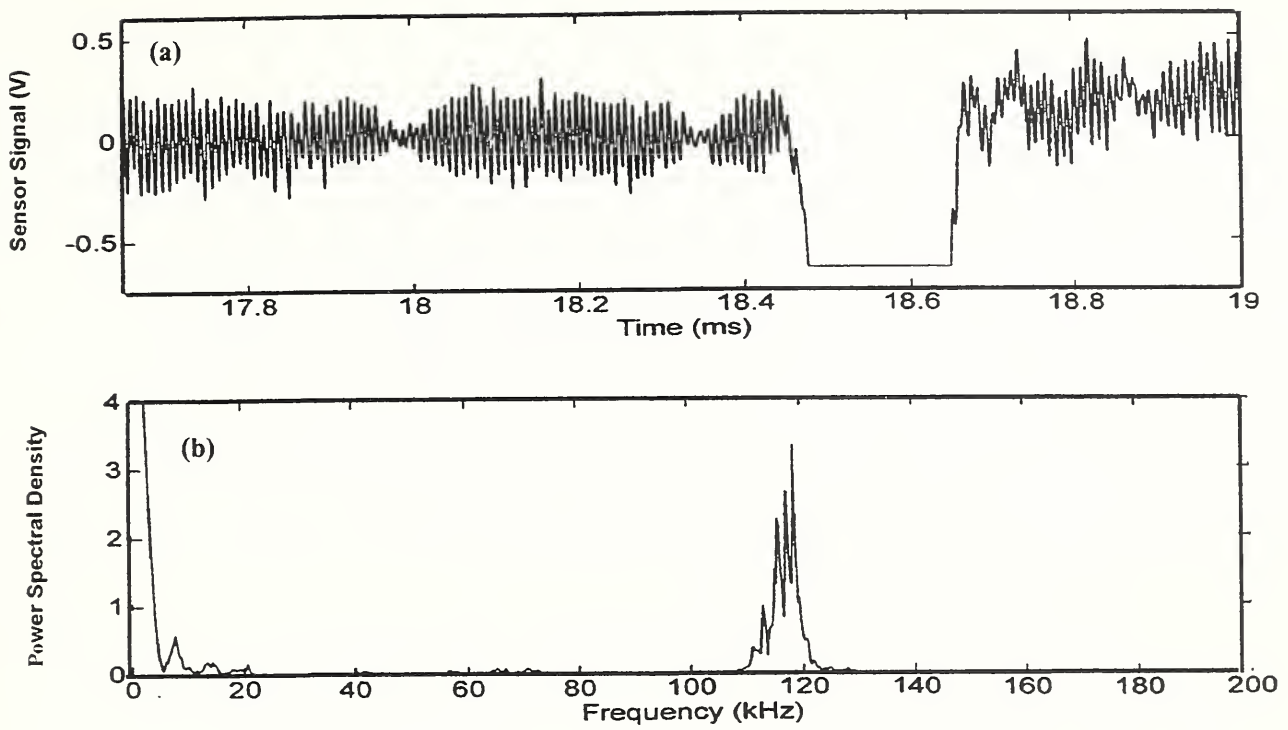
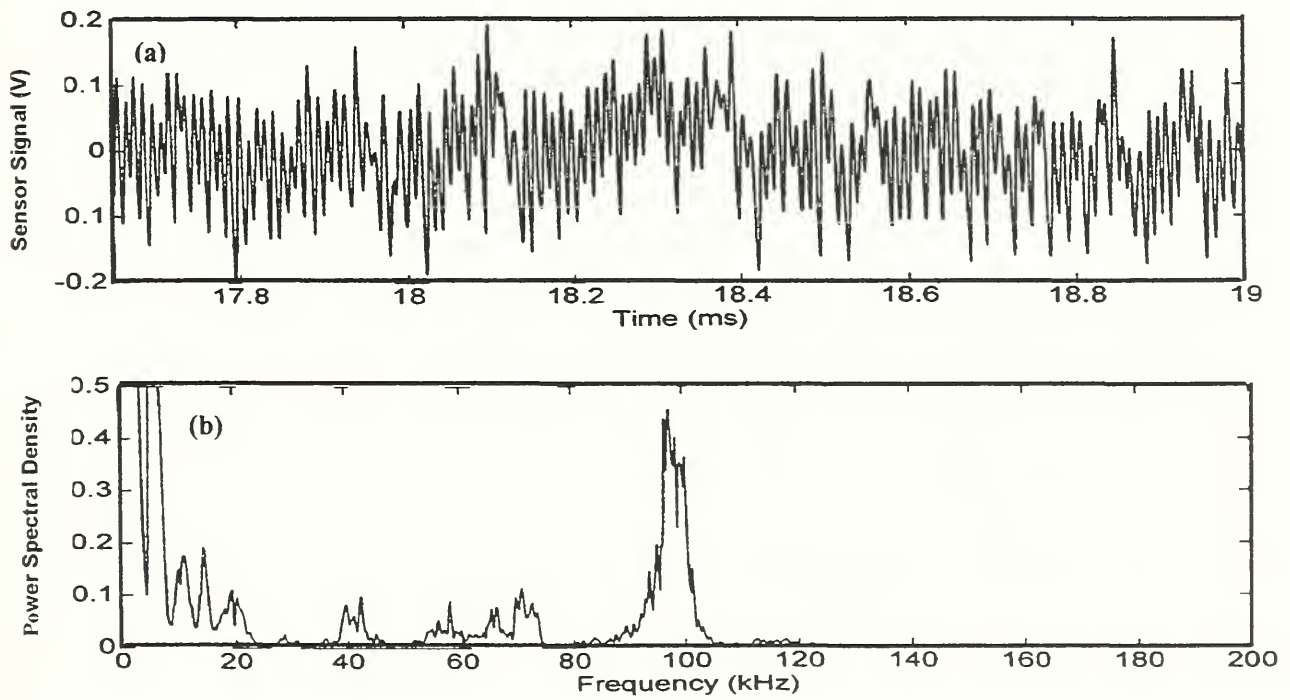Figure 2. (a) Time trace and (b) power spectrum of sensor signal for V = 4.3 m/s.

Figure 3. (a) Time trace and (b) power spectrum of sensor signal for V = 3.1 m/s.

Conclusions

We have demonstrated the use of a simple inexpensive $PVDF_2$ sensor to monitor the segmentation of chips in hard turning by detecting oscillations in tool stress due to thermoplastic instability in material flow during cutting. Several applications of this type of monitoring are apparent. First, because it represents underlying characteristics in real time, sensor data can be used to verify or refute models of material flow dynamics. Second, because results of other experiments indicate that chip segmentation is highly sensitive to cutting edge geometry, it is likely that variations due to wear will lead to measurable changes in the sensor signals. This evokes the possibility of schemes to detect tool wear by examining the spectra of sensor signals that result from segmentation. And finally, data from an improved appropriately calibrated sensor could be used to extend characterization of the dynamic tool environment beyond the envelope of conventional dynamometers. This would enable commensurate improvements of the accuracy and reliability of conventional process control methods.

Acknowledgements

References

1. R. F. Recht, Catastrophic thermoplastic shear, *Journal of Applied Mechanics*. 31:189 (1964).
2. R. Komanduri and R. H. Brown, On the mechanics of chip segmentation in machining, *Journal of Engineering for Industry*. 103:33 (1981)
3. R. Komanduri and R. H. Brown, On shear instability in machining a nickel iron alloy, *Journal of Engineering for Industry*. 108:93 (1986)
4. R. Komanduri, T. Schroeder, J. Hazra, and B. F. VonTurkovich, On the catastrophic shear instability in high-speed machining of an AISI 4340 steel, *Journal of Engineering for Industry*. 104:121 (1982)
5. T. J. Burns, Approximate linear stability analysis of a model of adiabatic shear band formation, *Quarterly of Applied Mathematics*. 43:65 (1985)
6. L. S. Costin, E. E. Crisman, R. H. Hawley, and J. Duffy, On localization of plastic flow in mild steel tubes under dynamic torsional loading, *Institute of Physics Conference, No. 47, Oxford U. K*: 90 (1979)
7. K. A. Hartley, J. Duffy and R. H. Hawley, Measurement of the temperature profile during shear band formation in steels deforming at high strain rates, *Journal of the Mechanics and Physics of Solids*. 35:283 (1987)

Session IV. Abrasive Machining

Flow Control Machining for Automotive Intake Manifolds
R.L. Resnick, S. Rynerson, and D. Yanda, Extrude Hone

Neural Network Modeling of Abrasive Flow Machining
A.E. Smith and W.S. Slaughter, University of Pittsburgh

Computerized Modeling of Electrodeposition Between
Conducting Boundaries Having Variable Geometries
J.R. Bush and M. El-Shazly, Abrasive Technology Aerospace

Experiences in the Grinding of Silicon Nitride on a Lower Cost
High Speed CNC Grinder
T.W. Hwang, E.P. Whittenton, and C.J. Evans, NIST
Manufacturing Engineering Laboratory

Nitrides in Materials Science: Some Thermodynamic Aspects
P.A.G. O'Hare, NIST Chemical Science and Technology
Laboratory

Flow Control Machining for Automotive Intake Manifolds

R.L. Resnick, S. Rynerson, D. Yanda
Extrude Hone Corporation

Introduction

A reciprocating internal combustion engine is in many respects a pump, the function of which is to ingest and compress air to establish ideal conditions for combustion of the injected fuel upon ignition. The passages that control the volume of air ingested are the focus of this project.

These passages and cavities are complex "shapes of flow"—too complex to be economically machined by conventional machining or grinding. They are consequently cast (normally sand cast) and, in modern engines, are most likely to be aluminum. Sand cast combustion chamber cavities typically vary in shape and position by more than 0.25 mm (0.010 in.) and intake ports and intake manifold passages by considerably more than this. In addition, the sand cast surfaces are rough and irregular, generating "boundary layer turbulence" that retards air flow, particularly at the passage walls.

These variations have significant impact on the performance, fuel efficiency and emissions of automotive engines. This effort seeks to develop a nontraditional finish machining method to provide high precision, conveniently adjustable, passage flow capacities. If successful, it will enable building higher performance, more fuel-efficient and cleaner burning engines at lower cost and in lower economical production volumes than currently possible. Once developed, this process would permit the economical and precise adjustment of these functional features on standard cast components to a specifically desired functional performance. This capability would allow a precisely calibrated engine "package" to be "assembled" from standard cast component "modules" that are flow (or volume) "tuned" to be appropriate to a broad menu of niche vehicle configurations.

Abrasive Flow Machining (AFM) is a process used to deburr and smooth the internal surfaces of a wide variety of flow passages. The concept of this process relies on passing a known volume of special flowable, abrasive filled, visco elastic media through the internal passage of a workpiece at a specific flow rate and pressure to attain a desired geometry and surface finish of the passage. The application of this process has been shown to significantly enhance the performance of a variety of engine components.

Extrude Hone Corporation has recently been granted U.S. and foreign patents on a variety of methods of "real-time" Flow Controlled Machining. Through correlation of the flow of the flow machining media to the orifice fluid, a significant enhancement in the performance of engines could be realized. This new process could provide dramatic advantages in the performance and efficiency of many products, particularly aircraft, truck and automotive engines.

This paper outlines aspects of an NIST Advanced Technology Program Project entitled "Flow Controlled Machining – Precision Nontraditional Finish Machining Methods to Improve Motor Vehicle Performance, Economy and Emissions." This is a four year collaborative project

cost on significantly higher quantities. Preliminary efforts on motor vehicle components have shown sufficient results to warrant high expectations and potential economic benefit, but technical barriers need to be overcome to meet the stringent cost requirements of high production motor vehicle manufacturing.

To enable predictive process control of the flow machining process, an understanding of the relationship between the flow of the machining fluid to the flow of the passage functional fluid is paramount. The problem is complicated by the highly non-Newtonian nature of the machining fluid. There is a possibility that no usefully predictive correlation will be found between the passage air flow in use and AFM media in processing. Due to the high complexity of the machining medium "system," coupled with the complex geometry and flow characteristics of the parts to be machined, it is unlikely that an analytic solution to the flow controlled machining process can be reached (at least not without prohibitively high economic costs and a lack of generality). For this reason, the project has an especially strong experimental and statistical component. Extensive data collection on before and after air flows of specific component test pieces coupled with continuous process flow data provides a large information base that, with proper analysis techniques, should permit the derivation of a correlation between air and AFM media. In addition the use of the developed correlation process algorithm on a generous sampling will verify the experimental approach. Subsequent iterative algorithm development on different components in decreasing sample sizes will be directed to reducing the correlation algorithm development cycle, reducing the time and cost of "start-up" for new applications. Extending the range of flow adjustments that can be precisely and cost-effectively achieved is a further objective.

In addition to the risks of the overall technical approach, there are major issues in the basic process that need to be addressed and solved. Monitoring the flow rate of the machining medium in real time to the resolution required provides a significant task. Preliminary efforts monitoring displacement over time have shown promise but substantial improvement is needed. Acoustic emission sensing has shown promise and is receiving considerable development to determine its practical feasibility.

The central objective of Flow Control Machining (i.e., correlation of the flow behavior of the medium in the AFM process with flow bench test data for air and/or viscous Newtonian fluids) represents a formidable scientific problem of high technical risk. While the AFM process has a broad range of commercial applications, neither a comprehensive analytical nor empirical model exists for the behavior of the AFM media. In short, we know the process works, but we don't know (in a scientific and technical sense) *why* it works. Although this research is greatly expanding the fundamental understanding of the AFM process, it is unlikely that a complete analytic, predictive model for AFM will be developed during this research effort. Nevertheless, through a carefully planned and synergistic research effort (encompassing rheology, constitutive modeling, experimental fluid mechanics, and the statistical Design of Experiments (DOE), as well as extensive flow bench databases), the research team believes that a correlation algorithm for a specific geometry and flow range can be built, given a sufficiently large supply of test parts to develop and verify the algorithm. If this effort is successful, the next challenge is to utilize the experience of the initial correlation effort to accelerate and enhance the correlation algorithm development cycle for subsequent component geometries and flow ranges. The ultimate objective is the development of a method and equipment that will be capable of generating engine aspiration

dedicated to developing advanced manufacturing processes to enhance automotive cast passages. The project participants have been carefully selected for an integrated process development approach consisting of:

End Users/Component Suppliers:	Ford, GM
Supply Chain Advisors:	Roush Industries, Uniboring
Equipment Builder/Process Developer:	Extrude Hone Corporation
Research Organizations:	University of Pittsburgh, University of Nebraska

Using Functional Feature Performance for Manufacturing Process Control

Traditionally, process results are verified and process control established by first machining the geometry and then using indirect inspection techniques to verify that a proper geometry is the result. The functional utility of this methodology, used extensively in manufacturing, is based on a number of questionable assumptions. First is the assumption that the functional performance of the part can be described and subsequently inspected to a geometrical representation. In reality, the majority of past design practices have been based on experienced knowledge that has not always been transferred completely or efficiently, leading to design rationales that have no functional basis.

To further exacerbate the problem, procedural tolerance errors are compounded—from the functional performance to the established geometrical representation to the prescribed inspection procedure to the performed procedure to the inspection data analysis. If a small deviation is encountered, especially in the early translation of function to geometry, it can have a significant impact on the final functional requirement of the part. In addition, the functional requirement of a part cannot usually be described by a single, simple geometrical representation, but rather consists of a complex combination of part dimension, geometrical integrity, edge condition, surface integrity and surface finish. The flow resistance of flow passages is a prime example of the difficulty in relating functional requirement to defined geometry. A motor vehicle aspiration passage is comprised of complex geometries with difficult-to-define bends, contours, edge conditions, and surface characteristics. Trying to define and control the manufacture of motor vehicle aspiration passages with geometrical features has been an ongoing task in frustration that is still more art than science.

The central concept of AFM "Flow Control Machining" is that the flow of the "machining fluid" through the flow passage to be "sized" can be used to predict the flow of the orifice fluid (e.g., fuel or air) that the passage will meter in use—essentially, "machining to function" rather than a geometrical representation. Flow conditions are different, of course, just as the flow tests used in production inspection procedures are different from the flow conditions in product use. Nevertheless, correlations can be made. Some useful range of "machining flow" parameters (i.e., machining media consistency, pressure, flow rate, etc. over time) can be found to predict the resultant "in-use" flow resistance for a specific component design.

The hurdles to be overcome in the proposed effort stretch the capabilities of the present process technology. The AFM process has shown significant benefit in those applications that require high precision, high quality, high performance characteristics but at a high price with low component quantities. The challenge is to approach the benefit levels at a significantly reduced

passages with easily adjustable precision flow capacities at low cost from standard, modular cast aluminum components.

The behavior of the AFM process is broken down here into somewhat arbitrary categories to clarify the technical and scientific concerns of the project. These effects are additive in some sense, but their relative significance is not known *a priori*. Furthermore, previous internal research and program research to date indicates that most of these effects "add" in a non-linear way.

Due to the high complexity of the AFM medium "system," coupled with the complex geometry and flow characteristics of the parts to be machined, it is particularly challenging to develop an analytic solution to the flow controlled machining process. Nonetheless, the potential for an analytic solution offers to advance the fundamental understanding of the effects of passage geometry on engine performance, so it is important to continue to pursue such a solution in this effort. Fortunately, in order to build useful correlation algorithms, an analytical model is not required. Empirical models that aid the analytical modeling effort have a strong experimental and statistical content requiring the processing and testing of a relatively large number of components.

These components could be test pieces with configurations similar to engine passage geometries and of like cast material requiring special manufacturing for the project. Fortunately, discussions with Ford revealed a special vehicle program development which might benefit from abrasive flowed components. This program was being conducted by the Special Vehicle Team

(SVT) at Ford in conjunction with Roush Industries to produce an enhanced performance Contour in limited quantities of 5,000 vehicles a year, a similar quantity determined to be necessary for the Flow Controlled Machining project. Preliminary testing by Roush indicated that intake manifolds enhanced with AFM could provide engine performance improvements similar to those required for the SVT Contour. A unique arrangement was initiated with Ford where they provide intake manifolds to Extrude Hone for abrasive flow processing; Extrude Hone processes the parts and collects process and flow data; and Ford in turn provides resultant performance, emissions, and fuel mileage data from vehicles utilizing the processed components.

Figure 1 - 2.8 Liter Ford Manifold being abrasive flow machined under the project

These data are then provided to project partners from the University of Pittsburgh for analysis in two aspects. First, polymer and rheological scientists analyze the data for relationships and insights into research of fundamental abrasive flow processing and its flow rate relationships to functional air flow. Although obtaining an analytical model of the total relationship would be difficult if not impossible to obtain, it is hoped that a relationship can be obtained within the processing regime of the particular components. It is then envisioned that with the limited correlative model and with processing equipment capable of providing adequate process information that subsequent process model development for different component configurations will not be as extensive or time consuming.

Because analytical models that explain a highly non-linear relationship with interactions among process variables are difficult to obtain, artificial intelligence techniques such as neural networks and expert systems have been increasingly used to model complex behavior. Project partners from the University of Pittsburgh expert in neural network development are also using the collected process and test data to develop a neural net model of the abrasive flow process.

In order to accurately control the AFM process acoustic emission monitoring techniques for sensing media flow are being developed by project partners from the University of Nebraska.

Project Progress

An AFM processing cell was constructed consisting of an Extrude Hone "one-way flow" machine; product inspection equipment including a surface roughness measurement system, mass
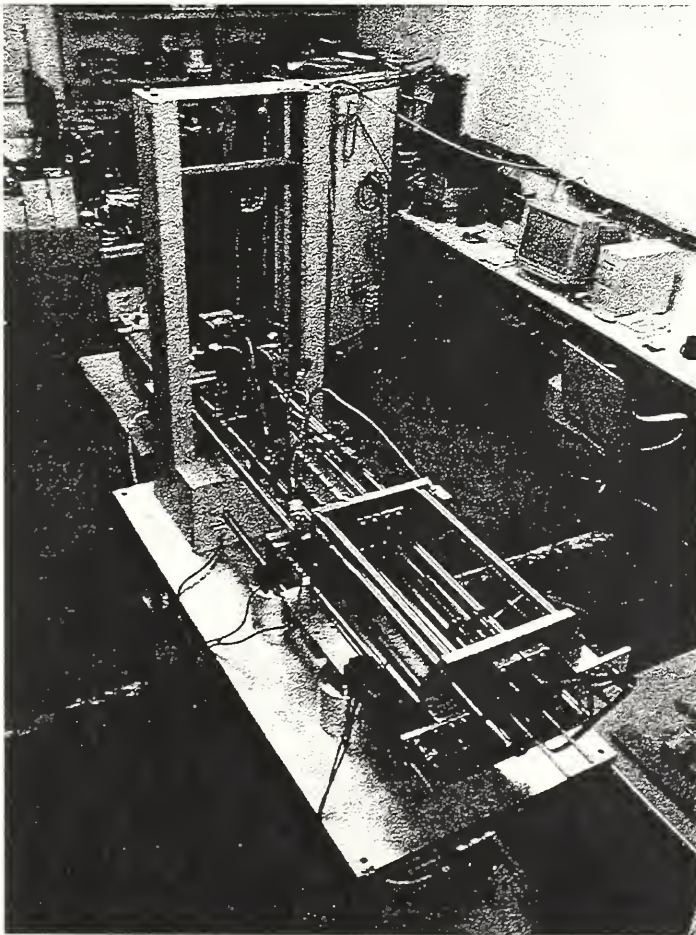
air flow stand, and a digital scale; parts cleaning apparatus; and appropriate part marking and handling equipment. A process was established to receive, identify and track components through the system with appropriate associated process and test data. Internal (LAN) and external (FTP) communication links were established to permit timely and seamless data transfer to university data analysis partners.

To permit a high degree of precision experiments and fundamental analysis of the abrasive flow process, a highly instrumented and controllable AFM test rig was constructed (*Figure 2*). This rig contains precision lead screw-driven media pressurization and receiving cylinders with linear feedback encoders for accurate displacement measurement and control. It also has several pressure/temperature transducers placed strategically throughout the media processing path for media pressure and

temperature measurement and control. Flow rate transducers were also incorporated to monitor media flow rate real-time. The rig is controlled by a PC-based data collection and control system that provides the capability to close the process control loop around any one of the process variable sensor feedbacks and subsequently monitor the remaining variables. The result is a highly controllable machine which permits the analysis of distinct process variables independently. Results from experiments on this machine can then be implemented on the AFM cell with a high degree of confidence and understanding.

Trials were run on a limited run of Ford-provided intake manifolds in a process-test-process-test iterative fashion to determine AFM process limits to reach needed specifications. Initial results indicated that the AFM process could indeed produce the technical objectives of performance enhancement needed for the Ford SVT Contour. This vehicle has been released to the market and has received very favorable feedback from the automotive press.

The project, however, continues to pursue the goal of true flow controlled machining without the need for the iterative process mechanism. Only with this technique will the automotive industry and the commercial public realize the benefit of aspiration passage tuning in an affordable way.

Neural Network Modeling of Abrasive Flow Machining

Alice E. Smith
Department of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA 15261
aesmith@engrng.pitt.edu

William S. Slaughter
Department of Mechanical Engineering
University of Pittsburgh
Pittsburgh, PA 15261
wss@upsm1.me.pitt.edu

Abstract

This paper discusses the preliminary development of a neural network-based process monitor and off-line controller for abrasive flow machining of automotive engine intake manifolds. The process is only observable indirectly, yet the time at which machining achieves the specified air flow rate must be estimated accurately. A neural network model is used to estimate when the process has achieved air flow specification so that machining can be terminated. This model uses surrogate process parameters as inputs because of the inaccessibility of the product parameter of interest, air flow rate through the manifold during processing. The primary project participants are Extrude Hone, Ford Motor Company and the University of Pittsburgh.

Keywords: Abrasive flow machining, neural networks, process monitoring, process optimization, engine intake manifold.

Introduction

The manufacture of precision parts emphasizes final finish machining operations, which may account for as much as 15% of the total manufacturing costs [12, 13]. Abrasive flow machining (AFM) is a nontraditional finishing process that is used to deburr, polish or radius surfaces of critical components. It has been applied in the aerospace, automotive, electronic and die-making industries. AFM can process many selected passages on a single workpiece or multiple parts simultaneously. Inaccessible areas and complex internal passages can be finished economically and productively [12, 13]. AFM is being used to finish air intake manifolds for the Ford Contour SVT [14] as part of this project.

The air intake manifold (Fig. 1) is that part of an engine that directs the flow of air from the throttle body to each of the (in this case twelve) intake valves. Air flows into the manifold through a single large orifice, where it is then divided into twelve "runners" that lead to the intake valves. The complex geometry and internal passages of the manifold dictate manufacture by sand casting, if the component is to be metallic. Injection molded plastic is a generally more expensive option. It has previously been determined that the Ford Contour SVT intake manifold will be an

aluminum alloy.

An optimal intake manifold would deliver precise, predetermined quantities of air to each of the intake valves during each cylinder's intake cycle. To obtain such an outcome for each of the many thousands of intake manifolds to be produced demands an innovative final machining method. Sand casting is not capable of producing finish parts to the high dimensional tolerances required and traditional final finish machining processes are not suitable for the interior passages of the intake manifold. In addition to the low dimensional tolerances of the air flow passages in the as-cast manifold, sand casting leaves rough, irregular surfaces that retard air flow and generate turbulence making the manifold even less optimal. It has been demonstrated that AFM can finish the sand cast manifolds so that precise air flow specifications are met and the roughness of internal passages is greatly reduced. However, the AFM process is not currently economical for mass production of manifolds because the target specification — air flow rate through the manifold with a given pressure differential — is not measurable during processing. This necessitates operator intervention and iterative processing of parts.
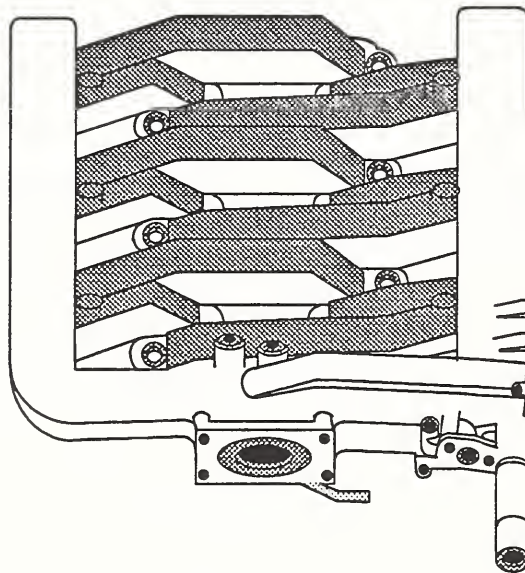

Figure 1. Drawing of Air Intake Manifold.

This paper describes the preliminary development of a neural network-based process monitor and off-line controller for abrasive flow machining of engine manifolds. For a given set of as-cast manifold characteristics and a set of machine parameters this model will predict when the AFM process has achieved the specified air flow requirements for engine intake passages. This project is currently underway and this paper reports preliminary results.

Background

The Abrasive Flow Machining Process

AFM is the removal of material by a viscous, abrasive laden semi-solid grinding media flowing, under pressure, through or across a workpiece. Generally, the media is extruded through or over

the workpiece with motion, usually in both directions. The velocity of the extruded media is dependent upon the principal parameters of viscosity, pressure, passage size, geometry and length [3]. Four types of abrasives are commonly used in AFM. These are aluminum oxide, silicon carbide, boron carbide and diamonds. The AFM process acts in a manner similar to grinding or lapping where the extruded abrasive media gently hones edges and surfaces. It is particularly useful when applied to workpieces containing passageways that are considered to be inaccessible with conventional deburring and polishing tools [2, 8, 11, 13].

Previous research on the AFM process include Fletcher et al. [4], Williams and Rajurkar [16, 17], Williams et al. [18], and Petri et al. [10]. Fletcher et al. studied the thermal and fluid flow properties of polymers used in AFM. They showed that the rheology of the media contributes significantly to the success of the AFM process [4]. Williams and Rajurkar showed that media viscosity and extrusion pressure significantly determine both surface roughness and material removal rate. The authors indicated that the major improvement in surface finish takes place within the first few cycles. Their later work proposed methods to estimate the number of dynamic active grains involved in cutting and the amount of abrasive grain wear per stroke [16, 17]. Williams et al. presented an experimental and qualitative analysis of the distribution of metal removal in multiple hole finishing applications. They also studied metal removal and surface roughness characteristics per cycle for a single hole part and found that the most pronounced change in the bore diameter and surface roughness occurred on the first cycle [18]. Petri et al. developed a predictive process modeling system for the AFM process that relates the critical parameters using strictly empirical techniques, namely neural networks [10]. Their system addressed process settings for AFM for a variety of products and material types. The research in this paper focuses on one particular product type (*viz.*, engine manifolds) but with the demand of precise control to meet stringent specifications.

Neural Networks for Process Modeling

Analytical models that explain a highly non-linear relationship with interactions among process variables are difficult to obtain. Moreover, there are no analytical models that capture the dynamics of the entire AFM process. Artificial intelligence techniques, such as neural networks and expert systems, have been increasingly used to successfully model complex process behavior in areas where analytical models are unavailable.

The use of neural networks is motivated because of their accommodation of non-linearities, interactions, and multiple variables. Neural networks are also tolerant of noisy data and can operate very quickly in software, and in real time in hardware. Statistical models, such as linear regression, require assumptions about the parametric and functional nature of the factors which may or may not be true. Neural networks do not require such assumptions and are data-driven models. Recent work in using neural networks for modeling manufacturing processes include [1, 3, 9, 10, 15, 19].

Model Development

Four major tasks were undertaken to develop the initial model: (1) identification of the key

process variables, (2) data collection, (3) preliminary neural network development, and (4) model validation.

Key Process Variables

The first step was to determine which process variables were critical to the AFM process and should be included as process input parameters to the neural network. Table 1 summarizes these process variables. Some of these variables may not be independent of each other. The development of the process model is an attempt to capture the behavior of both the independent and interaction effects of these variables in order to accurately predict the flow of the air through the manifold. The main categories of process variables are

- **Incoming part** - weight, surface finish, air flow, throttle body diameter
- **Ambient conditions** - temperature, humidity
- **Media condition** - grit, freshness, temperature
- **AFM machine setting** - pressure, number of passes

Table 1. Process Variables.

Process Variables	Definition
INCOMING WEIGHT	The shipped weight of the manifold
THROTTLE BODY DIAMETER	The diameter of the throttle body orifice
SURFACE FINISH	The surface roughness inside the throttle body orifice
RUNNER AIRFLOW	The airflow rate through each individual runner
VARIABILITY OF RUNNER AIRFLOW	The variability of airflow rate among the runners
AMBIENT TEMPERATURE	The temperature in the plant
AMBIENT HUMIDITY	The humidity in the plant
PRODUCTION SEQUENCE	The sequence of the production during the day
MEDIA CONDITION	The condition of the media (cutting ability, contamination level)
HYDRAULIC PRESSURE	The extrusion pressure of the media
MEDIA TEMPERATURE	The temperature of the media
NUMBER OF PASSES	The number of cycles of the AFM machine piston

There is variability among the incoming parts due to the limitations of the sand casting process. The number of passes of AFM processing is done interactively by the operator depending on his judgment of when the manifold reaches air flow specifications. The ambient conditions can impact the condition of the media, and characterizing media condition is important. The media starts new with an amount of grit and no impurities. Over time, impurities enter into the media from the metal being AFM'ed and the grit becomes less abrasive. This affects the AFM process. However, measurement of media condition during processing is not practical. Another change in

media condition occurs daily. The behavior of the media depends partially on its temperature. At the beginning of a day, the media is cold, however after repeated processing, it becomes heated. The production sequence (Table 1) is a rough approximation to this heating effect. Time of production was divided into six periods beginning in the morning (period 1) and ending with the work day (period 6). Each part was assigned to one of these periods depending on its time of production.

The outcome variables of interest are all specific to the manifold:

- **Total air flow**
- **Air flow per runner**
- **Surface finish**
- **Weight**
- **Throttle body diameter**

While all of these outcome variables define the state of the finished manifold, the primary specification is the first one - total air flow through the manifold. It is this specification that the model described here targets.

Data Set

Production data based on a subset of the variables described in Table 1 was collected by the company's technicians. Ninety eight observations were collected on the following seven input variables: incoming part total air flow, incoming part air flow variability over the runners, incoming part weight, total AFM processing time, total volume of media used during processing, average hydraulic pressure during processing and production sequence. The outcome variable studied was outgoing part total air flow.

Preliminary Neural Network Development

The neural network architecture, training parameters, and stopping criteria were selected through experimentation. The transfer function was the unipolar sigmoid and a traditional backpropagation learning algorithm was used because it is known as an universal approximator when used with a nonlinear continuous transfer function with at least one hidden layer [5, 6]. The final network architecture had seven inputs, one hidden layer with seven neurons, and a single output as shown in Figure 2.

Results And Discussion

A five-fold cross-validation approach which used the entire data set for model evaluation was used [20]. The available data was divided into five mutually exclusive groups after randomizing the entire data set. Five networks with parameters identical to those of the final network described above were built, with each using four groups of the data as a training set, and the remaining one as a test set. Figure 3 verifies that the final network has unbiased generalization to all combinations of the input variables used in this study. The x-axis is in units of cubic feet per

minute. Figure 4 shows the five-fold cross-validation networks and their predictions on each 20% test set against the actual observed outgoing average air flow. The x-axis is in processing order, and shows the predicted results of each of the five folds on the hold-out data set (the test set). The test sets are concatentated to produce Fig. 4. The y-axis is in cubic feet per minute of total airflow.
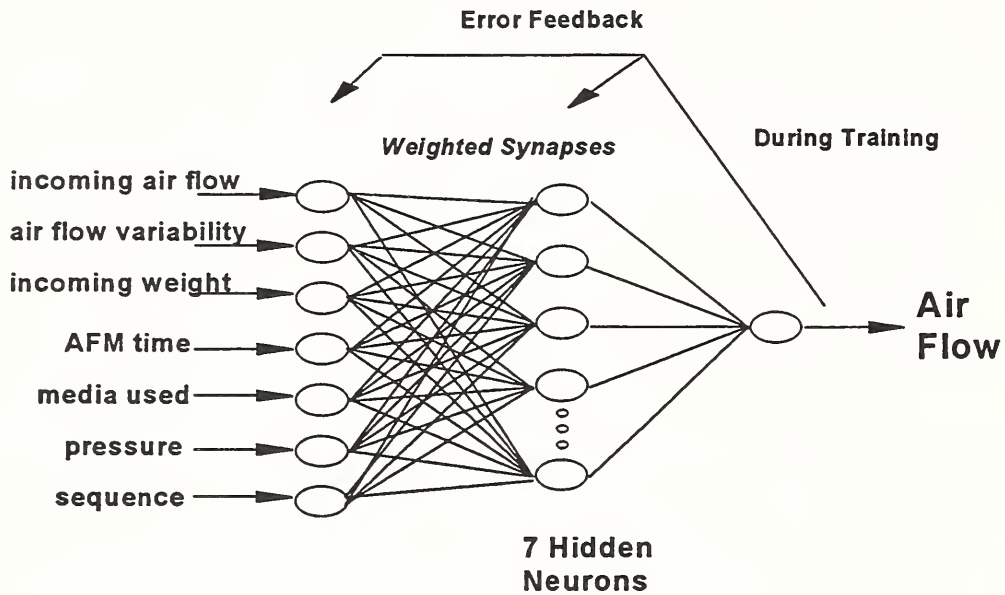

Figure 2. The Neural Network Architecture, Inputs and Output.
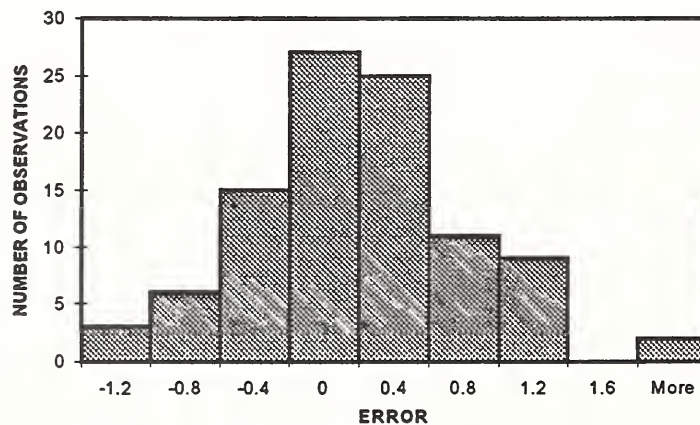

Figure 3. Histogram of Residuals of Five-Fold Cross-Validation.

The final network was able to predict the outgoing average air flow with a mean absolute error of 0.4873 (0.28%) and a root mean square error of 0.6386 (0.37%). The R-squared (coefficient of determination) value of the final network is 0.6527, which means that the network can explain about 65 percent of the variations in the outgoing average air flow. This figure is not high enough for use during production. The R-squared value for a stepwise linear regression using the same data set was slightly less than this value, indicating the appropriateness of a non-linear model.

The model will be improved with additional information from Table 1 that is expected to be monitored. One main area not addressed well by this preliminary neural network is the condition

of the media. The only input describing media condition is the categorical measure of production sequence. This is a very crude surrogate for media condition. It is anticipated that advanced techniques, such as acoustic emissions, may be able to characterize media condition quite accurately and in real time. This will significantly improve the accuracy of the controller. Also, as the data set expands to more observations, the neural network model should improve in precision.
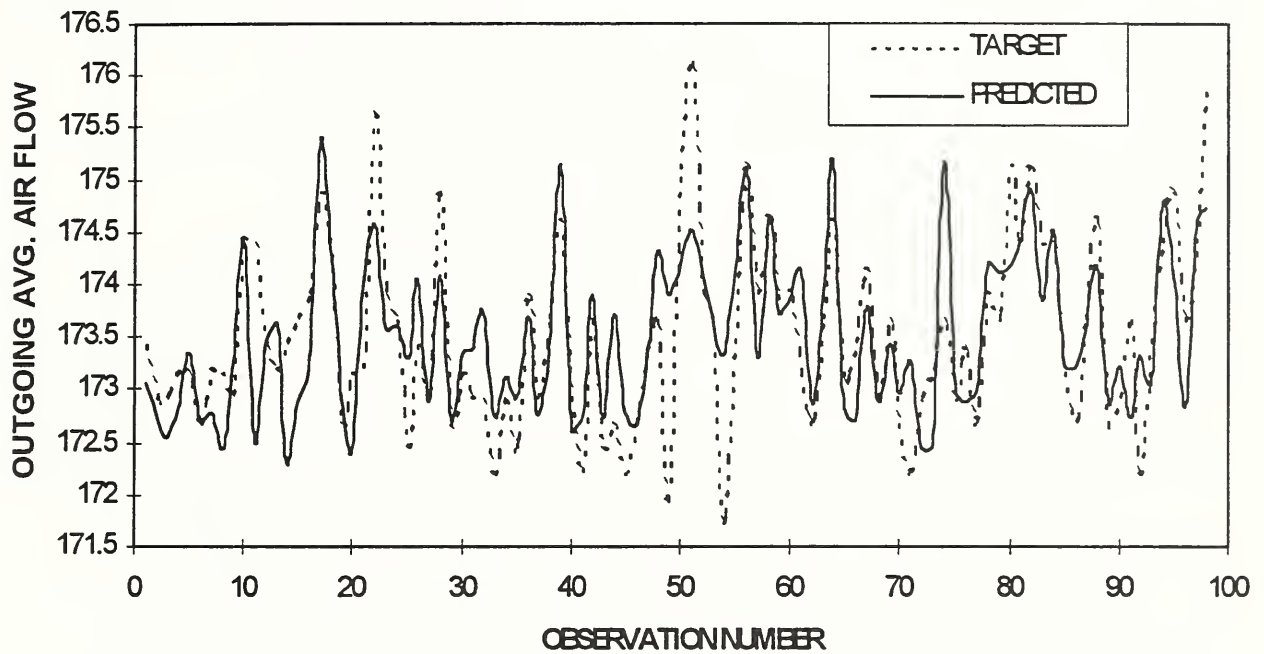

Figure 4. Performance on Five-fold Cross-validation: Predicted Against Target Outgoing Average Air Flow.

References

[1] Andersen, K., G. E. Cook, G. Karsai, and K. Ramaswamy, "Artificial Neural Networks Applied to Arc Welding Process Modeling and Control," *IEEE Transactions on Industry Applications*, 26(5), 824-830 (1990).

[2] Benedict, G. F., *Non-Traditional Manufacturing Processes*, New York, NY: Marcel Dekker Inc., 53-65 (1987).

[3] Coit, D. W. and A. E. Smith, "Using Designed Experiments to Produce Robust Neural Network Models of Manufacturing Processes," *4th Industrial Engineering Research Conference Proceedings*, Nashville, Tennessee, (IIE), 229-238 (1995).

[4] Fletcher, A. J., J. B. Hull, J. Mackie, and S. A. Trengrove, "Computer Modelling of the Abrasive Flow Machining Process," *Proceedings of the International Conference on Surface Engineering: Current Trends and Future Prospects*, Toronto, Ontario, Canada, (Abington, Cambridge: Welding Institute), 592-601 (1990).

[5] Funahashi, K., "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*. 2, 183-192 (1989).

[6] German, S., E. Bienenstock, and R. Doursat, "Neural Networks and the Bias/Variance Dilemma," *Neural Computation*, 4, 1-58 (1992).

[7] Hornik, K., M. Stinchcombe, and H. White, "Multilayer Feed-forward Networks are Universal Approximators," *Neural Networks* 2, 359-366 (1989).

[8] *Machining Data Handbook*, 3rd Edition, Vol. 2, compiled by the Technical Staff of the Machinability Data Center, Cincinnati, Ohio: Metcut Research Associates Inc. (1980).

[9] Martinez, S. E., A. E. Smith, and B. Bidanda, "Reducing Waste in Casting with a Predictive Neural Model," *Journal of Intelligent Manufacturing* 5, 277-286 (1994).

[10] Petri, K. L., R. E. Billo, and B. Bidanda, "Modeling the Abrasive Flow Machining Process: A Neural Network Approach," *4th Industrial Engineering Research Conference Proceedings*, Nashville, Tennessee, (IIE), 239-244 (1995).

[11] Rhoades, L. J., *Abrasive Flow Machining*, Irwin, PA: Extrude Hone Corporation.

[12] Rhoades, L. J., "Abrasive Flow Machining: A Case Study," *Journal of Materials Processing Technology*, 28, 107-116 (1991).

[13] Rhoades, L. J., "Abrasive Flow Machining," *Manufacturing Engineering*, 75-78 (1988).

[14] Smith, Steven Cole, "Four-Door Party Animals," *Car and Driver* 42 (9), 46-56 (March 1997).

[15] Stinson, M. E., O. W. Lee, J. S. Steckenrider, and W. A. Ellingson, "Recognition of Subsurface Defects in Machined Ceramics by Application of Neural Networks to Laser Scatter Patterns," *Ceramic Engineering and Science Proceedings* 15, 390-396 (1994).

[16] Williams, R. E. and K. P. Rajurkar, "Metal Removal and Surface Finish Characteristics in Abrasive Flow Machining," *Mechanics of Deburring and Surface Finishing Processes* (R. J. Stango and P. R. Fitzpatrick, eds.), New York, NY: ASME, 93-106 (1989).

[17] Williams, R. E. and K. P. Rajurkar, "Stochastic Modeling and Analysis of Abrasive Flow Machining," *ASME Journal of Engineering for Industry*, 114 (1), 74-81 (1992).

[18] Williams, R. E., K. P. Rajurkar, and J. Kozak, "Metal Removal Distribution and Flow Characteristics in Abrasive Flow Machining," *Transactions of NAMRI/SME*, XX, 145-150 (1992).

[19] Willis, M. J., C. Di Massimo, G. A. Montague, M. T. Tham, and A. J. Morris, "Artificial Neural Networks in Process Engineering," *IEEE Proceedings, Part D: Control Theory and Applications*, 138 (3), 256-266 (1991).

[20] Wolpert, D. H., "Combining Generalizers by Using Partitions of the Learning Set," *1992 Lectures in Complex Systems* (L. Nadel and D. Stein, eds.), SFI studies in the Sciences of Complexity, Lect. Vol. V, Addison-Wesley (1993).

COMPUTERIZED MODELING OF ELECTRODEPOSITION BETWEEN CONDUCTING BOUNDARIES HAVING VARIABLE GEOMETRY'S

John R Bush and Dr. Moe El-Shazly

abrasive technology aerospace, inc.

GENERAL BACKGROUND

Over the past 25 years, computer-based tools, graphics, and simulation have been successfully applied to a broad range of engineering, manufacturing and consumer applications. For example, computer-aided forging and extrusion die design was introduced to US industry as long ago as 1978. From Nintendo to Forest Gump, the application of computer implemented tools to contempary needs continues at an ever accelerating and demanding pace.

This paper discusses the development of a Computer-Aided Engineering, CAE, tool for rapidly and accurately solving a wide range of electrodeposition problems. This software development project is an integral part of Abrasive Technology's ATP Funded Project: "Rapid Fabrication of Superabrasive Grinding Tools" (95-02-0053). While the application of this software to Abrasive Technology's needs is proprietary, the generic form of this software is scheduled to become commercially available. It is the intent of this paper to describe the commercial version of this program.

A tool is "anything that serves as a means." For the software described in this paper to function as a truly useful tool, the code must meet several functional criteria including:
1. Easily loaded,
2. Rapidly executed,
3. Provides relavent output, and
4. Is accurate!

Within Abrasive Technology's ATP funded project, all of the above expectations are continuously being addressed; both for the generic commercial and AT specific versions. Computer-tool development encompasses two classes of work:
1. Data handling and report generation, and the
2. Technical maturity or sophistication of the mathematical algorithms characterizing the modeled process.

The next section of this paper overviews the electrodeposition process and the associated variables that must be accounted for in the CAE. The objective of this section is to acquaint the reader with the complexity of the scientific phenomena and insight into the flexibility being incorporated into the ATP "Toolset."

THEORETICAL BACKGROUND

Electroplating technology is commonly used in the automotive and the electronic industries to manufacture a wide range of metal and metal-alloy coatings for wear resistance, corrosion protection, decorative finishing, and for manufacturing net shape engineering products by electroforming techniques.

The manufacturing application of electrodeposition frequently requires very precise control of metal coating thickness, typically on the order of 5 - 10%. In order to achieve this high precision, precise control of the various relevant plating parameters is required.

Current state-of-the-art precision plating technology relies heavily on the plating operator (technician) to predict the required thickness using a combination of basic principles (calculations) and empirical techniques. In general, the electrodeposited metal coatings are not uniformly distributed over the workpiece part (cathode). The degree of coating uniformity depends on the current density distribution over the entire profile of the part (i.e., conductive work surface). To achieve better coating uniformity, it is critical to control the current density produced along the plating surface.

In addition to coating thickness, the current distribution is important in regulating the physical and the mechanical properties of the coating. Current density is kinetically related to the overpotential. Therefore, variation in the current distribution will lead to variation in the nucleation conditions and the resulting morphology of the coating surface. Therefore, if the magnitude of the local current density is outside the acceptable range for a given process application, many undesirable effects in the final coated product such as burning, hydrogen embrittlement and nodules formation can occur.

In all electrochemical systems the current distribution -- in the absence of polarization -- is determined solely by the geometry of the electrodes, their position relative to each other and the plating tank geometry. This is referred to as Primary Current Distribution. Polarization is a counter electromotive force which, in effect, increases the resistance to current flow between electrolyte and electrode. The effect of polarization on current flow, is referred to as Secondary Current Distribution. The increase in counter electromotive force (EMF) -- accompanying cathode polarization -- especially resists the flow of current to high-current density areas; such as sharp edges. Accurate use of polarization effects results in a more uniform current distribution over the cathode (conductive work surface). Polarization is dependent on the chemical composition of the plating bath, magnitude of the current density, cathode efficiency, temperature, and mechanical effects associated with agitation. For each plating system the sum total of these effects may increase or decrease the uniformity of the deposit. For example, agitation or an increase in temperature tends to destroy the polarization layers on the

cathode and thereby increase variations in metal distribution. On the other hand, organic additives such as brightners or levelers will generally increase polarization and thereby improve the thickness distribution.

All plating processes operate under a Secondary Current Distribution scheme where the current distribution depends on the cell polarization overpotentials (1). These potentials are commonly considered in terms of the: 1.   Activation Potential, η_a, associated with the kinetics of the electrode process,   2. Mass Transfer Overpotential, η_c, required to overcome transport limitations, and the 3. Ohmic Overpotential, η_Ω, due to electrolyte ohmic resisitivity. Accordingly, the total anode to cathode voltage drop is given by :

$$\Delta V = \text{EMF} + \text{total overpotential}$$

or

$$\Delta V = E + \eta_a + \eta_c + \eta_\Omega \tag{1}$$

Where η_a and η_c are present at the electrode surface and η_Ω is dissipated in the electrolyte across the anode and the cathode gap, E is the standard electromotive force.

The overpotentials are related to the current density, i as follows:

$$\eta_a = \frac{RT}{\beta F} \ln \frac{i}{i_0} \tag{2}$$

$$\eta_c = \frac{RT}{nF} \ln \left(1- \frac{i}{i_e} \right) = \frac{RT}{nF} \ln \frac{Ce}{C_b} \tag{3}$$

$$\eta_\Omega = i \frac{1}{k} \tag{4}$$

where
i	=	current density (amp/cm^2)
β	=	Cathodic transfer coefficient
i_0	=	The exchange current density

$$i_e \quad = \quad \text{The limiting current density}$$
$$K \quad = \quad \text{electrolyte conductivity } (\Omega^{-1}\,cm^{-1})$$
$$L \quad = \quad \text{path length that the current travels (cm)}$$

The current distribution is considered "Primary" when the Ohmic Overpotential dominates in the plating cell; and, therefore depends only on the geometry of the cell. Consequently, since it is difficult to design a practical plating cell where all current paths (L) in equation (4) are equal, most processes operating under "Primary" control conditions yield highly nonuniform thickness distribution.

Commonly, most industrial plating processes operate under conditions where <u>both</u> Ohmic and Activation Overpotentials are predominant over Concentration Over Potential ($\eta_\Omega - \eta_a \gg \eta_c$) and the Current Distribution is considered "Secondary." The effect of the activation resistance at the electrodes in this case, will lead to a more uniform current distribution.

Therefore, the general controlling modes in the plating cell can be summarized as:

	<u>Control Mode</u>	<u>Operating Conditions</u>
a)	$\eta_\Omega \gg \eta_\alpha + \eta_c$	Primary Current Distribution
b)	$\eta_\Omega + \eta_\alpha \gg \eta_c$	Secondary Current
c)	$\eta_\Omega \sim \eta_\alpha \sim \eta_c$	Mixed Tertiary

<u>Wagner Equation</u>

C. Wagner (2) has proposed a useful dimensionless number (W_a) to measure the current density uniformity based on the ratio of the Activation to Ohmic resistance.

$$W_a = \frac{Ra}{R\Omega} = \frac{k}{l}\left(\frac{d\eta}{di}\right) = \frac{k}{l}\frac{RT}{\beta Fi} \qquad (5)$$

where k is the Electrolyte Conductivity ($\Omega^{-1}cm^{-1}$) and ($\frac{d\eta}{di}$) is the rate of change of the electrode Over Potential with respect to Current Density. The characteristic (L) represents the distance from the electrode to a point beyond which the equipotential lines become almost parallel and no longer bend with the shape of the part. This distance represents the effect of varying the geometric size of the cell.

The larger the value of Wagner Number ($W_a > 1$), the greater the <u>uniformity</u> of the current. For $W_a > 1$, the Activation Resistance dominates. On the other hand, when $W_a < 1$, the Ohmic Resistance dominates; and the current distribution is expected to be <u>less uniform</u>.

According to the Wagner (W_a) number, the following conditions will promote uniform deposit distribution:

      a - high plating solution conductivity
      b - low current density
      c - low cathodic transfer coefficient (β)

Due to the Faradric equivalence of the quantity of metal deposited and electrical charge, the principle factor controlling the uniformity of the coating thickness is the current distribution in the cell. The current distribution is governed entirely by Laplace's equation:

$$\nabla^2 \phi = 0 \tag{6}$$

where $\nabla^2 \phi = \nabla (\nabla \phi)$ div (grad ϕ) and potential ϕ is a continuous scalar function of position.

$$\frac{d^2 \Phi}{dx^2} + \frac{d^2 \Phi}{dy^2} = 0 \tag{7}$$

Equation (7) is in a two-dimensional Cartesion form, and is solved based on the boundary conditions.

At insulating surfaces or boundaries, the boundary condition is that no current flows.

$$\frac{d\phi}{dn} = 0 \tag{8}$$

Where n is the direction normal to the insulating surface or wall.
However, at conducting boundaries such as electrodes,

$$\phi = V - E - (\eta_a - \eta_c) \tag{9}$$

where the net current density on the secondary electrode surface is represented by Volmer-Butler equation (1) :

$$i = i_0 \left\{ \exp^{\left(\frac{\alpha F}{RT}(V - E^\circ - \phi)\right)} - \exp^{\left(\frac{\beta F}{RT}(V - E^\circ - \phi)\right)} \right\} \tag{10}$$

Solving for the Secondary Current Distribution is considerably more difficult than obtaining the Primary Current Distribution in a corresponding geometry because the boundary condition at a polarized electrode is a non-linear one; i.e. it depends on the potential ϕ which is sought as a result of calculation and is not a prior known. Typically, a solution would often require an iterative numerical procedure.

There are three advanced numerical approaches to obtain analytical solutions for a complex cell electrode system.

     a - Finite Element Method (FEM)
     b - Finite Differences Method (FDM)
     c - Boundary Elements Method (BEM)

In simple terms, the finite element method obtains a solution by varying the potential throughout the problem domain so as to minimize the potential energy. In the boundary element method a solution is obtained by integrating the effect of the sources imposed at the problem boundary. In both methods a model is formed by subdividing the problem configuration into small elements over which the potential is assumed to vary in some prescribed manner; the simplest being a linear approximation.

The accuracy of the methods depend on the size and density of the elements, particularly in the vicinity of recesses and at the interface between conductive and non-conducting surfaces; since they introduce singularities in the electrical field. Detailed discussion of numerical techniques is beyond the scope of this paper. However, the preceding discussion has been compiled to prepare the reader for the sophistication that is being "built" into the this ATP project CAE tool.

COMPUTERIZED MODELING OF ELECTROPLATING CELL

As stated earlier, electrodeposition techniques offer a wide range of industrial applications. However, due to the empirical nature of the plating cell process, the design process mandates trial and error procedures requiring lengthy and costly set up. In addition, the process is highly dependent on the artisan expertise which is growing more difficult to develop in today's modern manufacturing environment.

The fundamental principles of electrochemistry have been extensively investigated by many researchers over the years. The problem of modeling the electrodeposition process has been difficult to solve and has met with only modest success even for simple geometrical cells. In recent years, solutions have been developed for computer modeling of electroplating cells using numeric methods (3,4, 5).

The first successful computer aided program was developed by U. Landau (5, 6), based on Disk Operating System (DOS). However, this software was successfully limited to academic applications and was not user-friendly as an industrial tool for daily engineering applications.

This next section of the report discusses the comprehensive CAE software developed for predictive modeling of electrodeposition between conducting boundaries having complex geometry. A problem that is typical of many manufacturing processes.

The software is based on a modified finite difference numerical solution of the boundary elements of the potential field which follow Laplace's equation within the electrolytic cell. In this unique software, once the potential field is solved, the current density is determined by a differential expression for Ohm's Law:

$$i = -K \ d\phi/dn \tag{11}$$

The deposit thickness can then be calculated from Faraday's Law;

$$\text{Deposit thickness} = \frac{i.M}{FA\xi} \tag{12}$$

Where

I = The total Current in Amperes.

M = Metal equivalent weight.

F = Faraday constant , 96484.6 $coulombs/equivlent$

ξ = Metal density.

The accuracy of the predictive modeling will depend on the input of the cell's Electrokinetic parameters. Therefore, in order to run the computer simulation program, the process electrochemical parameters (i_o, α, β and κ) must be first determined. Generally, the Kinetic parameters can be independently measured by running polarization curves on a rotating disk electrode technique (1). The conductivity of the plating process can be measured directly by conductivity techniques.

To enter the process chemical and physical properties, the data can be imported from a data base or entered manually. The data base contains a number of common plating chemistries such as copper, chrome, nickel, tin, gold, etc. The data base can be expanded and customized based on the applications required.

The computer algorithm is robust and the software under development is versatile, user-friendly and designed to run on Microsoft Windows NT work stations. Figure 3 demonstrates interfacing the predictive CAE modeling tool to the user's manufacturing CAD /CAM system for ease and rapid geometry input. This interface capability provides engineering manufacturing operations with a rapid predictive tool for process simulation and design modification at an efficiency that is acceptable for use as a manufacturing tool. This computer-based predicative tool provides the following :
   1.   Advanced graphic user interface.
   2.   Relevent output graphic and tabulated computations, see Figures (4,5).

APPLICATION OF THE COMPUTER MODELING TOOL FOR ELECTROFORMING MANUFACTURING.

To demonstrate utilization of this software as predictive manufacturing tool, the following example application of nickel electroforming of cylindrical shape product is shown (Refer to the schematic in Figure 6). The typical conventional manufacturing approach requires extensive labor for selective masking and part-specific fixture development. Then, the operator's skill is required to predict an addequate anode arrangement (around the parts) and the best setting to achieve the desired current density.

 The ATP funded CAE computer modeling tool can rapidly and accurately predict the best manufacturing conditions to produce acceptable thickness distributions on the plated part. Figures 7 and 8, presents the computed output for the potential map and the expected deposit thickness distribution for the plated tank system having two anodes around the part (cathode). Figure 8, further shows the low predicted thickness at the top side of the part. On the other hand, if a third anode is placed on the top side as shown in Figure 9, the coating thickness on the top side of the part can be inproved (Refer to Figure 10).

Further, in order to minimize or eliminate the high deposit build up on the edges (commonly known as edge effect), the modeling tool can easily incorporate candidate shielding arrangements and compute their impact within the plating cell. See Figure 11 for the cross section model of the cell and Figure 12 for the graphic output for the reduced edge effect.

The expected thickness profile achieved over 60 minutes by applying different current densities can also be readily simulated as shown in Figure 13. This future provides an excellent tool for manufacturing personnel to estimate the time requirements and more accurately schedule product delivery.

SUMMARY

Industrial Technological Benefits

This unique computerized tool provides engineering precision and predictability to electrochemical process design and can be integrated into the users existing or contemplated CAD/CAM manufacturing process design.

The newly developed computer-aided software can be used as a rapid manufacturing tool for design, modeling and process simulation of electroplated fabricated products spanning a broad range of industrial applications including:

1. Electronic industries
   - Predictive selective plating patterns for Semi-conductor and
   - Circuit board manufacturing

2. Automotive and Aerospace industries
   - Electroform net shape and fine precision parts
   - Corrosion protection coatings.

3. Electrochemical machining
   - Design of cathode tools for Electrochemical machining.
   - Simulate feed rate; shape of cut

4. Metal finishing industries
   - Position of the anode in the plating cell and shielding requirements.
   - Predict the influence of chemistry additives in the bath.
   - Selection of optional cell design and process parameters.

5. Electropolishing and anodizing
   - Determination of the material removal rate and anodizing rate.
   - Simulate film build up and distribution.

REFERENCES

1. Allen J.Bard and Larry R. Faulkner , " Electrochemical Methods ." John Wiely & son , Inc.,New york (1980) .

2. Carl Wagner , J .Eectrochem .Soc., 107,445 (1960).

3. S.Y.Parn ,M.McCormick , D. Howe and A.J.Naismith ." The Modeling of the Electroplating baths and its application to jig design ", Trans .I.M.F., 59, 2, 61, (19981).

4. H.P.Helle.,et al. "Numerical Determination of Potentials Distributions and Current Densities in Multi-Electrode systems ." Corrosion. 37, no.9.522,(1981)

5. W.j .Cook and U.Landau, " Numerical Modeling of current and Potential Distribution- Personal Computer Implementation, " Engineering of Industrial Electrolytic Process . " U .Landau ,R.E.White and Varjian, Eds., The Electrochemical Society proceeding Series , Pennington ,New Jersey, October (1986).

6. M.M .Menon and U. Landau, J. Electrochemical Soc. 134 (8) , 2248-2253 (1987).

Figure 1
Standard electropositon cell

Cathode

Anode

V

+

−

Electrolyte

Applied potential (V) = EMF + Over potentials

Figure 2.

Different overpotentials in a standard electrodeposition cell
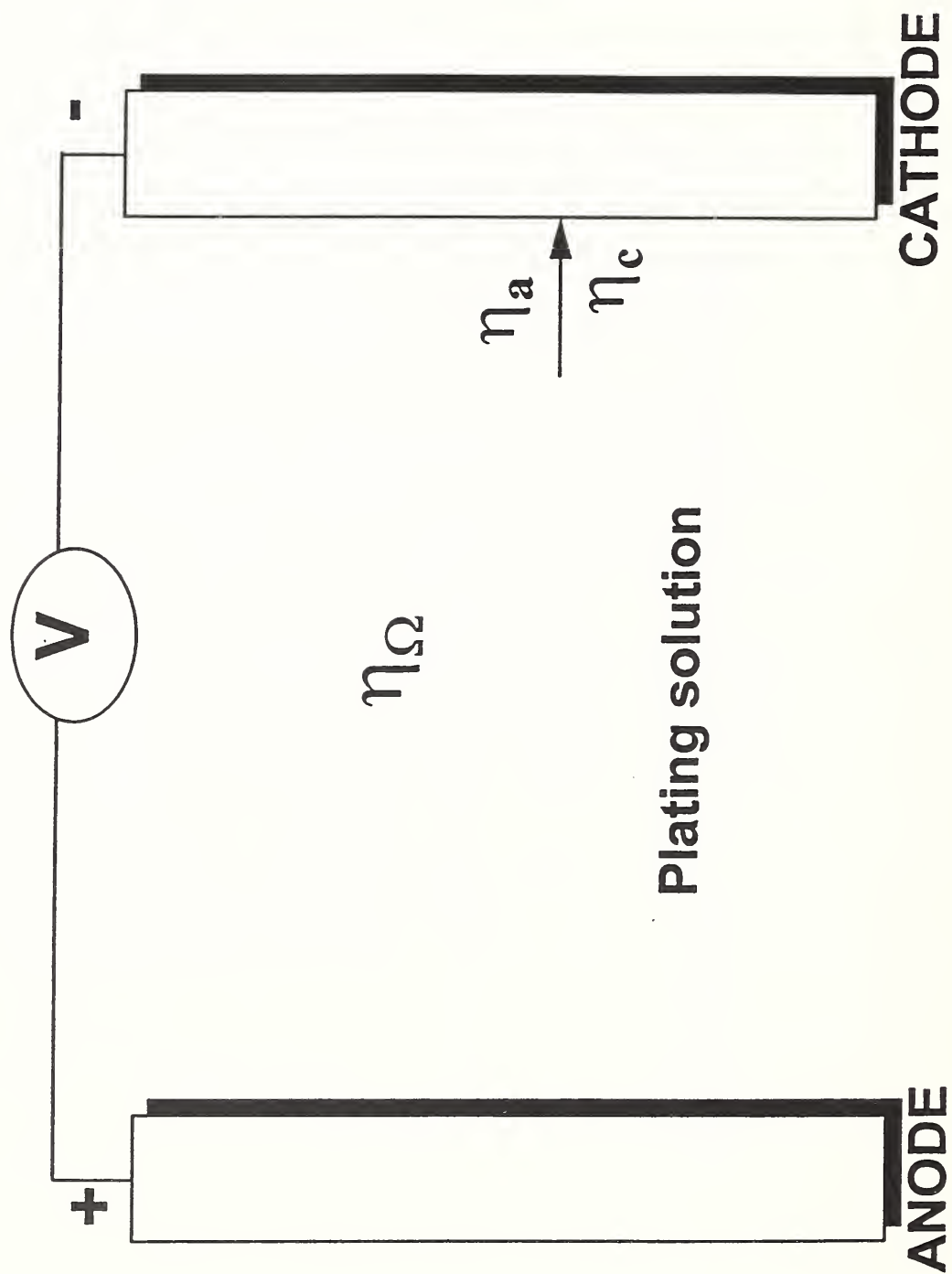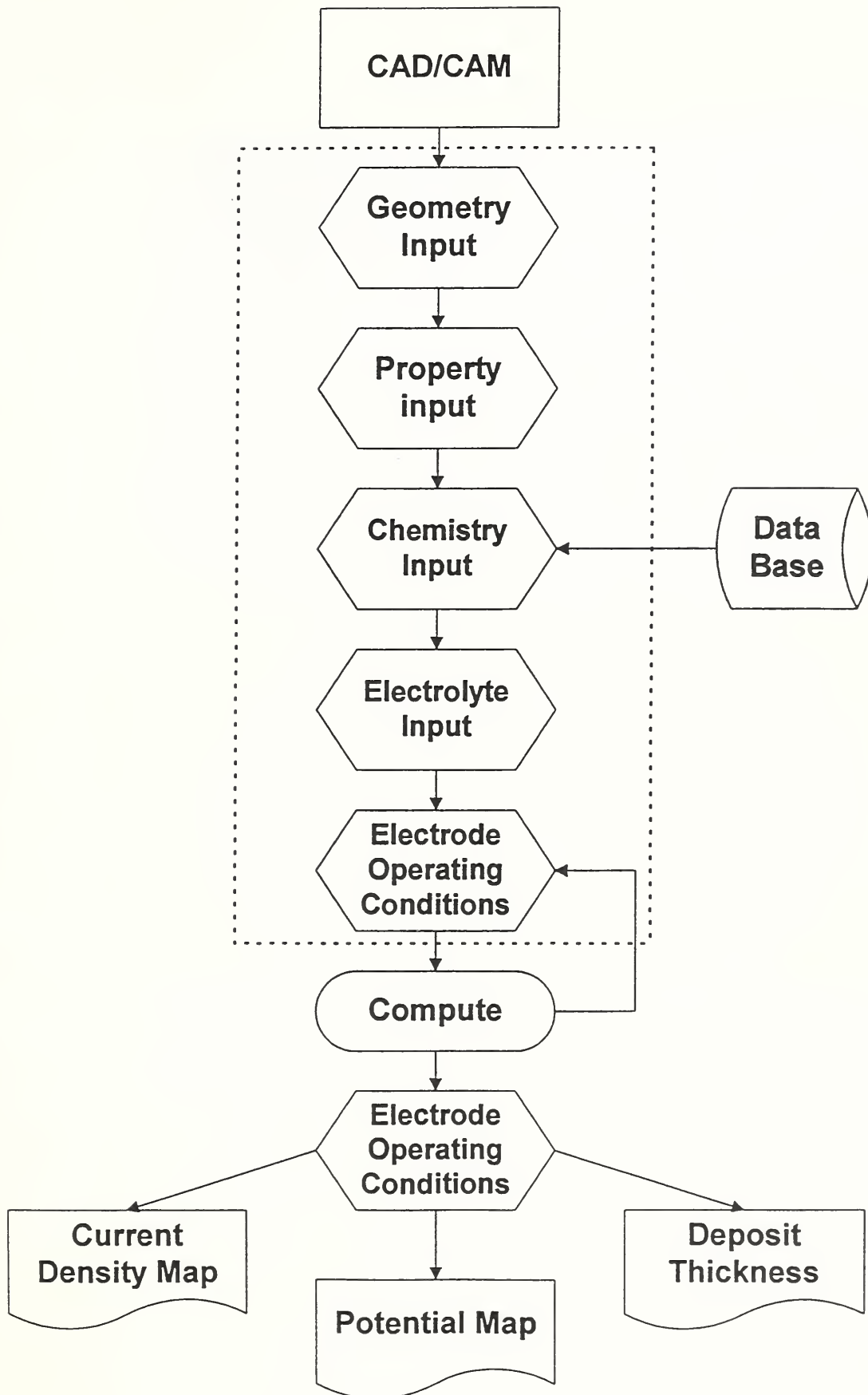
Figure (3): An overview of the computer modeling tool interface to the manufacturing CAD/CAM system.
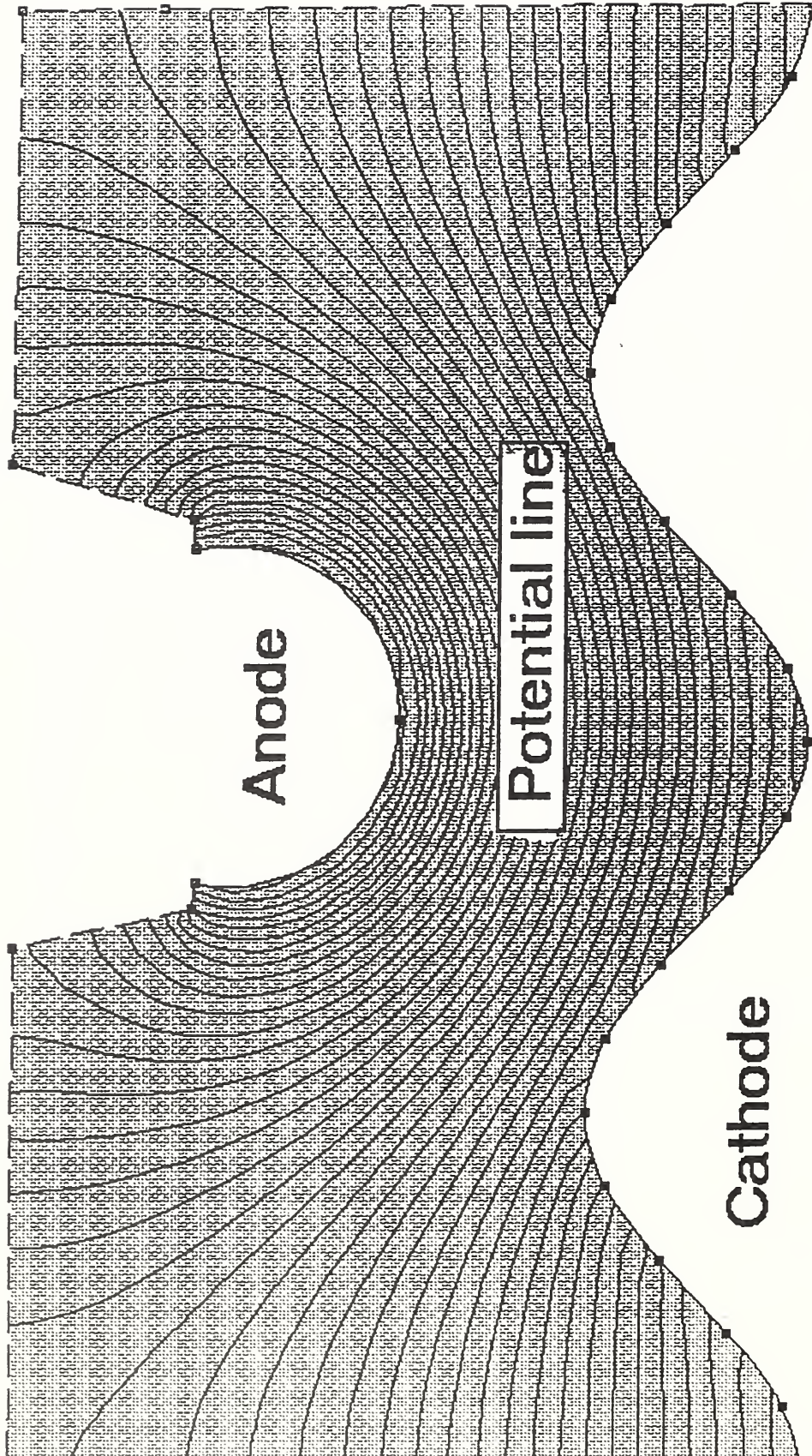
Potential map

Anode

Potential line

Cathode

Figure (4): General view of modeling simulation output for potential map.

Current Distributions map

Anode

Cathode

Current density profile at the cathode
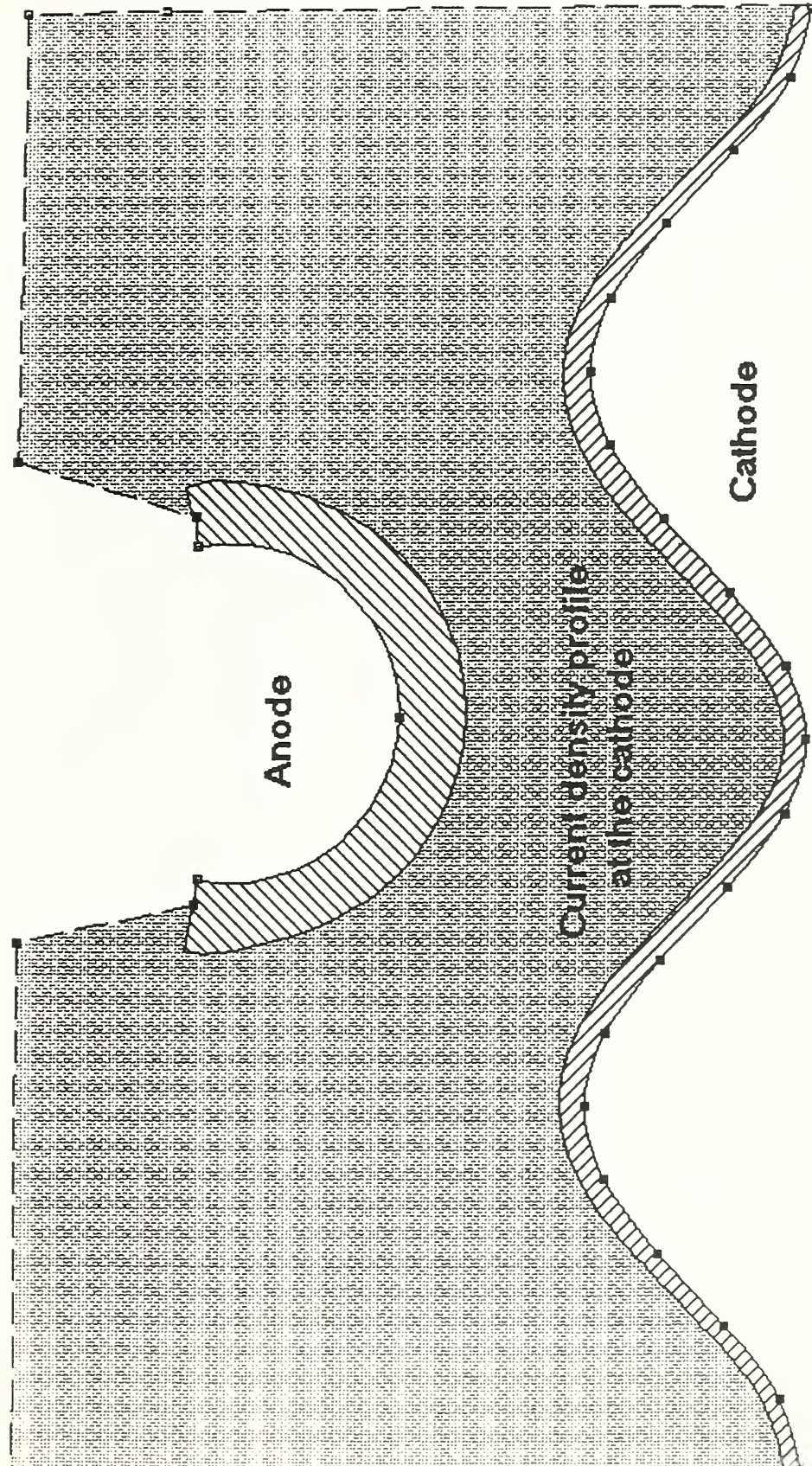
Figure (5) : General view of predicting output for the current density
profile at the cathode surface.
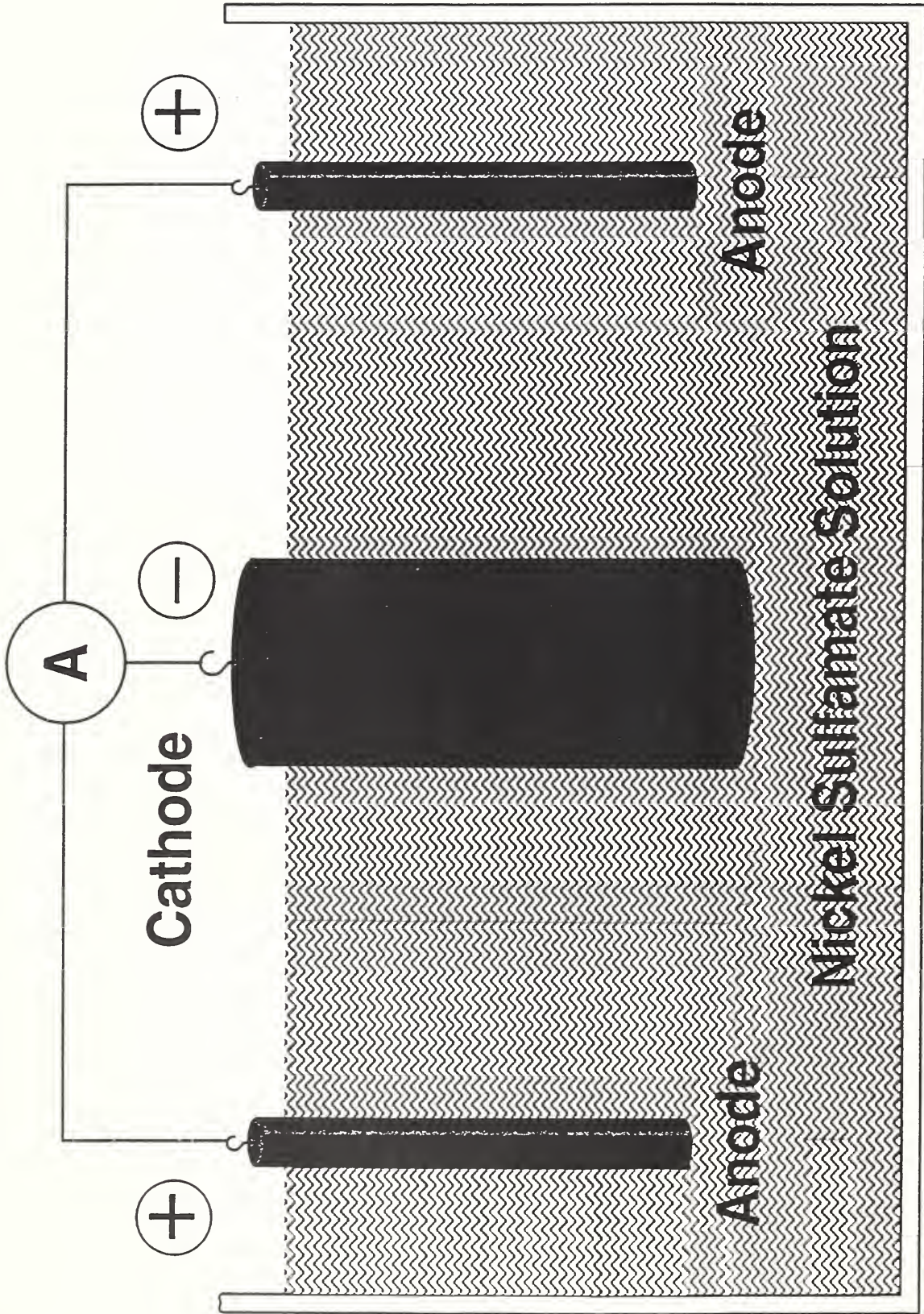
Electroforming of cylindrical shape cathode

Figure (6) : Nickel electroforming of cylindrical shape cathode.

174
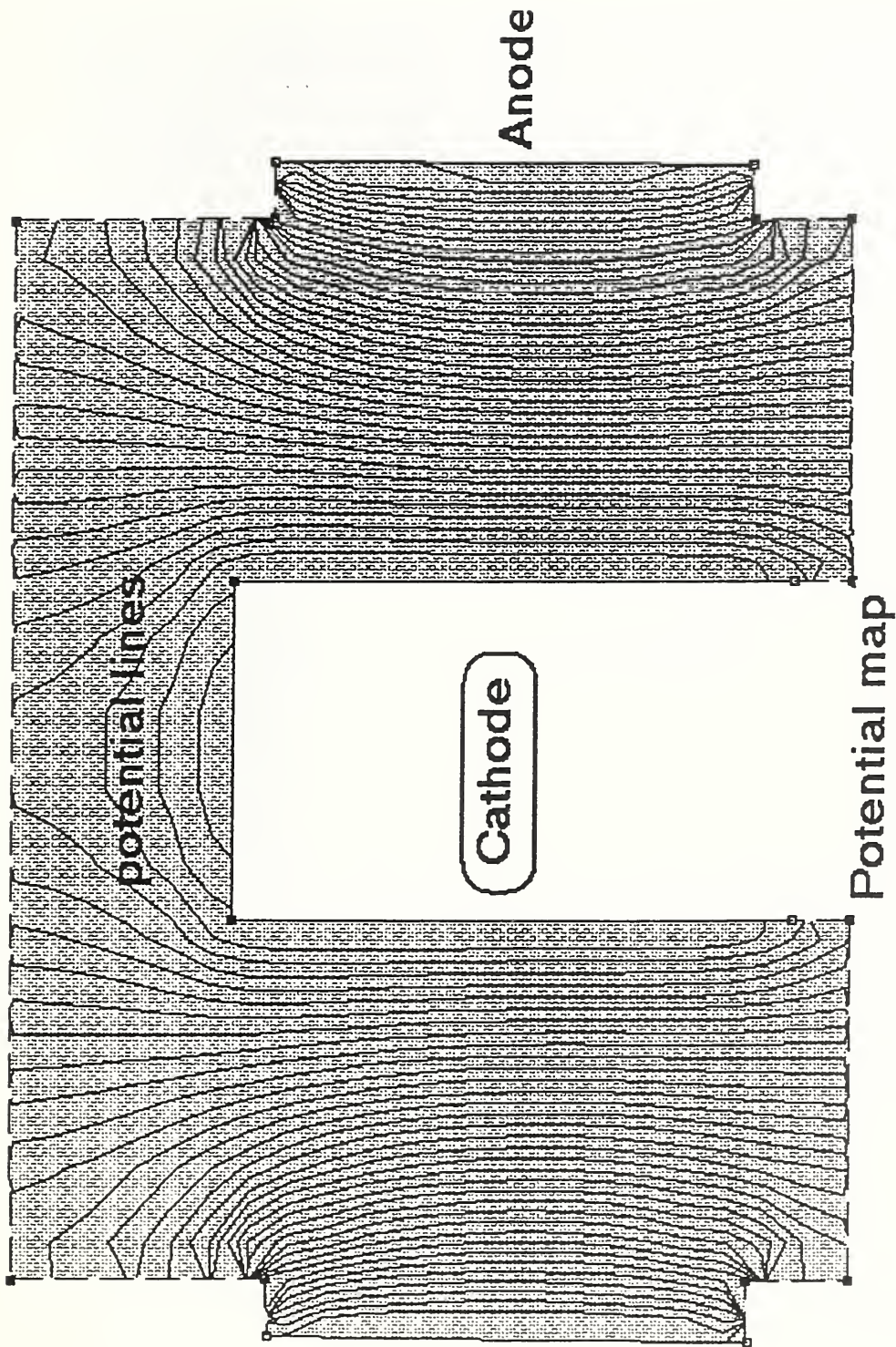
Figure (7) : Cross-section view of the model potential map for the nickel electroforming cell.
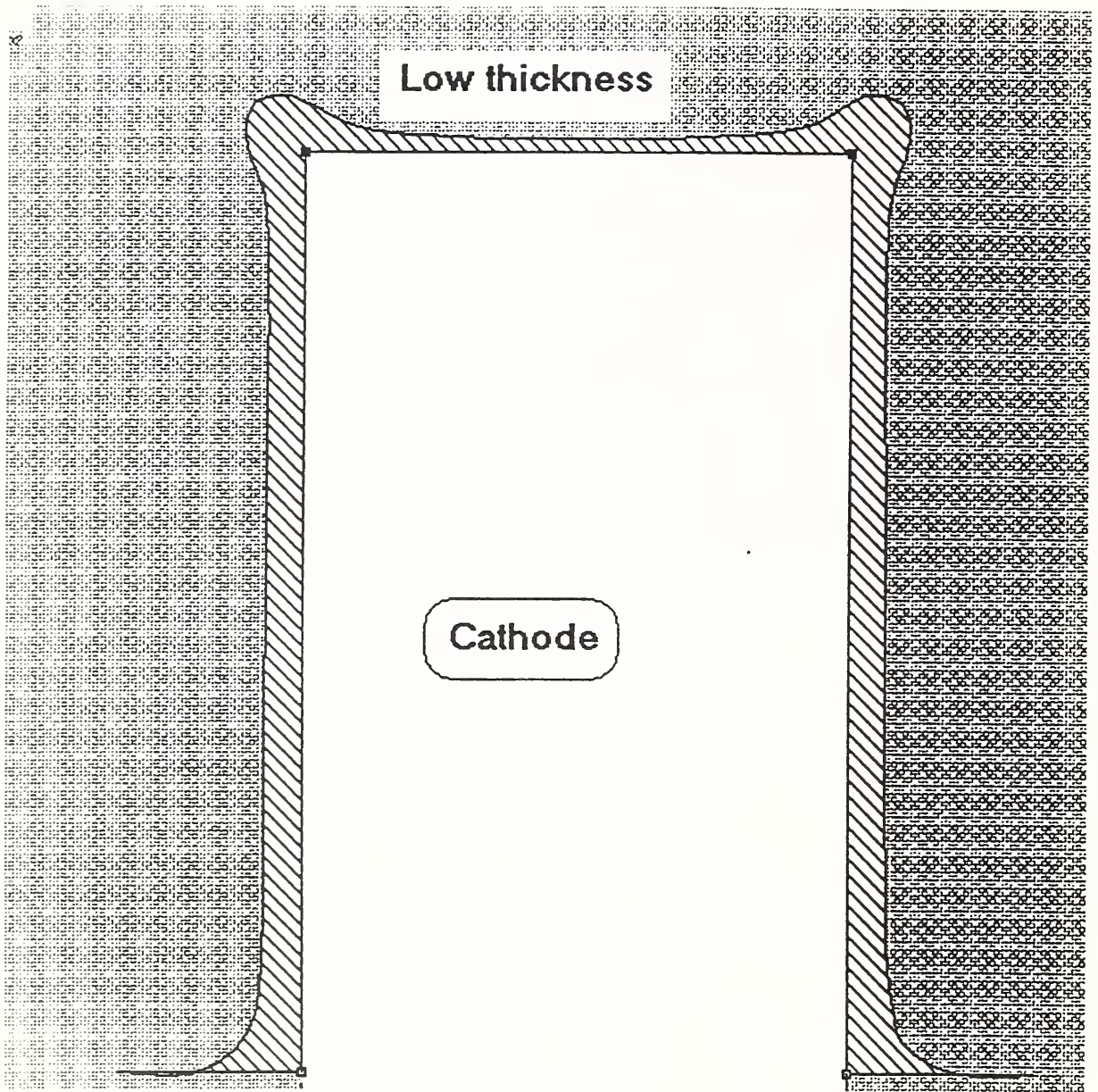
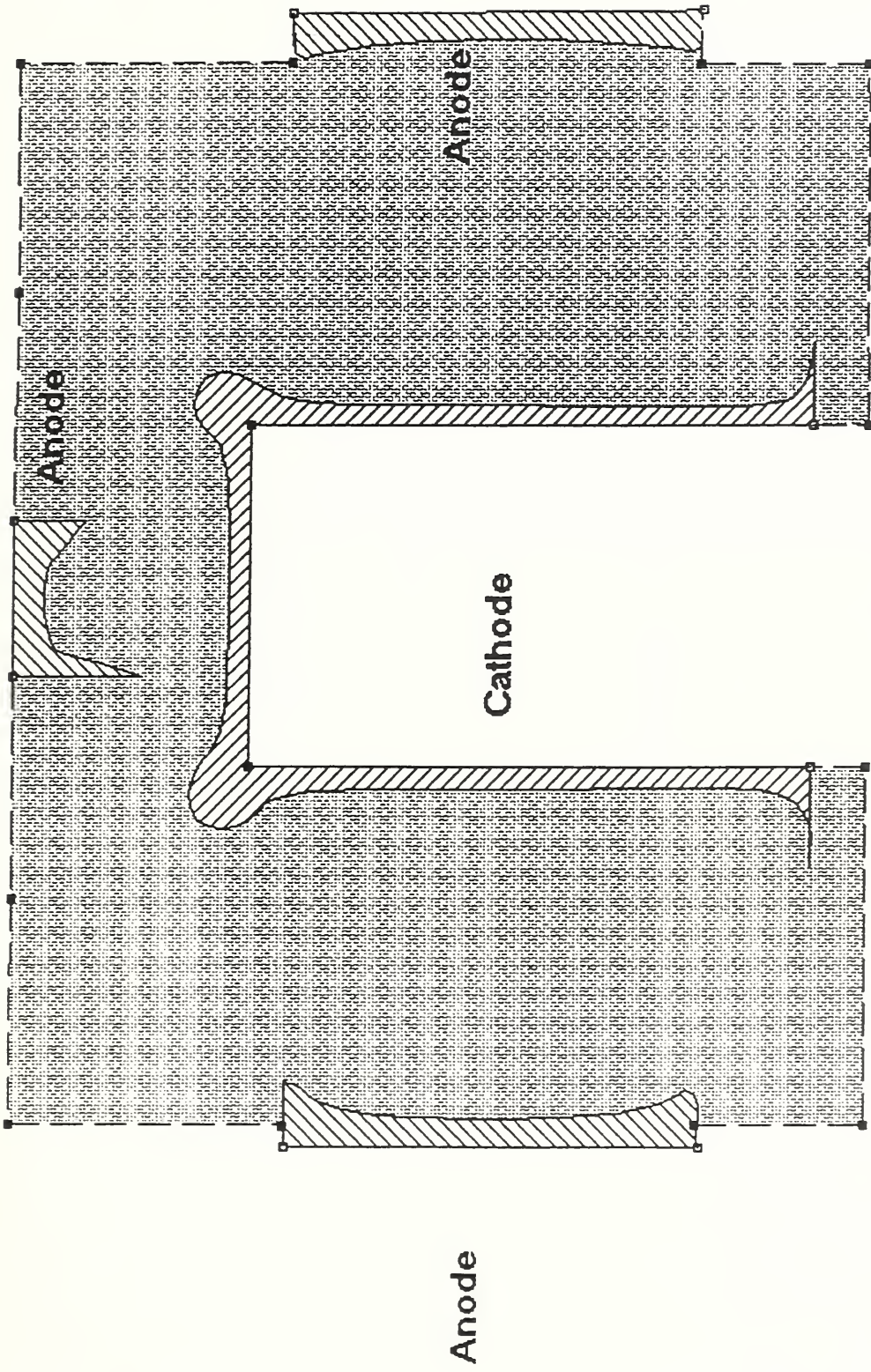**Figure (8): Prediction modeling of the electroforming nickel thickness distribution on the cylindrical cathode.**

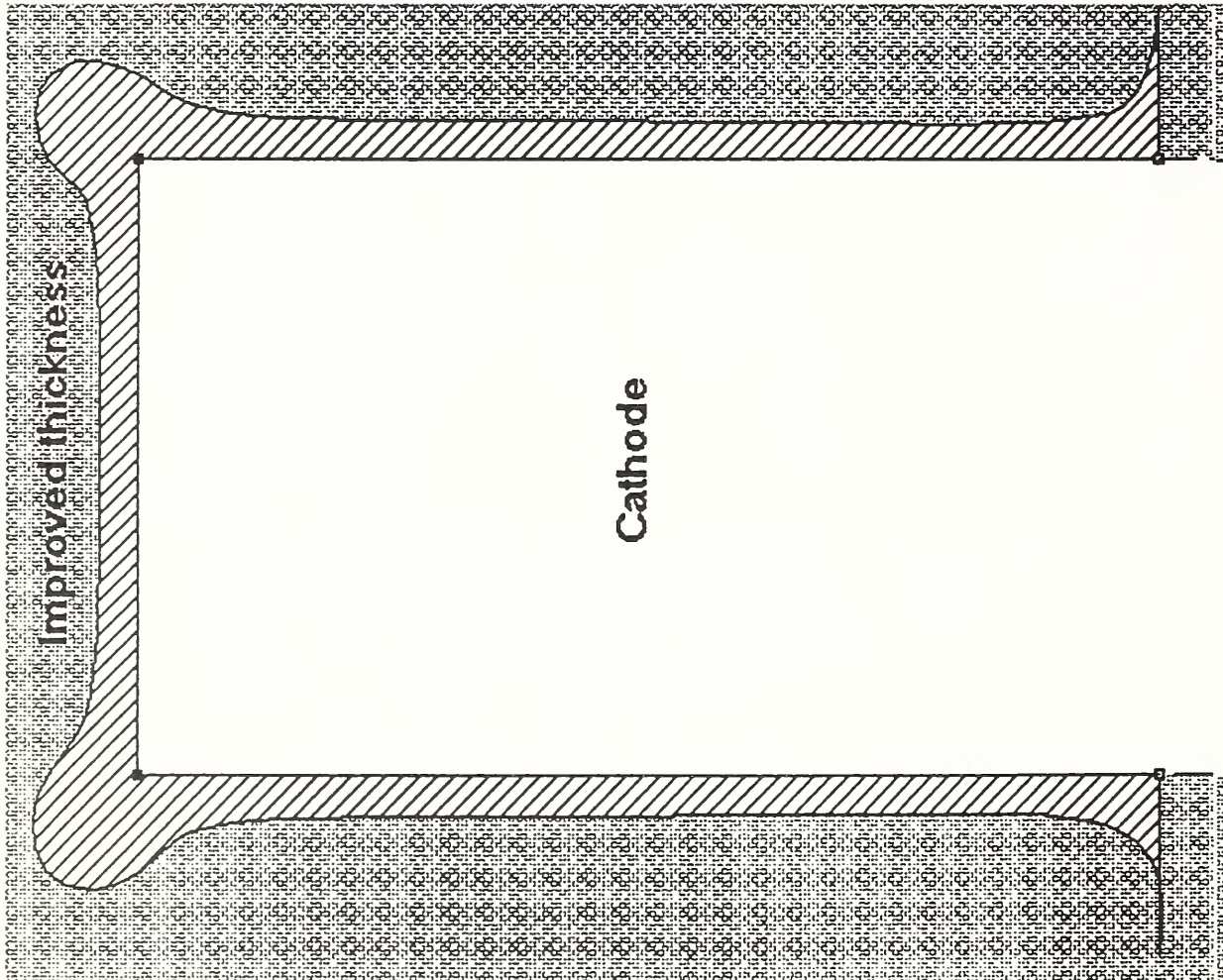**Figure (9) :** Cross-section output simulation of the electroformed cell with three anodes arrangement.

Figure (10) : Cross-section view showing improvement in the nickel thickness distributions by introducing top anode.
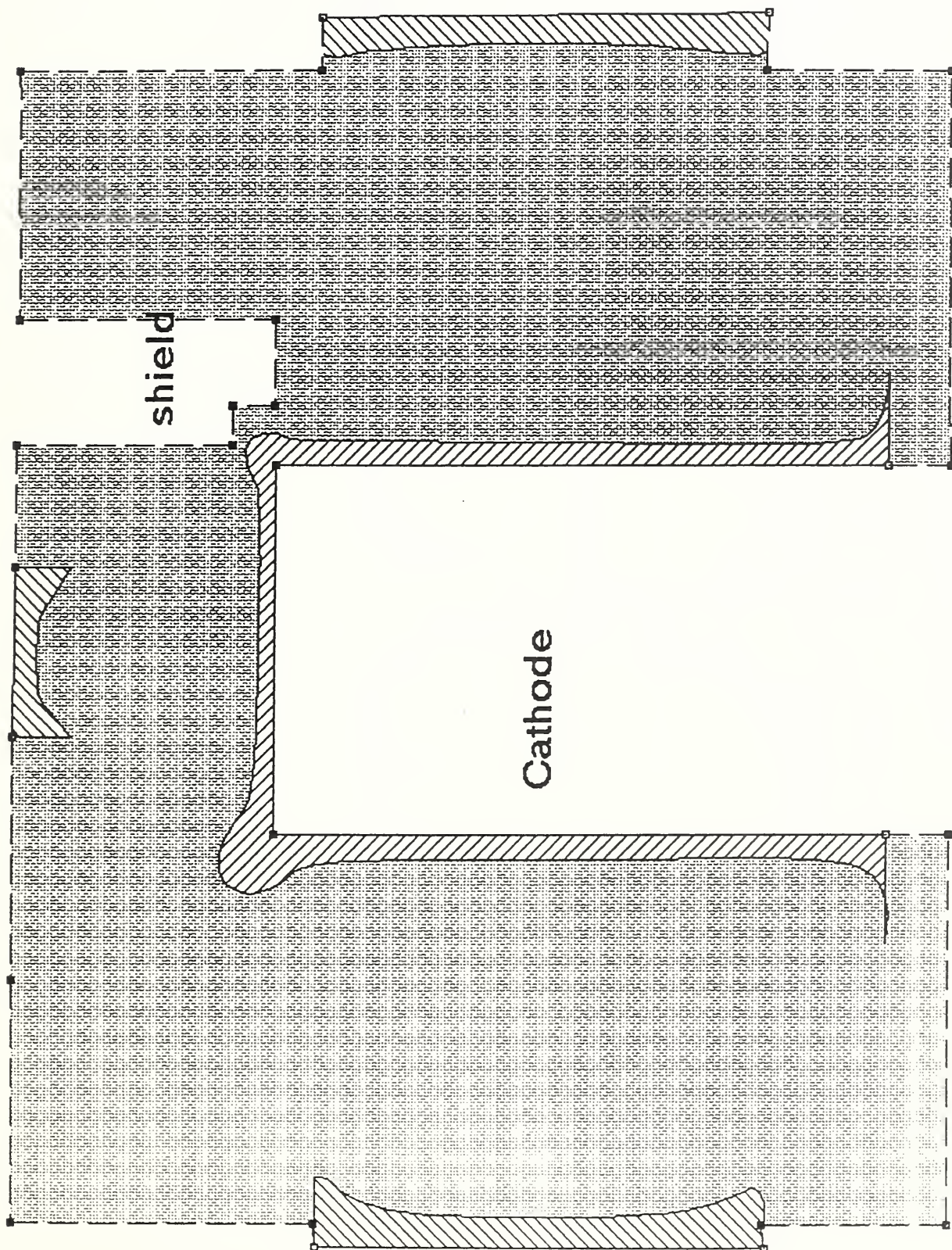
shield

Cathode

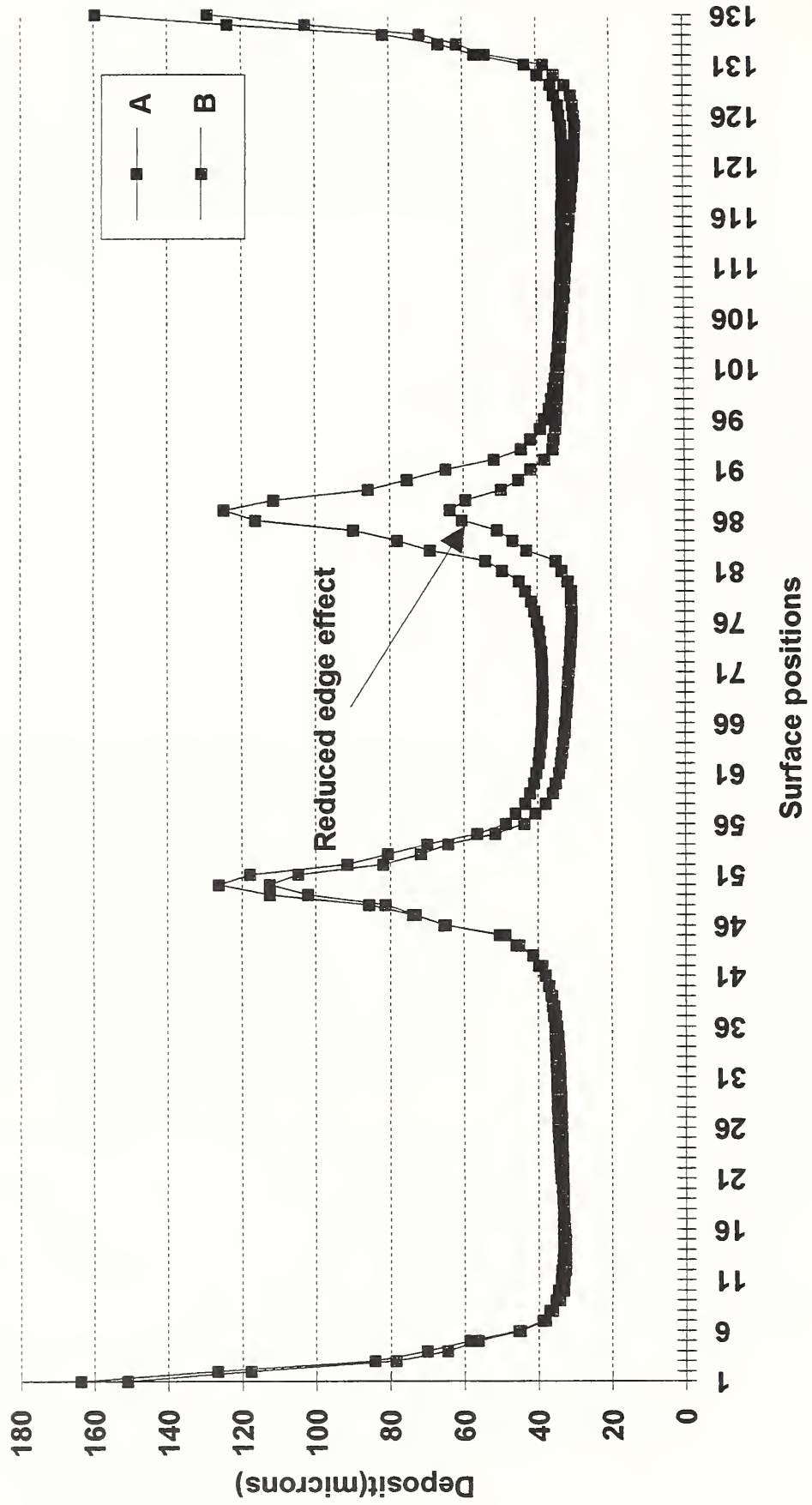Figure (11): Prediction modeling of shield design and location for reducing the thickness build up at the edges.

Figure (12): Graphical output showing reduction for plating edge effect with proper shielding.

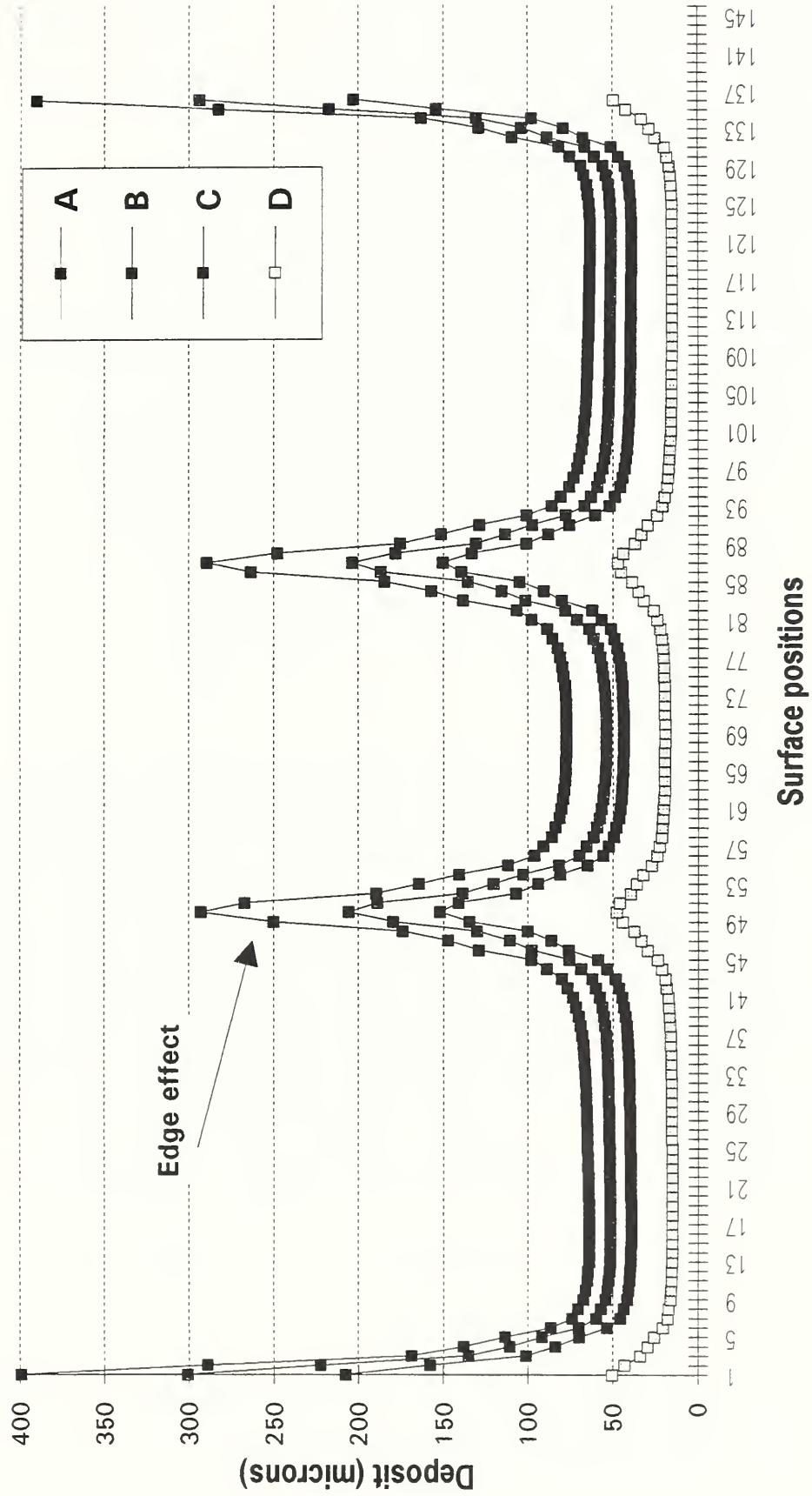# EFFECT OF CURRENT DENSITY

Edge effect

Figure (13): The effect of the current density on the predicted deposit thickness build-up per time.

(Curve A(30 asf), Curve B (40 asf), Curve C (50 asf) and Curve D (15 asf)).

Experiences in the grinding of silicon nitride on a lower cost high speed CNC grinder

by

Tae Wook Hwang[1,2], Eric P. Whitenton[2] and Chris J. Evans[2]

1. Department of Mechanical & Industrial Engineering,
   University of Massachusetts, Amherst, MA 01003

2. National Institute of Standards and Technology,
   Precision Engineering Division,
   Gaithersburg, MD 20899, USA

Abstract:

An investigation for high speed grinding of ceramics has been conducted to understand machine dynamics, wheel wear, and ground surface morphology. Some preliminary results are reported of modal testing and wheel wear together with microscopic observations of the ground surfaces. The modal testing for 4-axis CNC grinder reveals a vertical vibration of the wheel acting as a local "jack-hammer" originating from both the spindle and the motor, especially as the wheel speed increases above 10000 rpm. Wheel wear test using a single-layer plated diamond wheel for grinding of silicon nitride shows that the grinding ratio (G-ratio) is nearly constant of 840 mm^3/mm^3 and surface roughness (R_a) of ground specimen decreased from 1.6 μm at the beginning to 0.4 μm at the end of wheel life. However measured normal force per unit width was found to increase reaching 200 N/mm due to the development of wear flat areas on the active cutting points. Much remains to be done to separate the various affects alluded to in the high speed grinding of ceramics.

Keywords: high speed grinding, ceramics, plated diamond wheel, wheel wear

I. Introduction

Silicon nitride is the material of choice for a broad range of applications – including cam followers, hybrid bearings, pump seals, and blades in high temperature stationary turbo-generator sets. One primary limitation to the introduction of these

materials is the cost of machining parts with adequate form, finish, and strength using "conventional" grinding processes [1]. One approach to obtaining complex contours in conventional grinding is to dress the inverse of the required feature into a grinding wheel. Such a process is typically expensive, and non-uniform wheel wear tends to cause loss of profile accuracy.

Over the last decade, a great deal of work has been done investigating the use of single layer plated cubic boron nitride (CBN) wheels for form grinding of hard steels and superalloys. Such wheels are particularly well suited for high speed operation and, provided wheel wear is reasonable, can be used to economic advantage.

This paper is a preliminary report of our investigations into the use of plated diamond wheels operating at high surface speeds in the grinding of silicon nitride. The experimental work is being performed on a lower cost, commercially available, high speed 4-axis CNC grinder.

II. Motivation

Breaking strength of ceramic components may be limited either by the inherent strength of the material or by flaws induced by the machining process. "Ductile regime" grinding is alleged to produce fracture-free surfaces while "low damage microgrinding" produces surfaces where fine scale fractures do not propagate deep into the work material. In indentation of isotropic homogeneous materials, a clear transition between "ductile" and "brittle" behaviour occurs as a function of penetration depth. Transitions from "ductile" to "brittle" have been observed in grinding of a range of brittle materials as a function of uncut chip thickness, although the system behaviour is obviously more complicated than is suggested by this simplified model. Note that there appears to be a transitional range, where there is a mixture of plastic deformation and fine scale fracture taking place.

In peripheral wheel grinding, uncut chip thickness is a function of wheel diameter, depth of cut, feed rate (i.e. ratio of wheel speed to table surface speed), and grit spacing. For a given wheel speed and diameter (limits imposed in conventional ceramics grinding practice), the user can decrease uncut chip thickness by:

- Decreasing table speed or nominal depth of cut, both of which reduce stock removal rate linearly; or

- Decreasing grit spacing.

In practice, the grit spacing in conventional resin or metal bond grinding wheels may be decreased, to a limited extent, by increasing "concentration" or by decreasing the grit size. The latter strategy also decreases the "chip clearance", or volume available to remove the swarf since the individual grains typically require about half the grain to be "held" in the bond. This, practically, limits the achievable depth of cut.

There have been a number of advances in high speed grinding of super-alloys using plated CBN wheels. Wheel surface speeds now approach 200 m/s in industrial practice, with speeds up to 500 m/s reported in research institutions. For ceramics grinding, it has been suggested that increased surface speed leads to increased ductility through adiabatic heating and hence to increased strength in the product. High spindle speed, however, may lead to increased vibration with the individual grits acting as indenters at the frequencies and amplitudes given by the dynamic performance of the machine.

Plated wheels allow much higher abrasive packing densities than can be achieved in conventional metal or resin bond systems and the single abrasive layers allow for much larger grit exposure. Hence, for a given wheel diameter and grain size, a plated wheel should give smaller uncut chip thickness and hence higher ground part strength than would be obtained with a conventional wheel. Unlike conventional wheel bond systems, a plated wheel cannot be dressed or self-dressing; the single layer of abrasives is bonded directly to the wheel core. Wheel life is given directly by grit wear rates; economic application of such wheels depends, therefore, on grit wear rates.

Machining induced fracture depends, in large measure, on the extreme value of grit penetration, rather than the mean. The height distribution of grits on an as-plated wheel depends on the grain size distribution; its effect, however, may be mitigated in one of two ways;

- Using a "dresser" to mechanically pluck the high grits from the bond; or

- Grinding a sample which will chemically dress the high grits; for diamond wheels, good candidates are chrome and tungsten.

The NIST program in high speed grinding of ceramics is attempting to separate the various affects alluded to above using an Edgetek 4-axis CNC grinder[*] (see Figure 1). Work is also underway to understand the machine characteristics and their impact of grinding performance.

III. Machine dynamics

As indicated above, there is no point selecting process parameters which give small undeformed chip thickness if machine dynamics result in the wheel acting as a local "jack-hammer". From the dynamics perspective, there are two primary sources of relative motion between the work and the wheel: driven headstock vibrations, which are typically a function of spindle speed, and the lowest modes of the machine structure.

---

[*] Specific commercial products are identified solely to provide a complete description of the experimental work and does not imply an endorsement by National Institute of Standards and Technology, nor that they are necessarily the best for the purpose.

III-1. Modal testing

One important consideration in high speed grinding of ceramics is to reduce the undesirable vibrations which can potentially cause strength degradation in the workpiece. A slightly unbalanced wheel at high spindle speed exerts high in-process impact energy on the workpiece, enough to reduce the strength although moderate process parameters are applied during grinding. In this section, the dynamic response of Edgetek grinder will be discussed from a various aspects including the effects of wheel balance, wheel mass, and coolant. Dynamic measurements made in-process will also be discussed.

A 26 kW motor drives the spindle using v-grove belts and pulleys, and rotates about 2.5 times slower than the spindle. For the measurement of vibration, two accelerometers were mounted on the spindle housing placed at end of spindle. One was sensitive primarily in the vertical direction whilst the other was sensitive primarily in the horizontal direction. During the test, the data collected was fed into a PC through a signal analyzer for further analysis as shown in Figure 2.

III-2. Results and discussion

Prior to the vibration testing during the cutting process, a hammer equipped with a load cell was vertically applied on the non-rotating spindle nose and frequency response function was obtained as shown in Figure 3. The result shows the presence of two dominant structural modes with the highest peak at about 400 Hz. Since the spindle is designed to be run from 8000 rpm (133 Hz) to 14000 rpm (233 Hz), one would expect that a spindle speed of around 11700 rpm, corresponding to about 78 Hz of motor frequency, may be troublesome if the motor vibrates enough to stimulate the 78 Hz peak. Later in this paper, the issue of the motor vibration will be discussed further.

Vibration testing analysis was conducted with the spindle rotating over a wide range of wheel speeds for unbalanced and well balanced wheels, light and heavy wheels. Furthermore the effect of coolant on the frequency response of the grinding machine was investigated. The results are presented in Figures 4 - 5. The top row of graphs are of the total integrated displacement as a function of the spindle speed. The bottom row of graphs are of the frequency at the maximum displacement amplitude as a function of spindle speed, showing which frequency is the primary contributor to the overall displacement. The label 'S' in the bottom low of the figures denotes the frequency at the maximum peak displacement where the spindle frequency corresponds to this frequency, label 'M' the maximum peak frequency corresponding to the motor frequency, and the label N the noise floor of the accelerometers and frequencies that are not a harmonic of motor and spindle. The vertical displacements as shown in these figures for the balanced wheel exhibit much larger amplitude compared to the horizontal ones as the spindle speed increases above 10000 rpm. For the results of unbalanced and balanced wheels as shown in Figure 4, the dominant mode of machine vibration changes from the spindle frequency to the motor frequency or both as wheel balance is improved. Therefore it would be expected that the motor causes undesirable

vibrations as the spindle speed increases above 10000 rpm even with a well balanced wheel.

Next the mass of the wheel was increased by approximately 50 %. Comparing the left column of Figure 5 with the right column of Figure 4 shows that, with this mass increase, what seemed to be one vertical displacement peak is actually two peaks which separate and move apart. This gives us the ability to move the peaks around to improve machine performance. If we are operating at 12000 rpm, we want a heavier wheel. If we are operating at 11000 or 14000 rpm, we want a lighter wheel. The right column of Figure 5 shows how the coolant has the effect of somewhat dampening the motor vibrations in the vertical direction. The two peaks become lower and just slightly broader when the coolant is applied.

Up to now, the dominant mode of machine vibration under the various conditions is at the frequencies originated from either the spindle or the motor, or both. Similar vibration modes can be seen in Figure 6 obtained during grinding of slipcast sintered silicon nitride with a depth of cut 25 μm and wheel speed 110 m/s (10300 rpm). From the frequency response function as shown in this figure, both the spindle and the motor frequencies exist together, contributing comparably to the overall displacement. However another set of experiments with a more aggressive process conditions shows that the overall displacement is mainly dependent on the spindle frequency. This indicates that wheel out of roundness and other synchronous sources of variation modulate the grinding forces at the spindle frequency, resulting into high impact forces on the workpiece.

The modal testing presented above suggests that two sources from the spindle and the motor mainly contribute the vibration of the machine during the process. An attempt was made to reduce the motor vibration for the well balanced wheel. For this purpose, weights were clamped to the rotor at the motor's free standing end, where we could get to easily access, to perform a single plane balance. Then one spindle speed of 14000 rpm was selected and weights are adjusted to obtain the minimum vibration at the given spindle speed. Once the balancing was achieved, the spindle speed was increased by every 200 rpm from the lowest 8000 rpm. The measured maximum displacement was plotted versus corresponding spindle speed in Figure 7. As can be seen from the results, the displacement steeply reaches the maximum peak at about 11700 rpm and decreases as the spindle speed increases. Since the motor frequency corresponding to the spindle speed of 11700 rpm is 78 Hz, the main source at this condition is likely to be one of the machine's structural modes found from the hammer test as shown in Figure 3. Similar behavior was also found for the spindle speed of 13000 rpm as shown in Figure 7. Hence, it is clear that there is no one "correct" balance which is optimal at all speeds. This suggests that motor balancing should be adjusted as a function of the spindle speed. To address this, a simple "proof of concept" experiment was performed. An electromagnetic voice coil shaker was mounted in the free end of the motor in the vertical direction as shown in Figure 8. The same frequency as that of the motor was applied to the shaker with a phase shifting to

eliminate the vibration from the motor. It was found that the shaker effectively reduced the vibration originated from the motor.

IV. Wheel wear and effect on ground surface topography

A wheel wear test was conducted using a 180 grit size electroplated diamond wheel. The results show that the grinding ratio (G-ratio) is nearly constant of 840 mm^3/mm^3 after an initial transient region. Surface roughness (R_a) of ground specimen decreased from 1.6 μm at the beginning to 0.4 μm at near the end of wheel life. However measured normal grinding force was found to increase reaching about 200 N/mm.

Surface profiles of ground specimens collected at various stages of grinding times were analyzed using both areal power spectrum density(APSD) and areal autocorrelation function(AACF) together with microscopic observations with scanning electron microscopy (SEM). The APSD reveals concentrated energy distribution across the grinding direction and the relative energy decreases with smoother ground surface. The AACF exhibits strong correlation along the grinding direction and steeply decays across the grinding direction.

IV-1. Experimental works

Straight surface grinding experiments were conducted under plunge conditions (no crossfeed) in the down mode on a Edgetek 4-axis CNC machine equipped with FANUC 16M* controller. One grinding condition was selected for the wear test: depth of cut a = 50.8 μm, workpiece velocity v_w = 63.5 mm/s, and wheel velocity v_s = 85 m/s. During grinding, a copious coolant flow was applied with additional coolant directly on the wheel surface to remove swarf. Measurements were made of the grinding power using a power meter, the grinding force components using a piezoelectric dynamometer (Kistler 9257A*), and spindle vibrations using accelerometers. The data acquisition system is illustrated in Figure 2.

The workpiece material was a slipcast sintered silicon nitride (AlliedSignal, AS800*). The mechanical properties at ambient temperature given by the manufacturer are fracture toughness K_c = 8.0 (Mpa m$^{1/2}$), hardness H = 16 Gpa, elastic modulus E = 310 Gpa, and Poisson's ratio v = 0.28. All workpiece specimens were machined to the same initial dimensions 21.6 mm x 12.7 mm x 50.8 mm. Grinding was in the longitudinal (50.8 mm) direction with wheel engagement across the entire width (12.7 mm).

The same 180 grit single layer nickel plated wheel (Abrasive Technology*) for the modal test was used of diameter d_s = 203.2 mm (8 inches) and width b = 25.4 mm (1 inch). The grit dimension measured using an optical microscope was $d_g \approx 90$ μm, which is very close to the expected value of 85 μm estimated as 60 % of the mesh spacing. The cutting point density C for the wheel, obtained by counting grains on the

wheel surface using an optical microscope, was $C \approx 67$ mm^{-2}, which is about three times bigger than what was found for a 100 concentration resin-bonded diamond wheel of the same grit size. This value of $C \approx 67$ mm^{-2} is about half the maximum grit density for areal packing in a rectangular array with a grit spacing equal to d_g .

For the measurement of wear depth on the wheel surface, a ceramic specimen with a smaller width than that of the wheel was first ground followed by grinding of a glass specimen of the same width as grinding wheel. A surface stylus trace was then made both on the glass specimen to measure the wear depth and on the ceramic specimen to measure the surface roughness. For each specimen, three measurements were taken at different places and averaged. Four replicas were taken using acetyl cellulose tape at marked places around the wheel surface after completing each ceramic specimen. The replicated samples were examined using an optical microscope to see whether the topography of abrasive grains changes. Furthermore surface profiles of ground specimens were measured using a white light interferometer (WYKO Corporation, Rollscope*) to characterize the spatial properties related to the APSD and the AACF. 300 passes were taken on each ceramic specimen.

IV-2. Results

IV-2-1. Grinding force and wheel wear

The grinding test was terminated due to wheel failure while grinding the seventh ceramic specimen when the plated abrasive layer totally separated from the steel hub. The total volumetric material removal was 71,775 mm^3 corresponding to a volumetric removal per unit width of 5,652 mm^3/mm and a grinding time of 1752 sec.

During grinding, the normal force per unit width F'_n tended to progressively increase as shown in Figure 9. Initially F'_n was only about 10 N/mm, slowly increased up to about 25 N/mm after 500 seconds, and then increased at a much faster rate reaching 200 N/mm at end of the wheel life. After the wheel failure, pieces of plated layer were collected and observed using both SEM and optical microscopy. It is obvious from Figure 10 that wear flats on the active grains are developed and wear tracks along the grinding direction can be clearly seen in Figure 10(b). Therefore an increase of normal force is mainly due to an increase of wear flat area as grinding proceeded, which is consistent with the observations reported by others[2,3]. The measured grinding power (%) per unit width based on the maximum machine power of 26 kW in Figure 11 follows a similar trend as that of normal force. It should be noted that the transient effect due to the machine deflection causes the actual depth of cut to be less than the controlled infeed input to the machine when replacing a ground specimen with a new one.

Surface profiles were measured after various amounts of grinding and the corresponding R_a values obtained are plotted versus grinding time in Figure 12. The surface finish progressively decreases with grinding time. The R_a value at end of the wheel life was found to be close to 0.4 µm, which is only about one fourth of the initial

R_a value at start of grinding. It should be noted that this improvement in surface finish does not necessarily guarantee less strength degradation of the workpiece because of the much bigger normal forces on the workpiece.

An attempt has also been made to obtain a better surface finish without causing such high grinding forces by "chemically" dressing a plated wheel with molybdenum [4]. Some preliminary results with the chemically dressed wheel (180 grit plated diamond) for otherwise identical grinding conditions as above indicate R_a value of about 0.5 μm with much lower normal force per unit width of about less than 10 N/mm. A cross-plot of normal force per unit width versus surface roughness as shown in Figure 13 shows a better performance for the chemically dressed wheel than the one as received. Much additional work is needed in this area.

Profilometry measurements on the ground glass specimen as shown in Figure 14 were used to obtain the wheel wear. The worn depth of the wheel taken from Figure 14 appears to linearly increase after an initial transient region as shown in Figure 15, which resulting into a similar trend on the plot of wheel wear per unit width versus volume of material removed per unit width in Figure 16. To characterize wheel-wear resistance, the grinding ratio (G-ratio), which is the volume of material removed (V_w) per unit volume of wheel wear (V_s), is computed as:

$$G = \frac{V_w}{V_s} \qquad (1)$$

Results for both the instantaneous and accumulated G-ratios as estimated from Equation (1) are plotted versus grinding time in Figure 17. The accumulated G-ratio was calculated by dividing the total volume of workpiece removed by the total volume of wheel wear at a given time. These results suggest an instantaneous G-ratio of about 840 which is almost constant with grinding times except an initial transient region. No measurement was made during an initial grinding time. It is necessary to measure the wheel wear during the initial stage to find out the complete behavior of G-ratio over the wheel life. The accumulated G-ratio increases but is much smaller than the instantaneous steady state value due to the expected initial high run-in wear.

IV-2-2. Topography of ground surface

The ground ceramic specimens collected at various stages of wheel wear were observed using SEM. Surface profiles measured using the white light interferometer were analyzed to characterize three - dimensional (3D) topography.

One of the most advantages for 3D surface assessment compared to two - dimensional (2D) analysis is to provide spatial properties which can distinguish isotropy and anisotropy of engineered surfaces. In a 2D profile assessment, the autocorrelation function (ACF) and the power spectrum density (PSD) have been considered as a good candidate to characterize spatial properties of the profiles [5,6]. Similarly for 3D surface, the areal autocorrelation function (AACF) and the areal power spectrum

density (APSD) have been widely used to characterize the spatial property [7-9]. Therefore both AACF and APSD are applied to characterize the ground surfaces in the present paper. The details were found elsewhere [9,10].

SEM photographs shown in Figures 18(a) - 20(a) were taken after the grinding time t = 264 s, 1048 s, and 1520 s, respectively. At initial stage of wear test as shown in Figure 18(a), the ground surface reveals deep characteristic scratches along the grinding direction which are mainly associated with extensive ductile flow. As the grinding continues and wheel wear increases, the ground surface becomes smoother with almost no presence of deep grooves as shown in Figure 19(a). Furthermore at near the end of wheel life, the ground surface in Figure 20(a) reveals somewhat different surface topography which can be observed on a severely worn surface by rubbing/sliding. The corresponding surface profiles in Figures 18(b) - 20(b) show deep grooves along the grinding direction at initial stage of wear test and turn into smoother surface with grinding times.

The APSD for the ground surfaces in Figures 18(a) - 20(a) are presented in Figures 18(c) - 20(c) with a coordinate system where angular frequencies in two orthogonal axes are zero at the center of the figure and maximum frequency is a half of sampled frequency. The APSD energy in Figures 19(b) and 20(c) are normalized with the maximum peak value of Figure 18(a). As expected, the APSD in all figures is concentrated across the grinding direction although relative energy concentration decreases with grinding times. The AACF are presented in Figures 18(d) - 20(d). A strong correlation is observed along the grinding direction and the AACF decay faster in the cross-grinding direction. As the ground surfaces become smoother with grinding times, the level of correlation along the grinding direction decreases and decay faster. The characteristics of the APSD and the AACF for the ground surfaces are similar to those of turned surfaces possessing the anisotropy along the cutting direction. To compare the spatial characteristics of an anisotropic surface to those of an isotropic one, the same ceramic specimen polished into the surface roughness (R_a) of 0.18 μm was also analyzed. The surface profiles together with the APSD and the AACF are presented in Figure 21 without normalization. The APSD energy for the polished surface in Figure 21(b) spreads out with much less directionality. For the AACF, it steeply decays in all direction representing a characteristic of an isotropic surface.

V. Conclusion

A cost-effective grinding of ceramics at high speed requires an understanding of machine dynamics and wheel performance while ensuring part quality. For this purpose, an investigation has been performed for the use of plated diamond wheel installed on a high speed 4-axis CNC grinder in the grinding of silicon nitride.

Modal testing showed the wheel acting as a local "jack-hammer" on the workpiece during grinding, which is mainly associated with the vibrations of the motor and the spindle as well as the lowest modes of machine structure. The vibration of the

wheel during grinding was observed to steeply increase as the wheel speed increases above 10000 rpm. An electromagnetic voice coil shaker mounted in the free end of motor was found to effectively reduce the vibration originated from the motor at high spindle speed.

A wheel wear test using the 180 grit size single-layer plated diamond wheel showed that the grinding ratio (G-ratio) was nearly constant of about 840 mm^3/mm^3 after an initial transient region. Surface roughness (R_a) of ground specimen decreased from 1.6 μm at the beginning to 0.4 μm at near the end of wheel life. However the behavior of measured normal grinding force and energy were found to linearly increase as grinding times reaching the end of wheel life after an initial transient region. The measured normal force at near the end of the wheel failure was about 200 N/mm, which is about as twenty times bigger as the one measured at initial stage of wheel life. The development of wear flats on the active cutting points was found to be directly related to the increase of normal force. Surface topography of ground surfaces at a various stages of grinding times also showed quite different characteristics revealing a severely worn surface at near the end of wheel life, associated with rubbing/sliding, but a relatively smooth surface with less APSD energy and level of AACF.

References

1. Jahanmir S., Ives L. K., Ruff A. W., and Petersen M. B., 1992, "Ceramic Machining: Assessment of Current Practice and Research Needs in the United States," NIST Special Publication 834.

2. Malkin S. and Cook N. H., 1971, "The wear of grinding wheels, Part I-Attrious Wear," ASME J. of Eng. for Ind., pp.1120-1128.

3. Hahn R. C., and Lindsay R., 1967, "On the effects of real area contact and normal stress in grinding," Annals of the CIRP, Vol. 15, pp. 197-204.

4. Paul E., Evans C. J., Mangamelli A., McGlauflin M. L., and Polvani R. S., 1996, "Chemical aspects of tool wear in single point diamond turning," Precision Engineering, Vol. 18, pp. 4-19.

5. Whitehouse D. J., and Archard J. F., 1970, "The properties of random surfaces of significance in their contact," Proc. R. Soc. London, Ser. A, Vol. 316, pp. 97-121.

6. ANSI, Surface Texture: Surface Roughness, Waviness and Lay, American Standard ANSI B.46.1, 1985.

7. Longuet-Higgins M.S., 1957, "The statistical analysis of a random, moving surface," Philos. Trans. R. Soc. London, Ser. A, Vol. 249, pp.321-384.

8. Nayak, P. R., 1971, "Random process model of rough surfaces," J. Lubr. Technol. Trans. ASME, July, pp.398-407.

9. Dong W. P., Sullivan P. J., and Stout K. J., 1994, "Comprehensive study of parameters for characterizing three-dimensional surface topography, IV: Parameters for characterizing spatial and hybrid properties," Wear, Vol. 178, pp.45-60.

10. Bendat J. S., and Piersol A. G., Random data: analysis and measurements procedures, Wiley-Interscience, 1971.
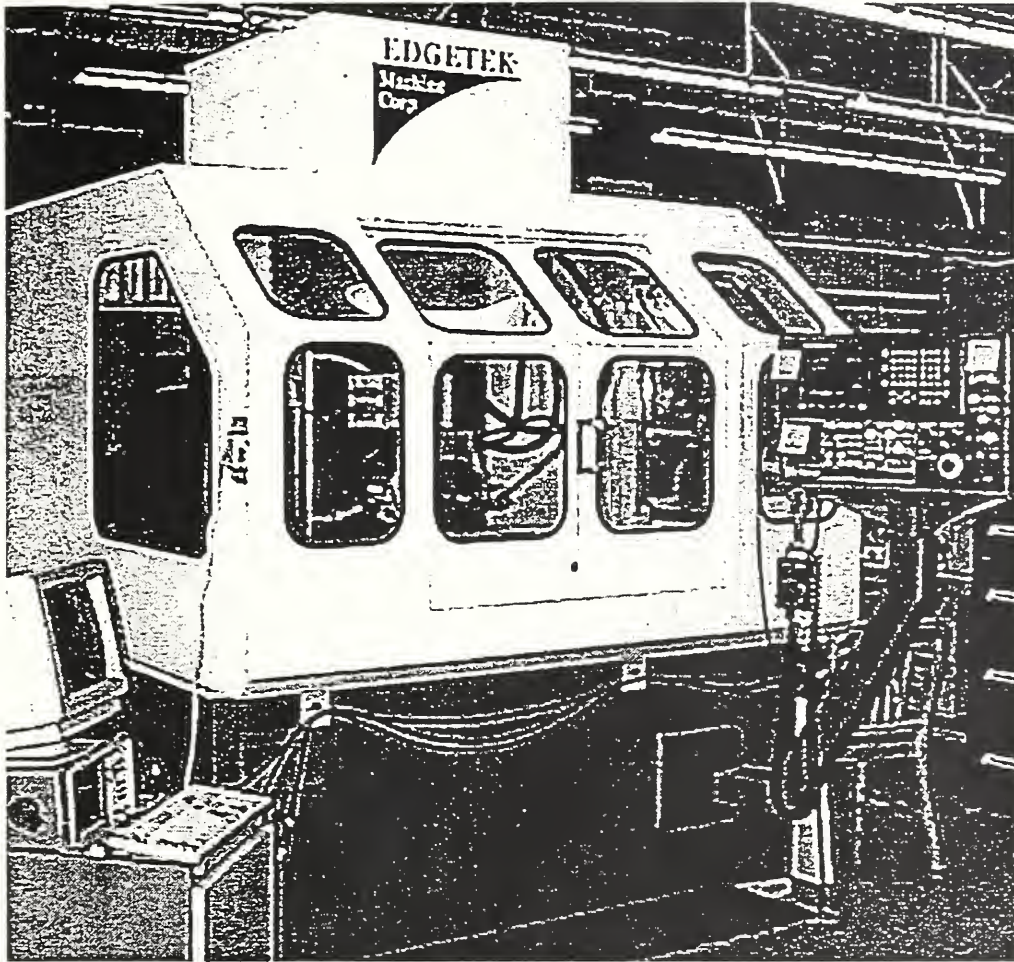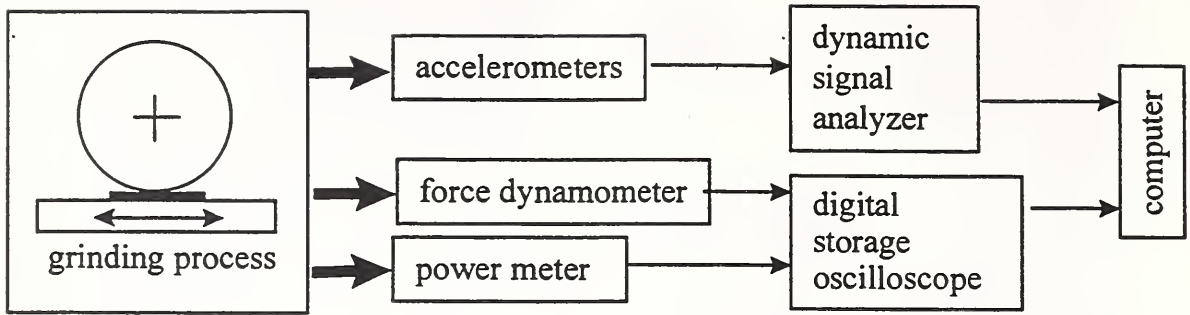
Figure 1   Edgetek 4-axis high speed CNC grinder
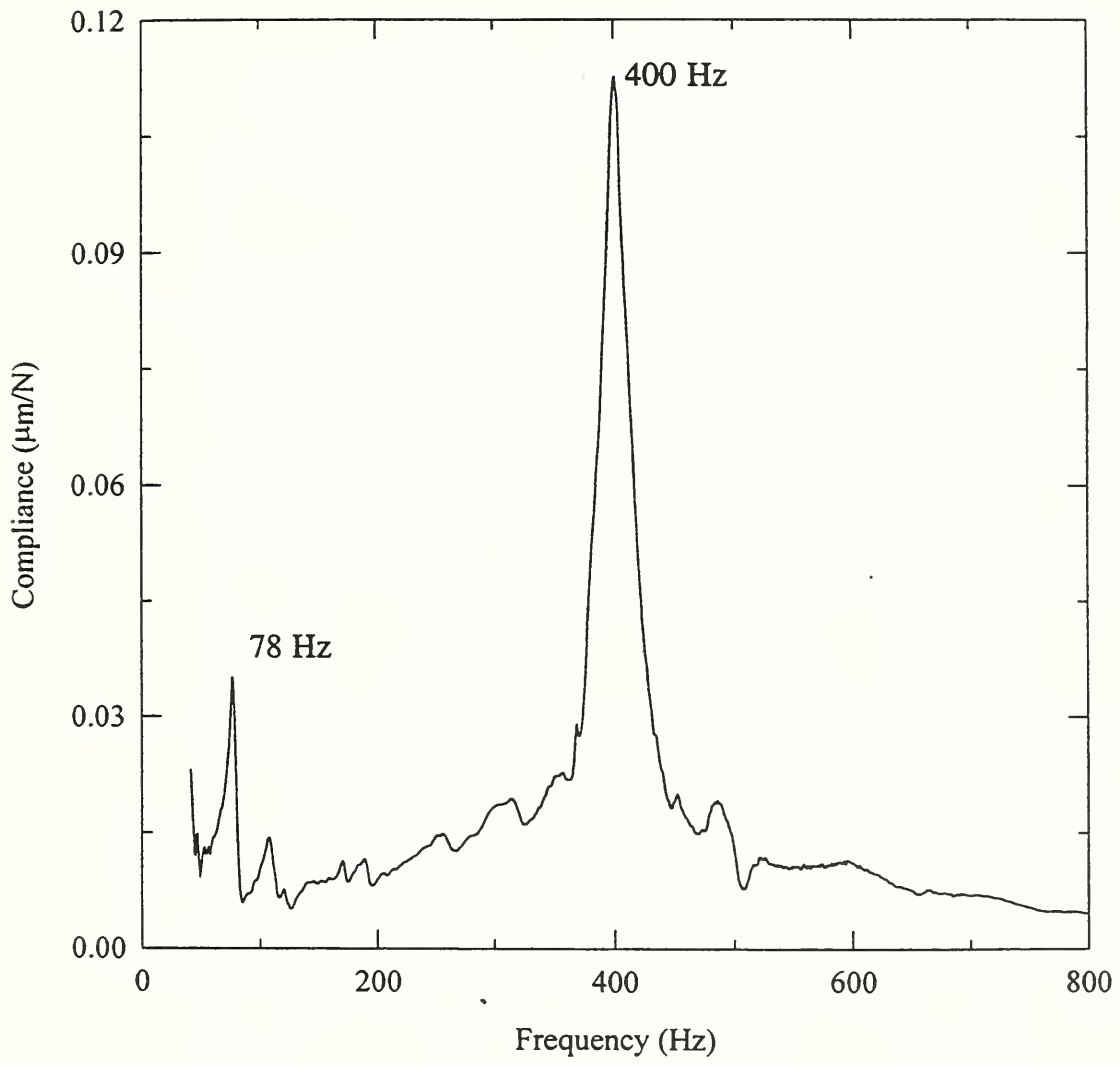
Figure 2   Schematic illustration of experimental setup

Figure 3   Frequency response function from the hammer test
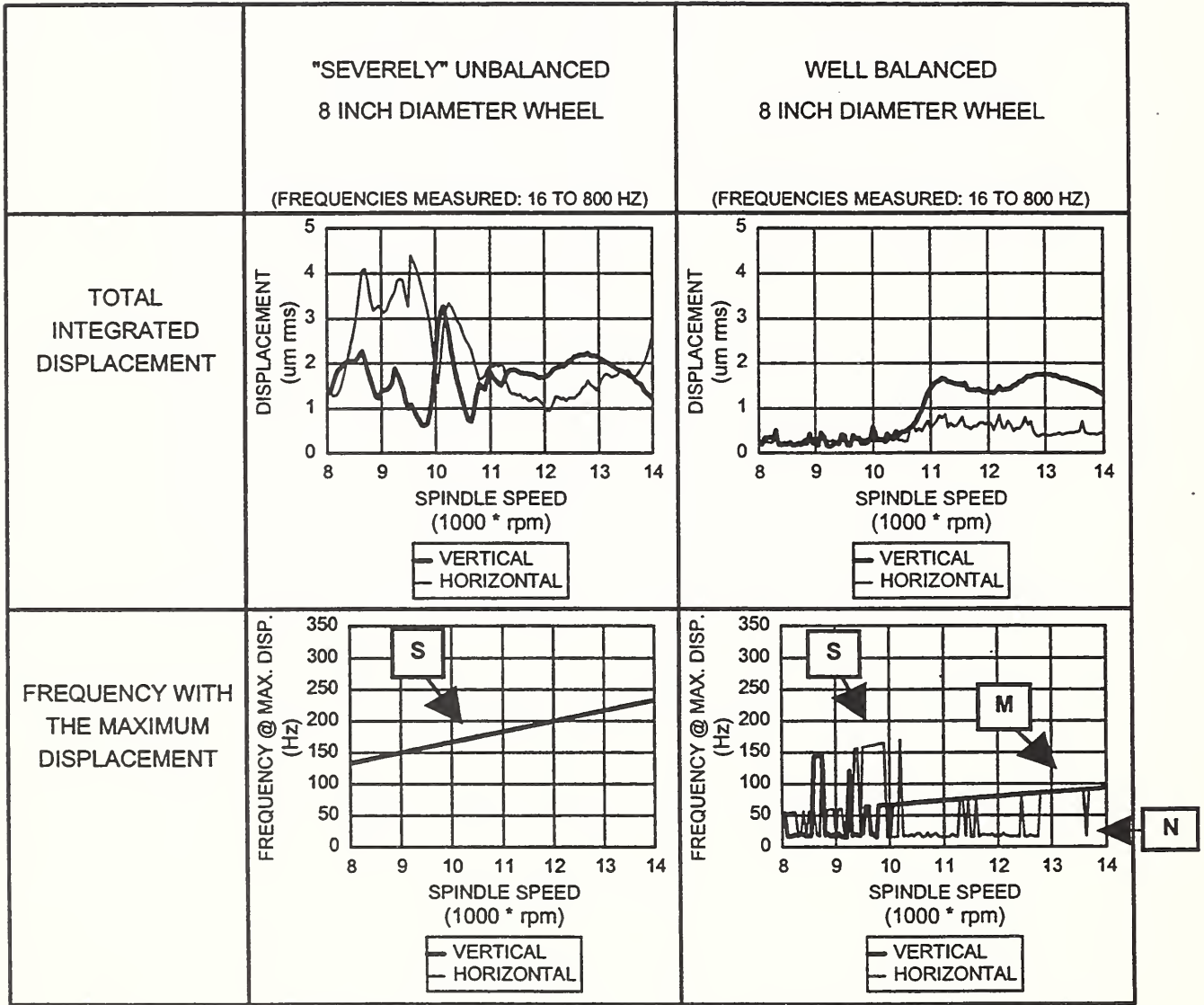
Figure 4   Effect of wheel balancing on spindle vibration
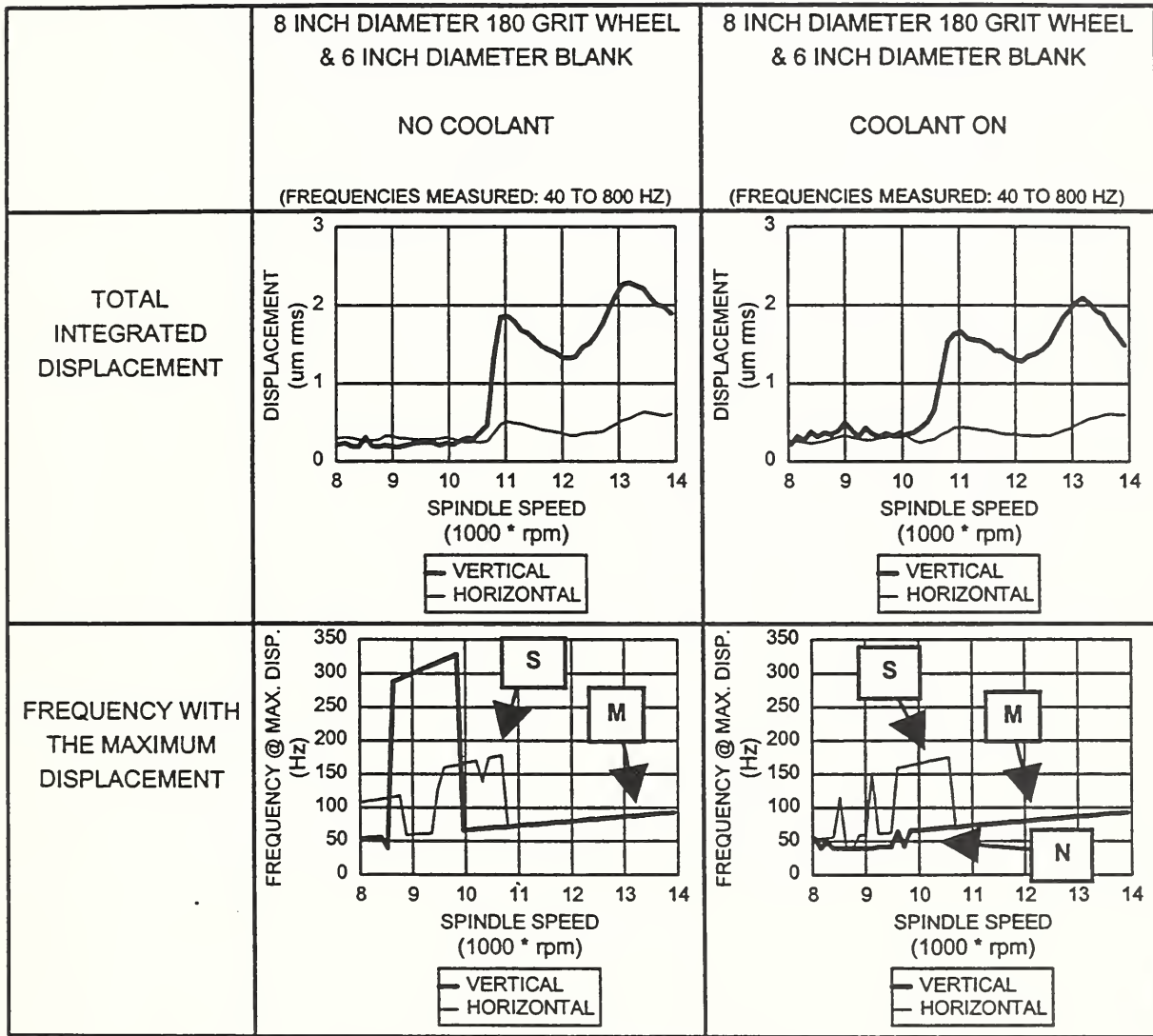
Figure 5   Effects of added mass and coolant on spindle vibration
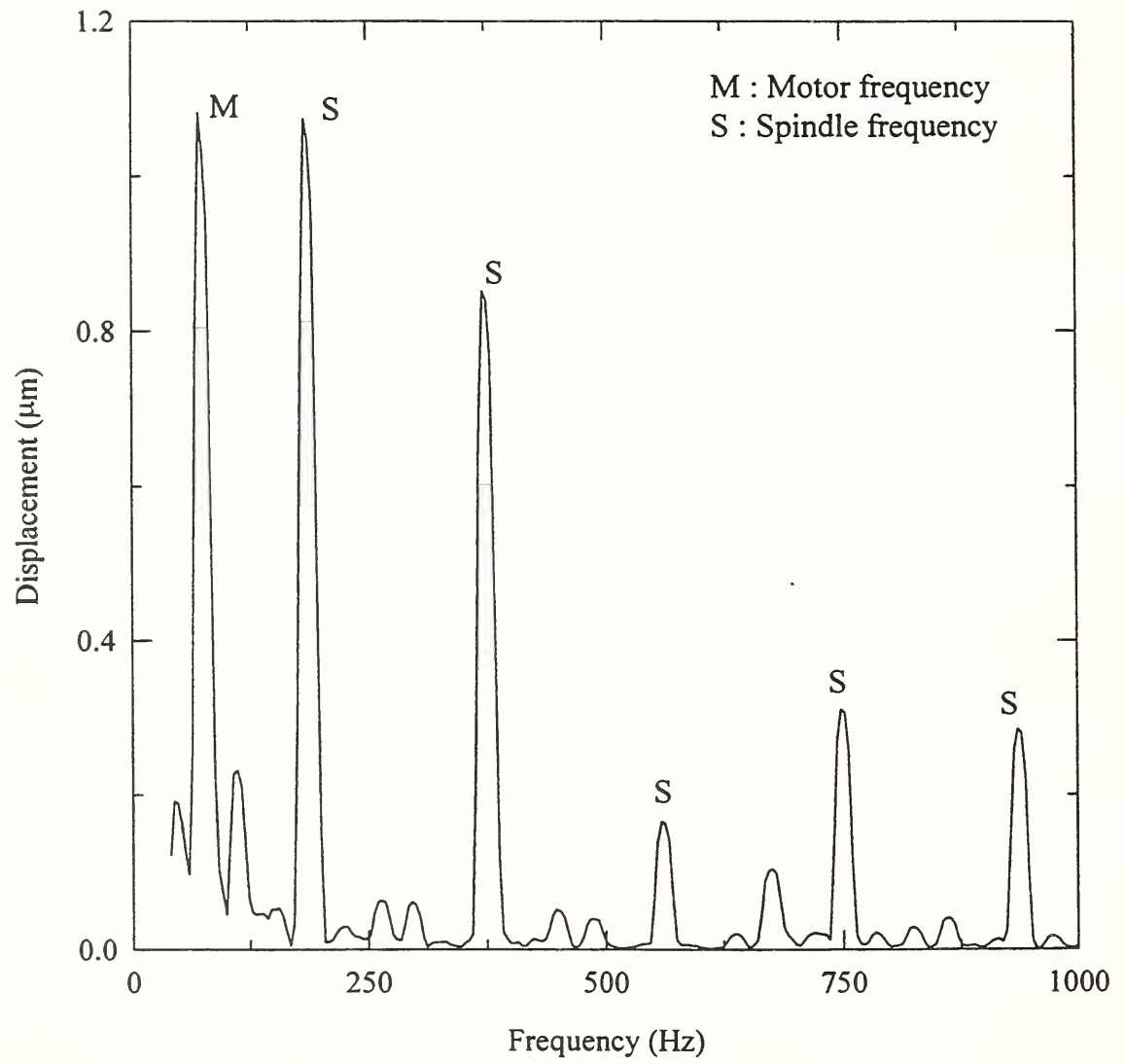
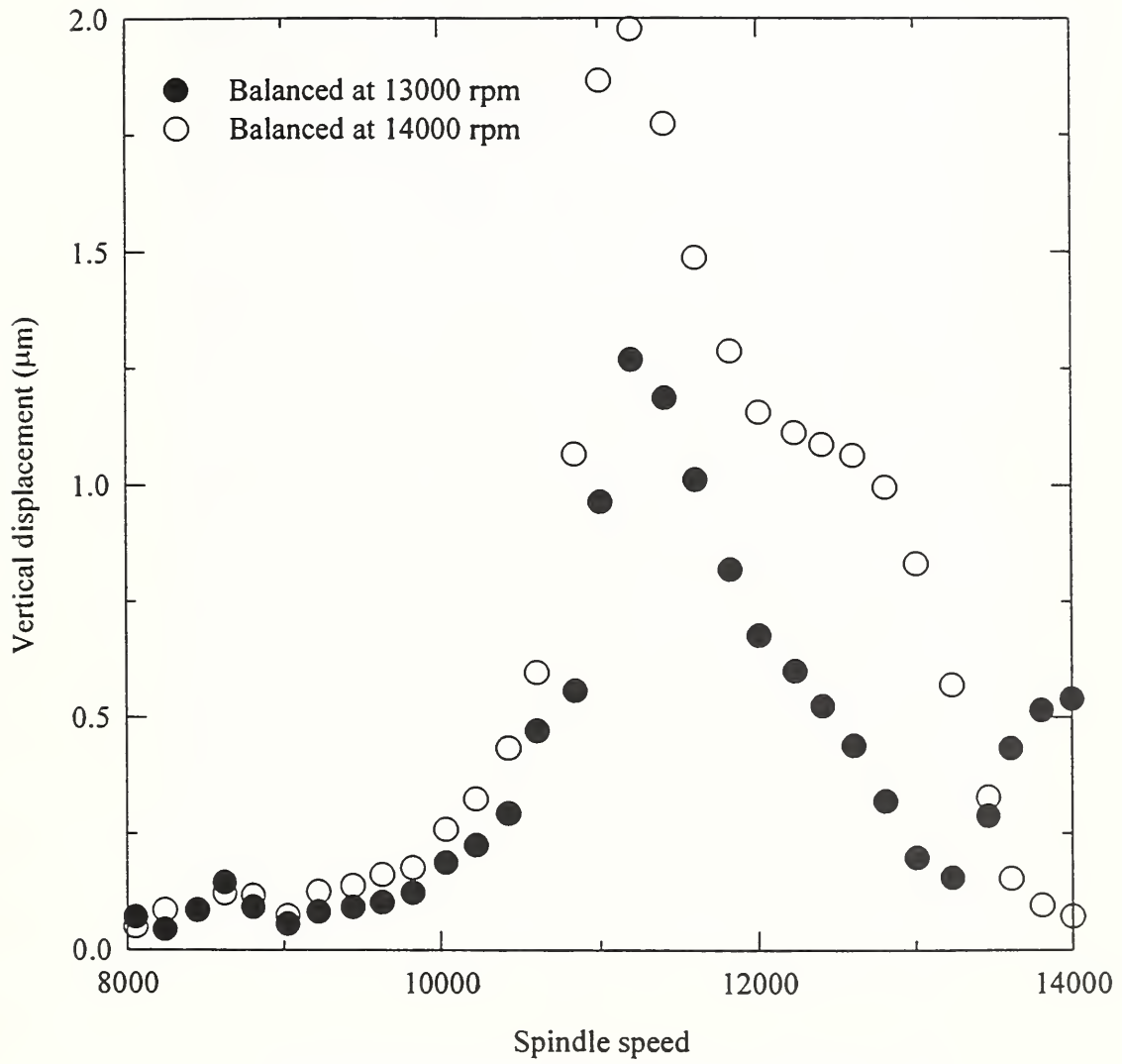Figure 6   Vertical response during grinding

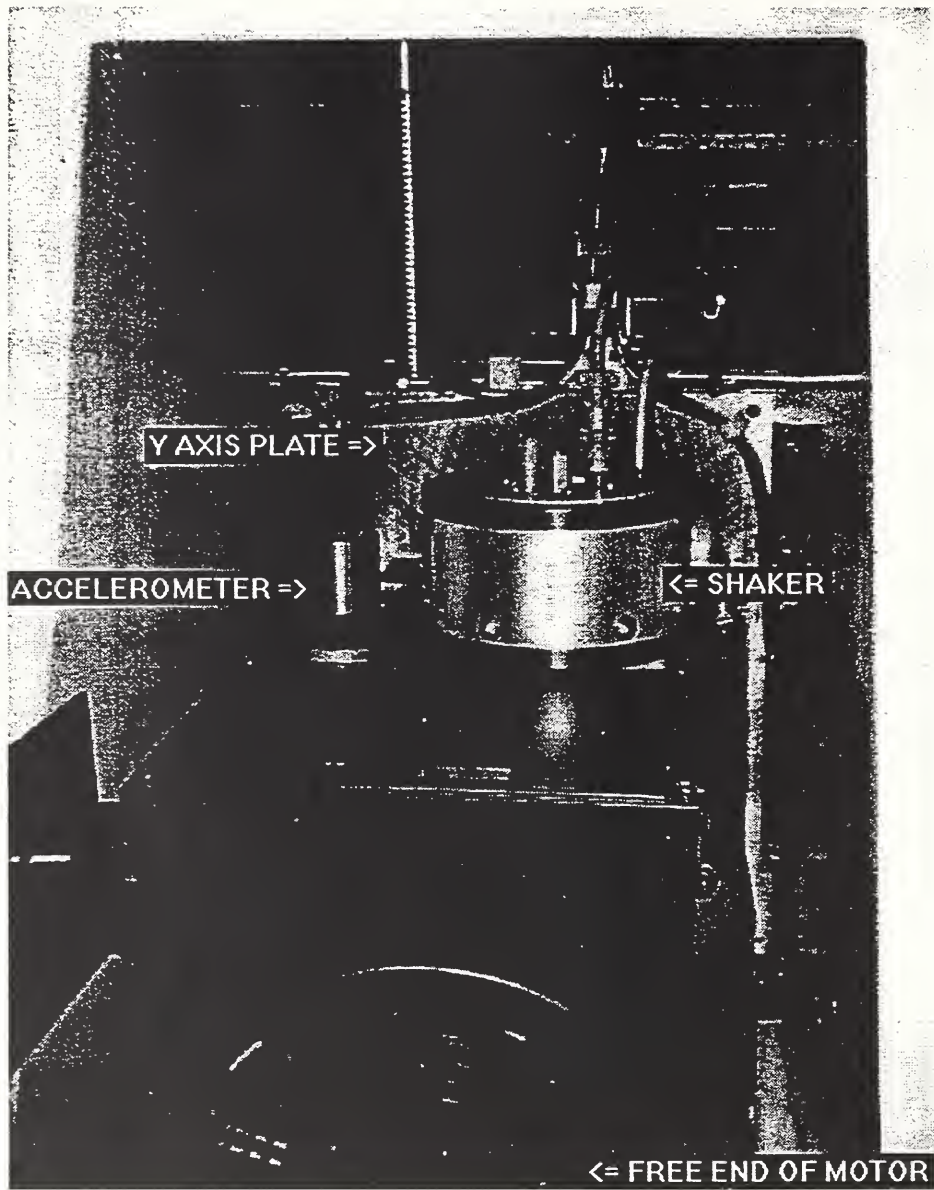Figure 7   Effect of balancing spindle motor at different speeds

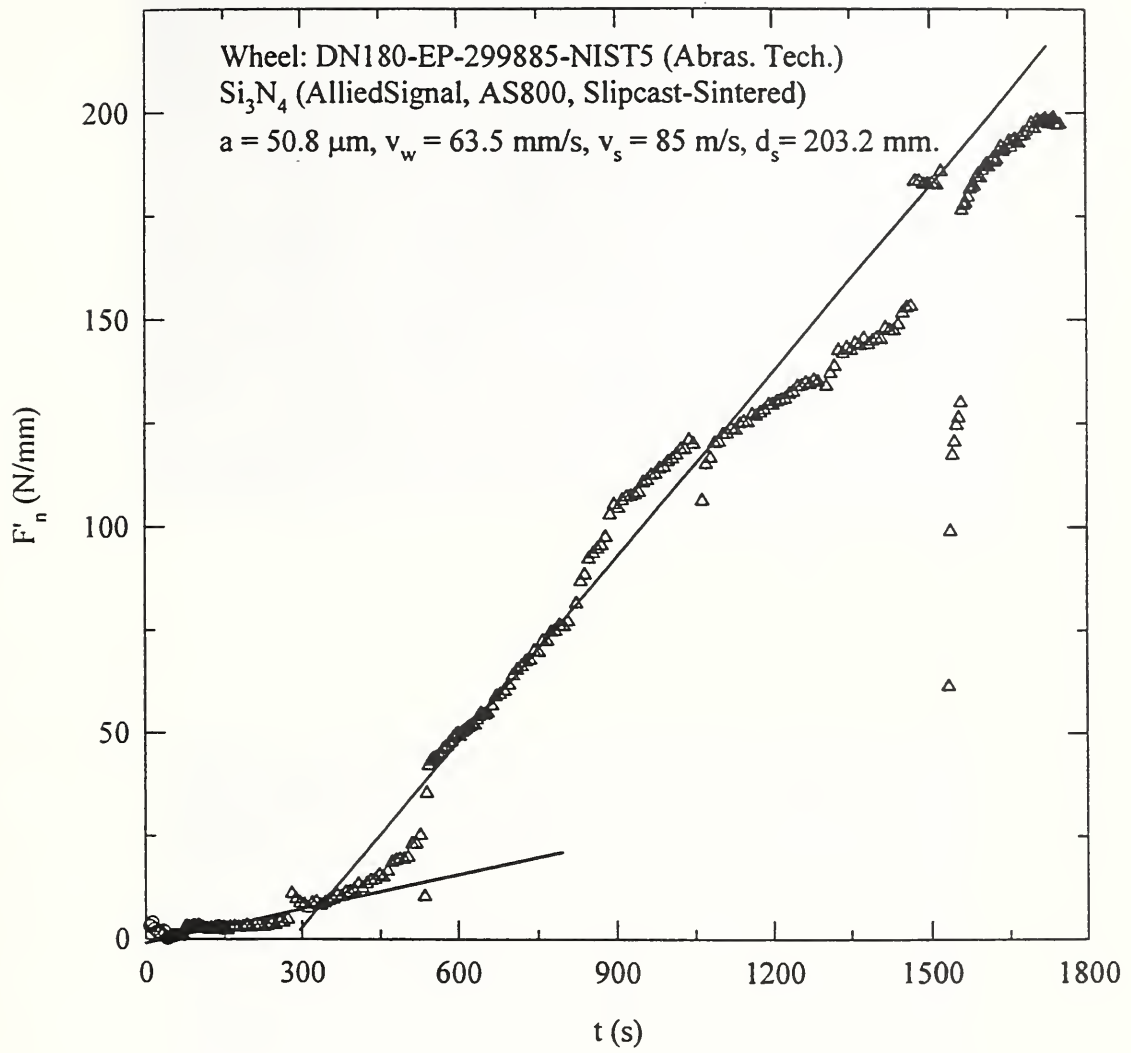Figure 8   Active vibration cancelation set up
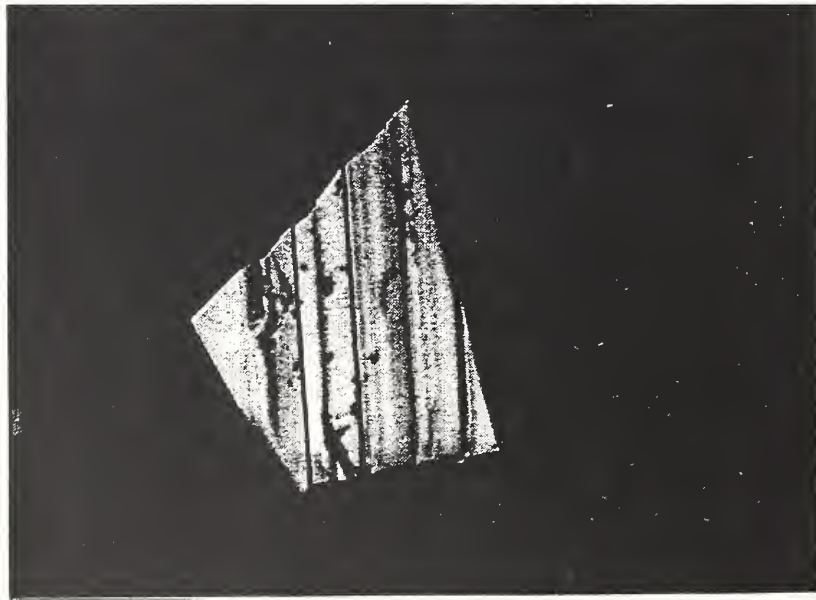
Figure 9   Normal force per unit width versus grinding time.

(a) SEM micrograph

(b) Optical micrograph

Figure 10   Wear flat on the diamond abrasive
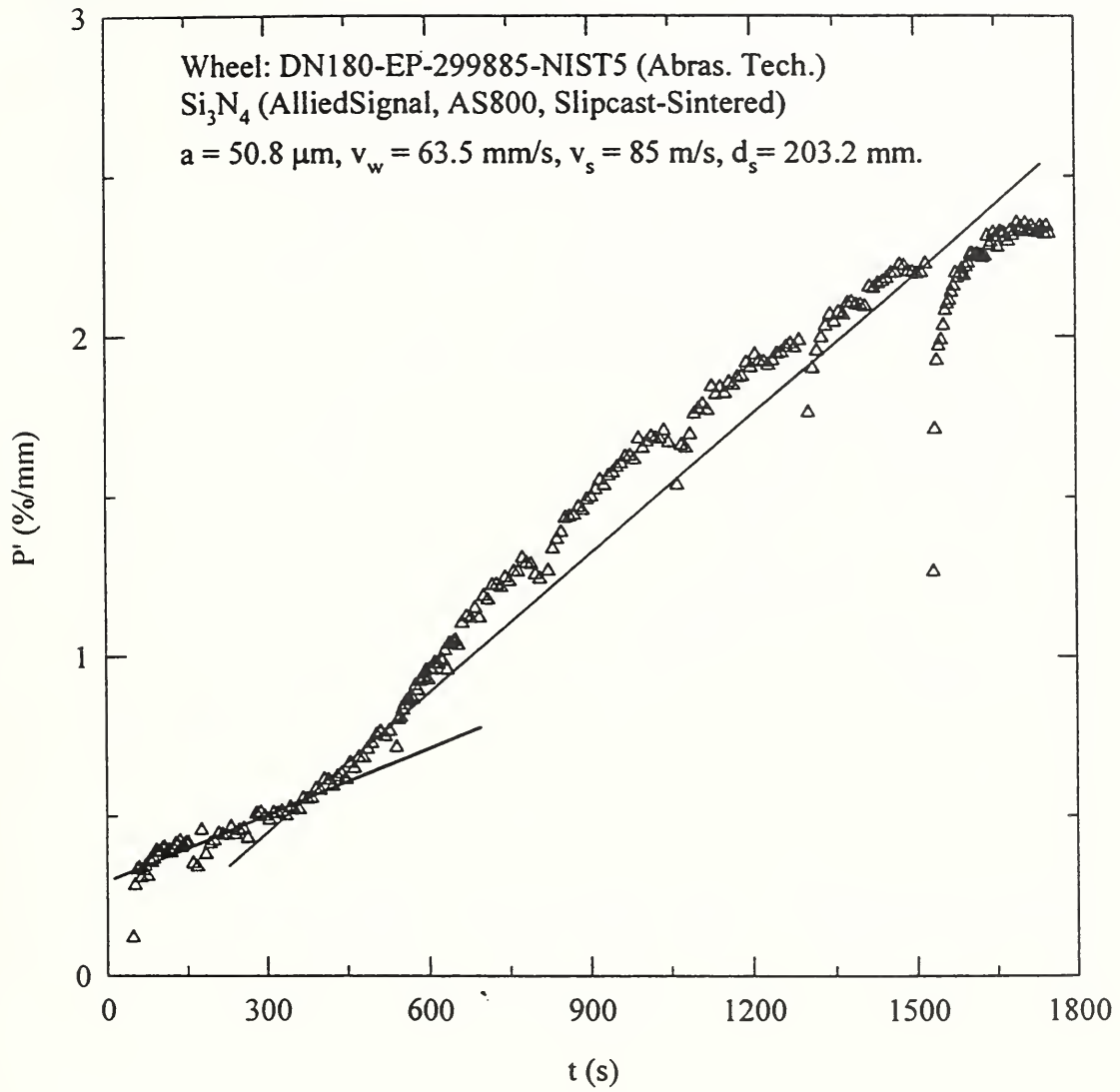
Figure 11 Grinding power per unit width versus time.
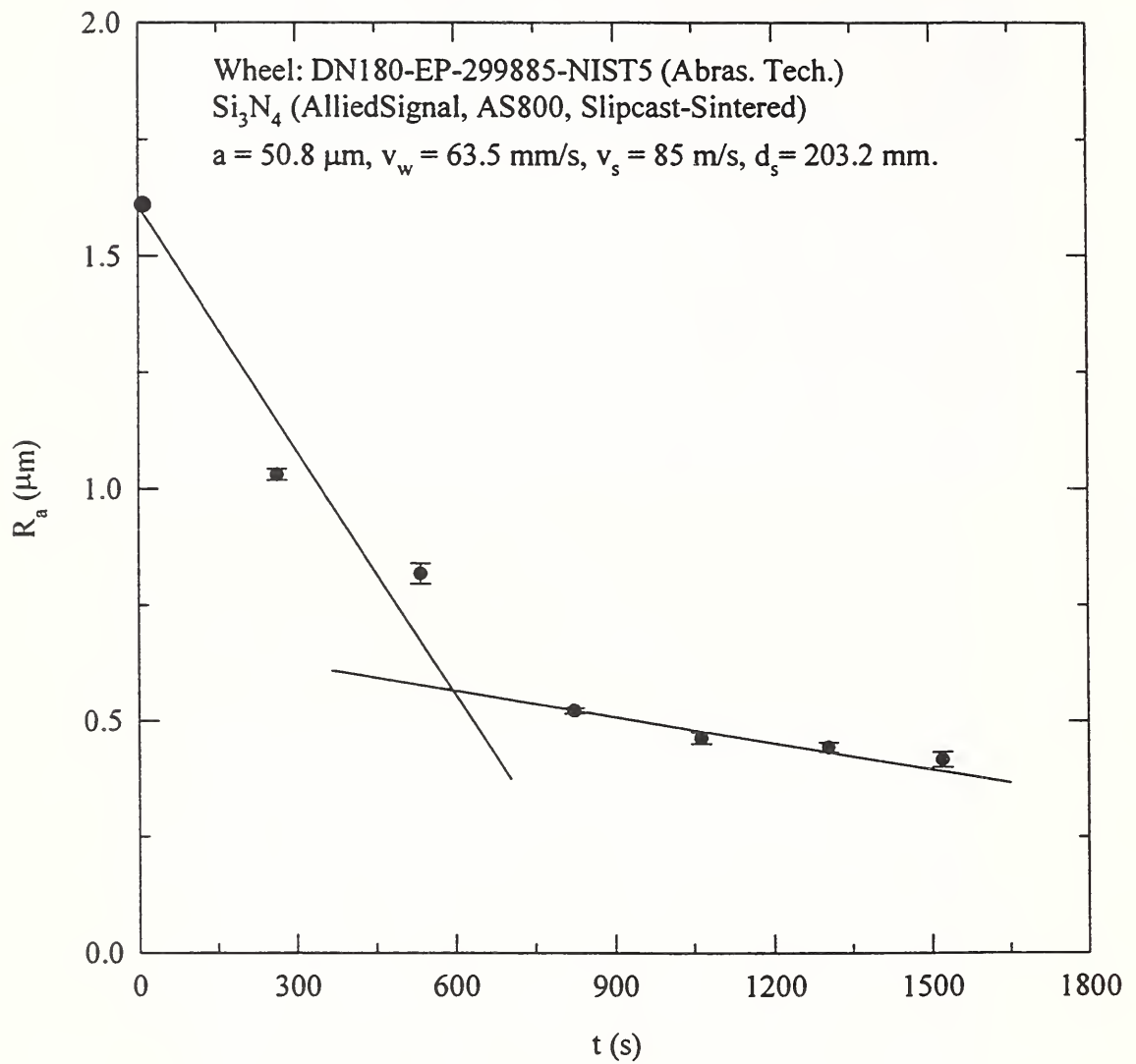
Figure 12 Plot of surface roughness (R_a) versus grinding time
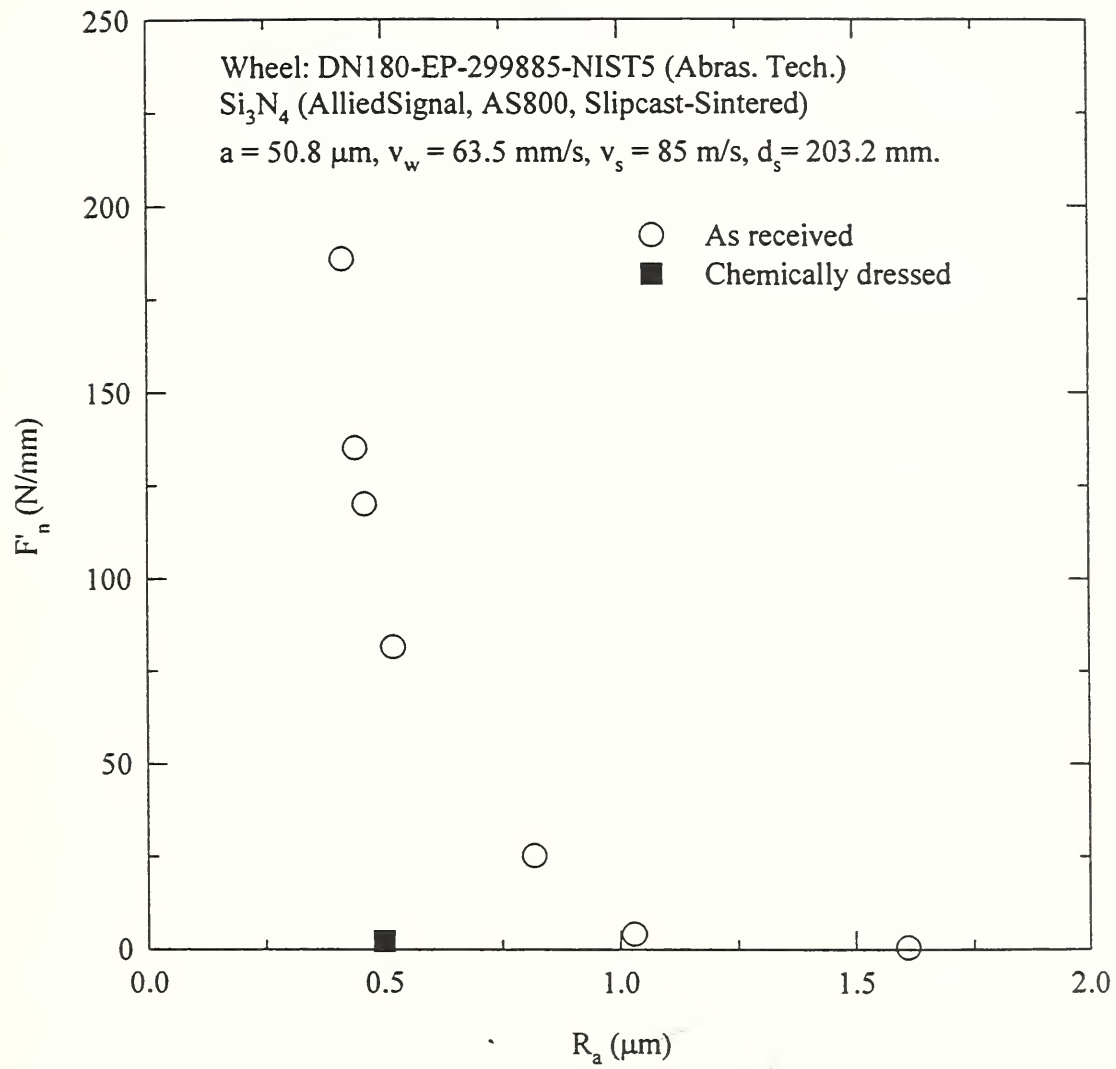
Figure 13 Plot of normal force per unit width versus surface roughness.

Worn depth

(a) t = 528 s

(b) t = 808 s

(c) t = 1048 s

(d) t = 1288 s

(e) t = 1520 s

60 μm

6 mm
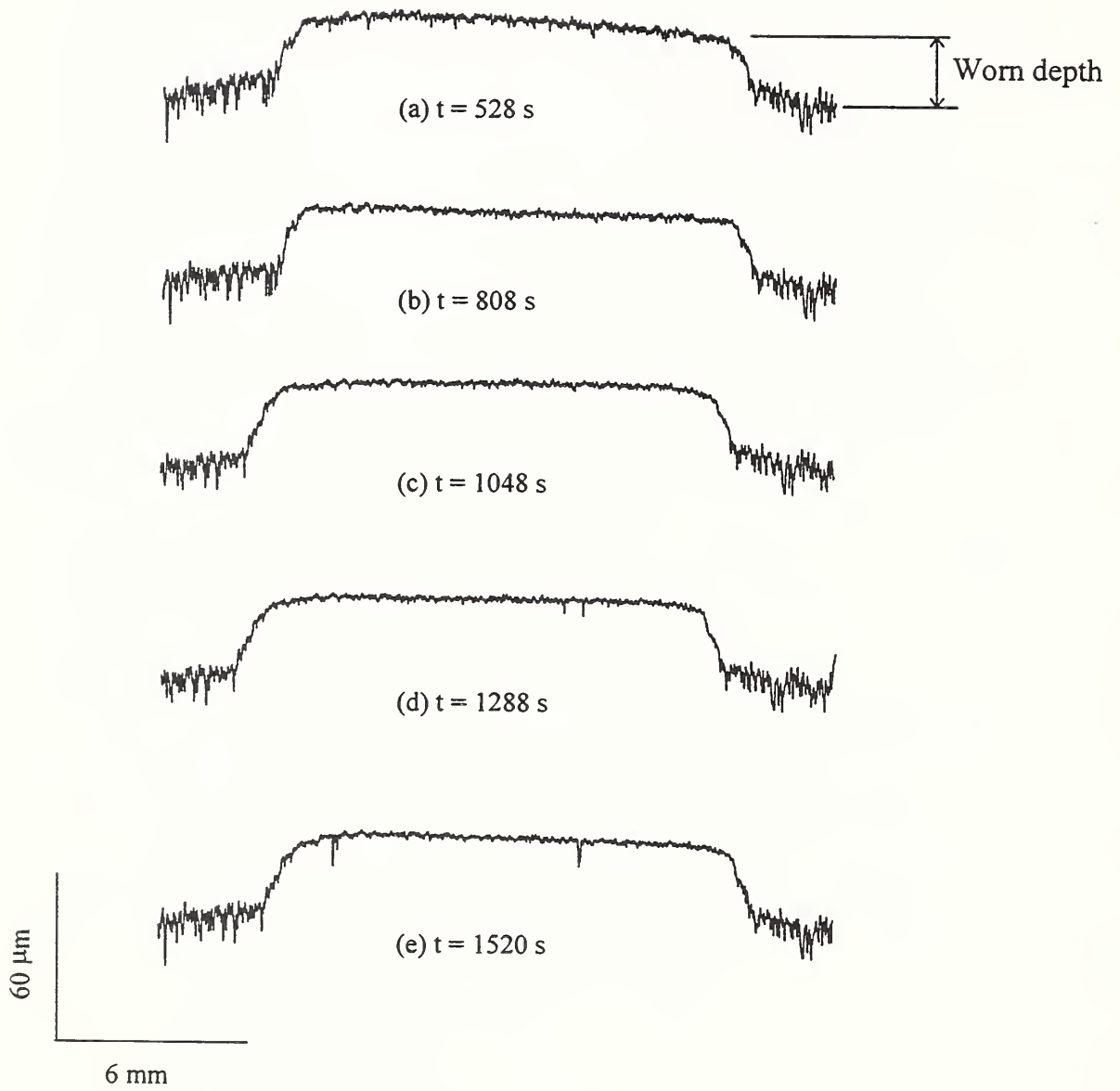
Figure 14   Surface profiles of wheel wear
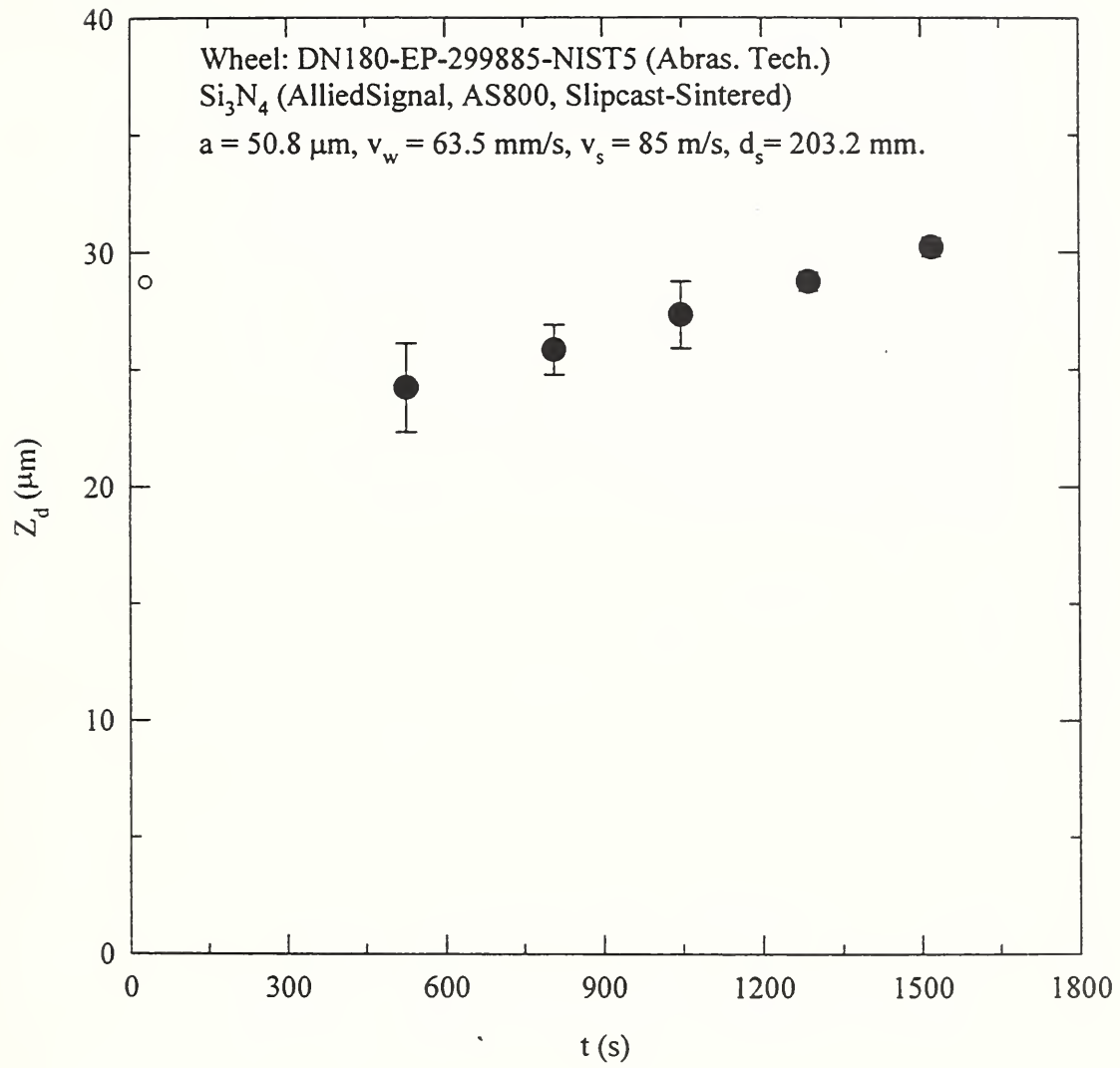
Figure 15 Plot of worn depth of the electroplated wheel versus grinding time
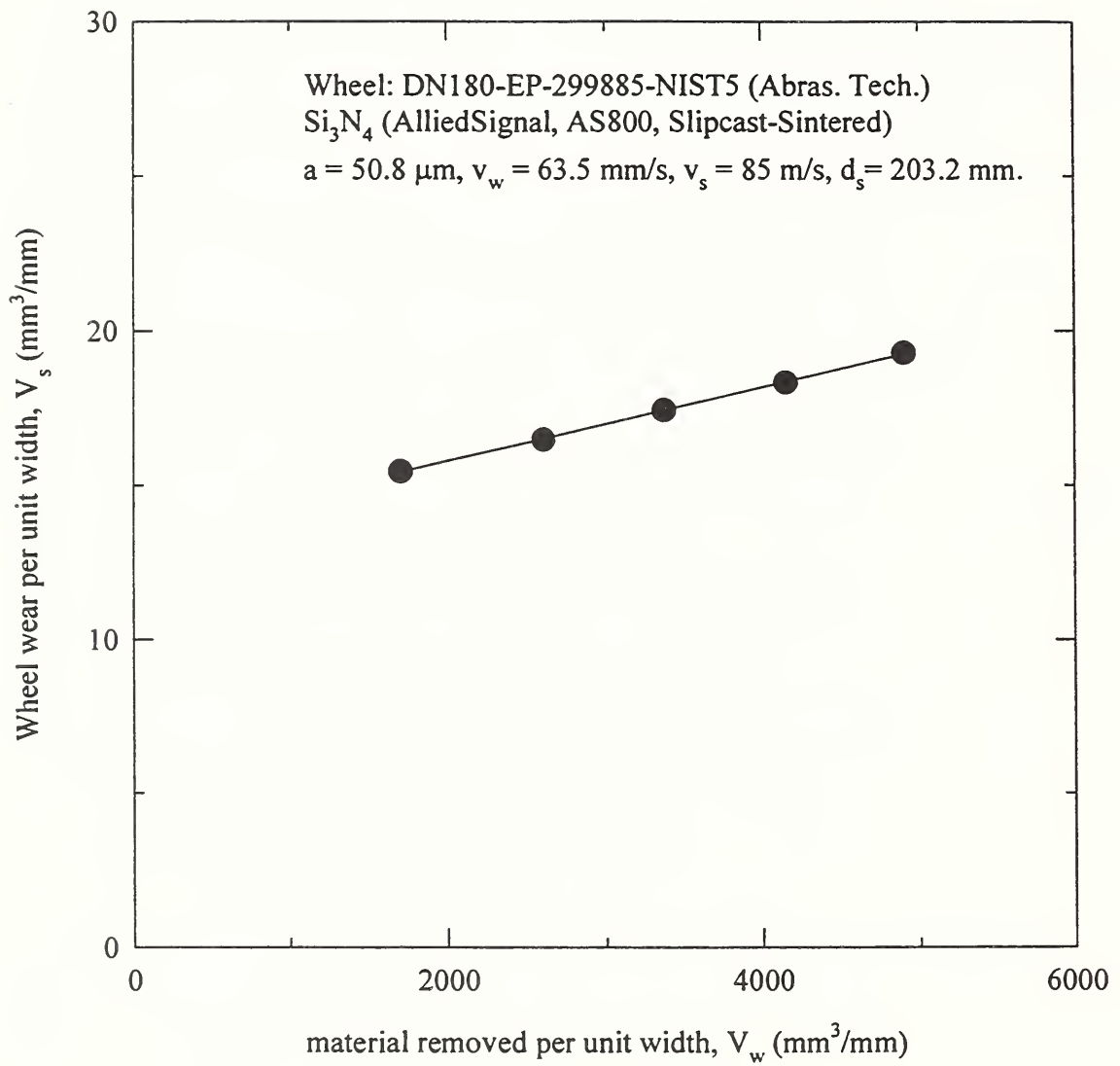
Figure 16   Wheel wear versus volume of ceramic removed.
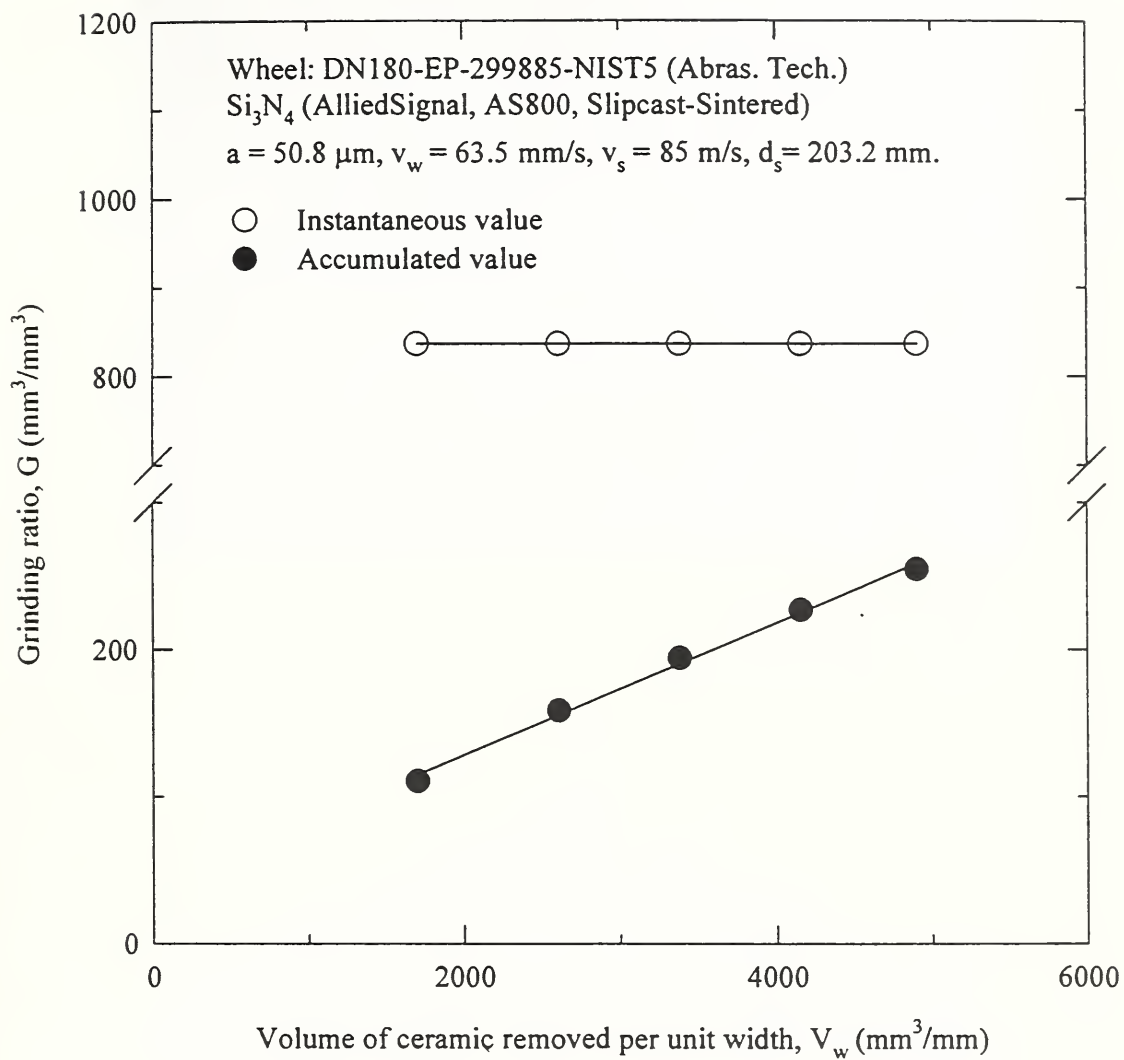
Figure 17 G-ratio versus grinding time.
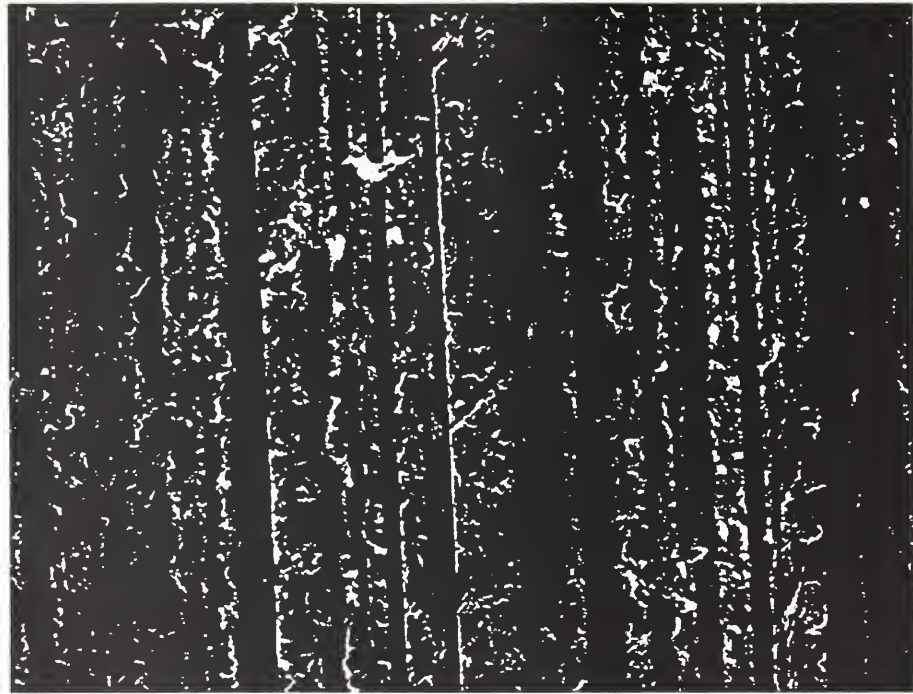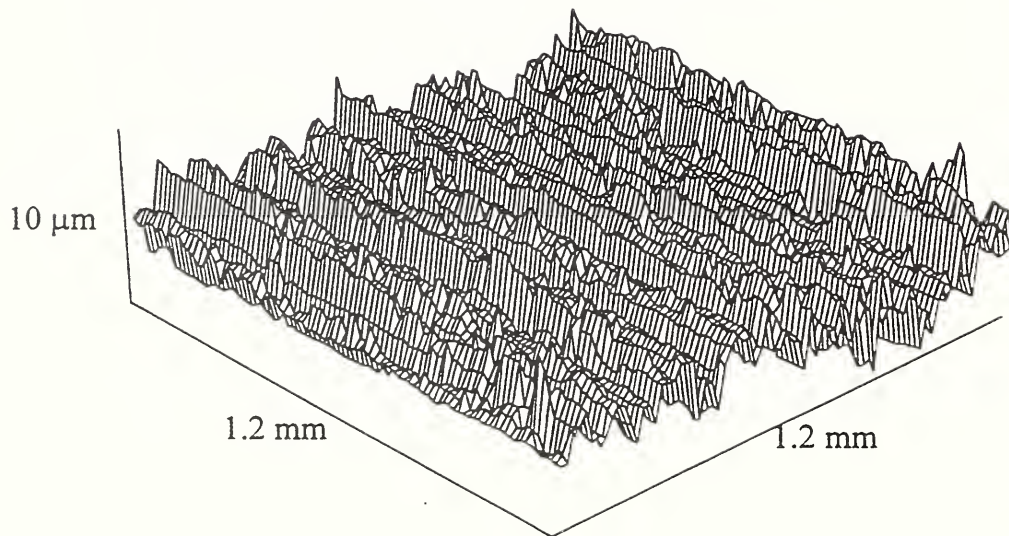
(a) SEM micrograph

(b) Surface profile

Figure 18   Ground surface after grinding time t = 264 s.
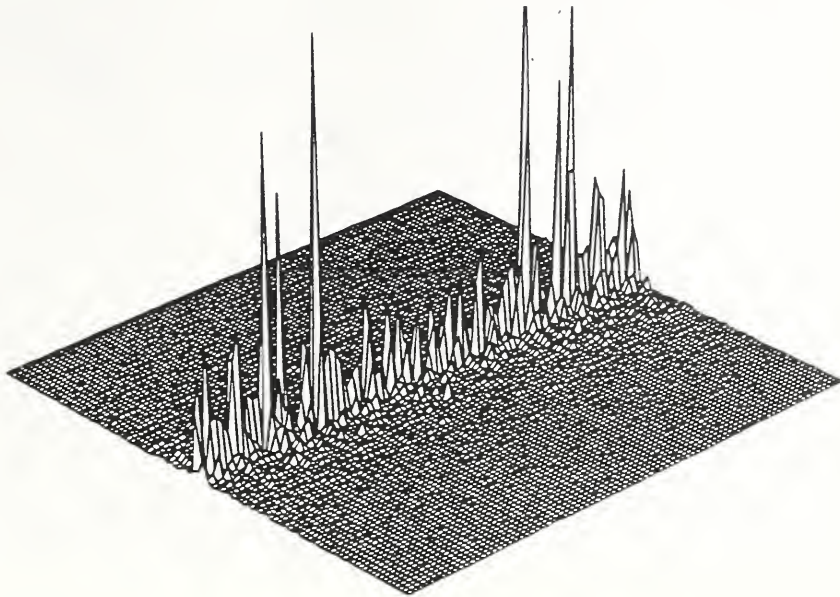
(c) APSD
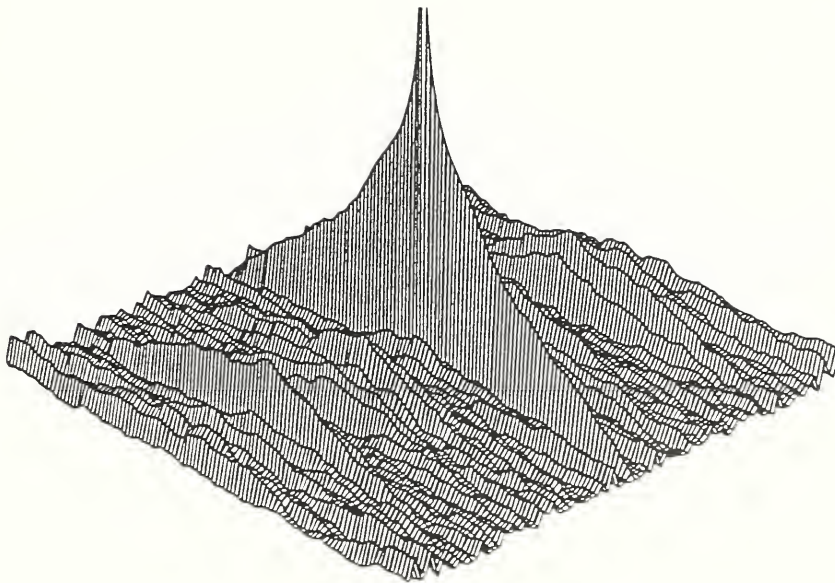
(d) AACF

Figure 18   Ground surface after grinding time t = 264 s.
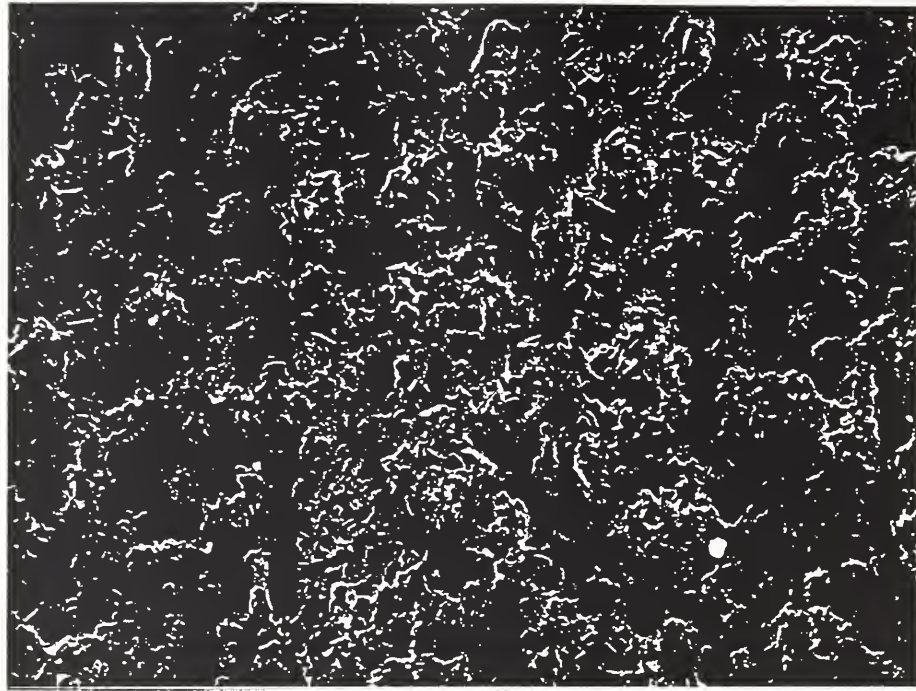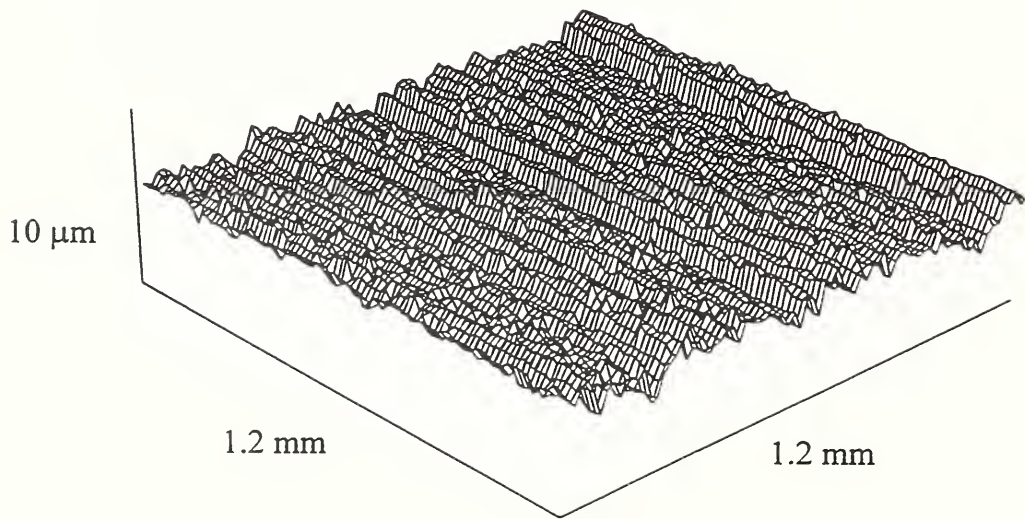
(a) SEM micrograph

(b) Surface profile

Figure 19   Ground surface after grinding time t = 1048 s.
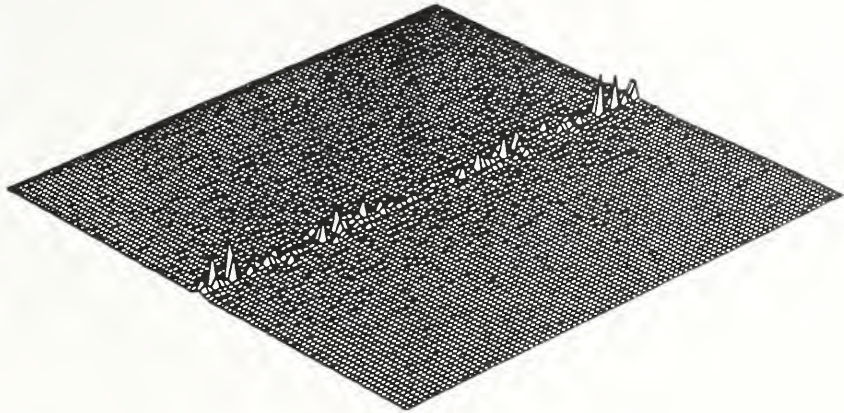
(c) APSD
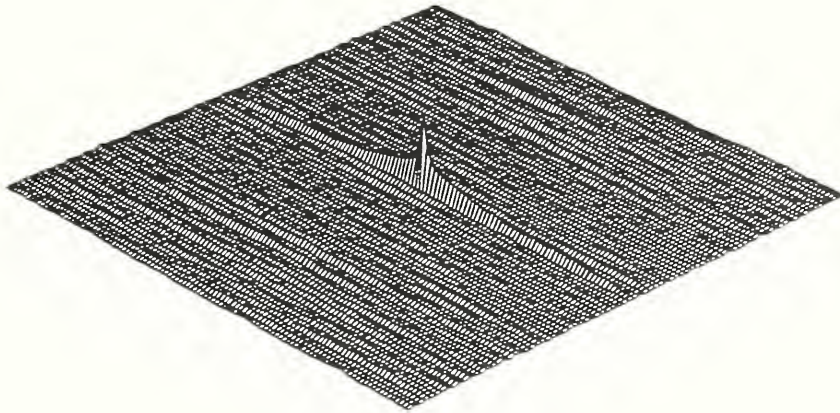
(d) AACF

Figure 19   Ground surface after grinding time t = 264 s.
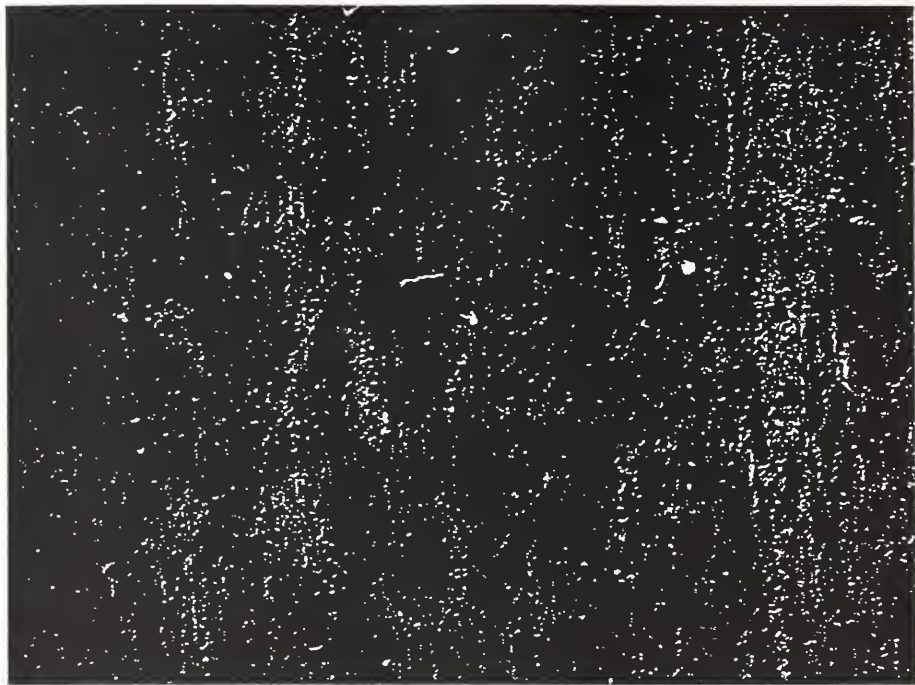
(a) SEM micrograph
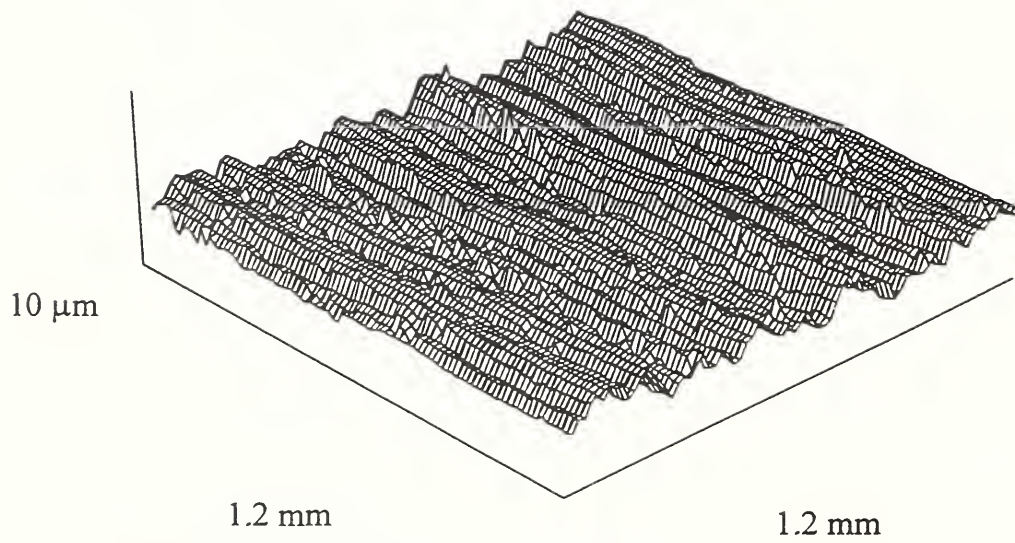
(b) Surface profile

Figure 20   Ground surface after grinding time t = 1520 s.
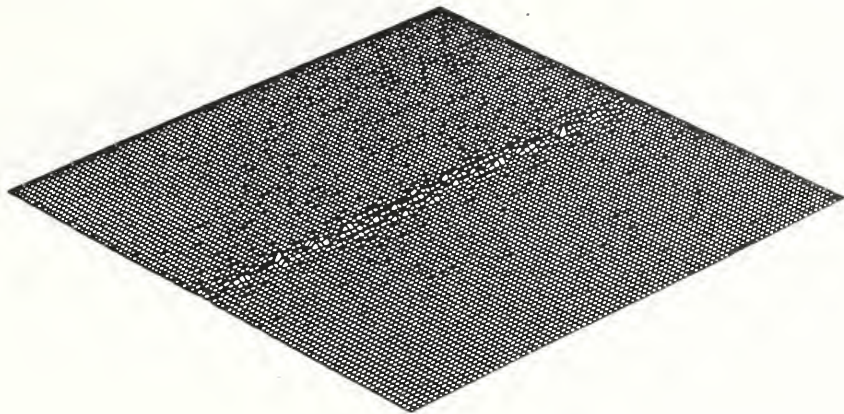
(c) APSD
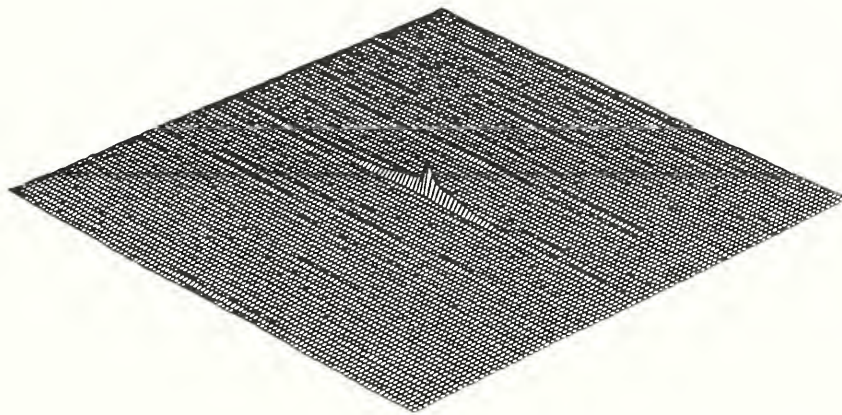
(d) AACF

Figure 20   Ground surface after grinding time t = 1520 s.
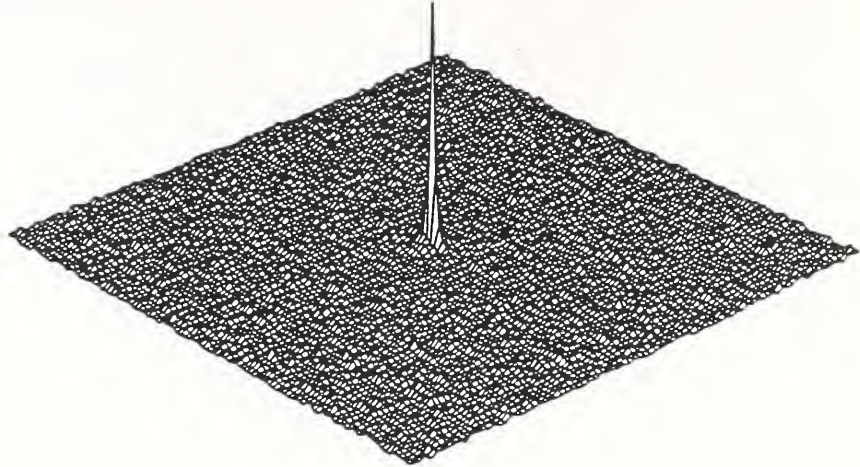
(c) AACF

Figure 21   Polished surface

Nitrides in Materials Science: Some Thermodynamic Aspects

P. A. G. O'Hare

National Institute of Standards and Technology, Physical and Chemical Properties Division, Gaithersburg, MD 20899-0001

ABSTRACT   Recent studies by fluorine bomb calorimetry are described that lead to the determination of the standard molar enthalpy of formation of silicon nitride, $\Delta_f H_m^\circ(Si_3N_4, 298.15\ K)$. Initial experiments on the combustion of hexagonal and cubic boron nitride are underway.

1. Introduction

Because of their unusual thermal stability, resistance to oxidation, and excellent physical properties at high temperatures, many inorganic nitrides are categorized nowadays as "high-tech" materials. Thermodynamic modelling is being used extensively to predict and understand the behavior of such substances and their compatibility with others. The effectiveness of this approach depends very much on the reliability of the undergirding basic thermodynamic information: much of it is, unfortunately, of questionable validity. Research work in progress at this laboratory is intended to provide a reliable database of thermodynamic quantities for the technologically important inorganic nitrides. In particular, this paper deals with recent and ongoing work on two important "high-tech" systems: (silicon + nitrogen) and (boron + nitrogen).

The thermodynamic quantity of greatest usefulness for the examination of high-temperature equilibria and stability of materials is the Gibbs free energy of formation as a function of temperature $\Delta_f G_m^\circ(T)$. This property is conventionally obtained by combining the standard molar enthalpy of formation $\Delta_f H_m^\circ(T)$ and the standard molar entropy of formation $\Delta_f S_m^\circ(T)$:

$$\Delta_f G_m^\circ(T) = \Delta_f H_m^\circ(T) - T \cdot \Delta_f S_m^\circ(T). \tag{1}$$

Here, $\Delta_f H_m^\circ(T)$ is frequently calculated by combining the quantity $\Delta_{T'}^T H_m^\circ$ (where $T' = 298.15\ K$), obtained on the basis of heat capacities, with $\Delta_f H_m^\circ(298.15\ K)$ from reaction calorimetry. Likewise, $\Delta_f S_m^\circ(T)$ is derived from the heat capacities as a function of T.

Studies of equilibria at high temperatures can also be used to derive $\Delta_f G_m^\circ(T)$:

$$\Delta_f G_m^\circ(T) = -R \cdot T \cdot \ln K, \tag{2}$$

where K is the equilibrium constant of a reaction. Such determinations of K usually require Knudsen effusion or mass spectrometry, or a combination of both techniques. Conversely, equation (2) may be used to derive $\Delta_f H_m^\circ(298.15\ K)$ from $\Delta_f G_m^\circ(T)$ by means of the so-called "third-law" calculation. That requires, for each component of the reaction, the function $(\Delta_0^T S_m^\circ - \Delta_{T'}^T H_m^\circ / T)$, derived from the separately measured molar heat capacity $C_{p,m}^\circ(T)$.

The present paper outlines our recent measurements of $\Delta_f H_m^\circ(298.15\ K)$ for two forms of silicon nitride, $\alpha\text{-}Si_3N_4$ and $\beta\text{-}Si_3N_4$, and compares the results with all previous investigations. It also describes some preliminary scoping studies on the combustion calorimetry in fluorine of the hexagonal and cubic forms of boron nitride, BN. It is planned to use this technique to determine

$\Delta_f H_m^\circ$(298.15 K) of three forms of crystalline BN (hexagonal, cubic, and wurtzitic), thus to assess their relative stabilities.

Fluorine bomb calorimetry and nitrides

Combustion calorimetry is one of the traditional techniques for the determination of the standard molar enthalpies of formation $\Delta_f H_m^\circ$ of substances. Fluorine bomb calorimetry, a descendant of the classical oxygen bomb calorimetry, has been used in studies of inorganic substances that cannot be made to combine readily or completely with oxygen. For example, uranium monosulfide reacts with oxygen in a combustion calorimeter more or less according to:

$$US(cr) + nO_2(g) = UO_x + SO_2(g) + SO_3(g) + (U,S,O). \qquad (3)$$
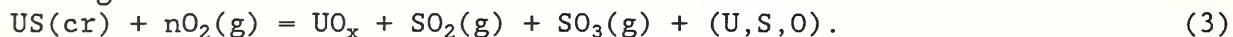
Here, ill-defined uranium oxides, a mixture of sulfur oxides, and possible ternary (uranium + sulfur + oxygen) are the products of combustion. The first law of thermodynamics requires that the initial and final states of the system be known precisely. To satisfy that demand, analytical characterization of the products of reaction (3), a daunting challenge, would be necessary.

But substitution of the more potent oxidizer, fluorine, greatly simplifies the combustion:

$$US(cr) + 6F_2(g) = UF_6(cr,g) + SF_6(g). \qquad (4)$$

If the enthalpy of reaction is denoted by $\Delta_r H_m^\circ(4)$, then:

$$\Delta_f H_m^\circ(US, cr) = \Delta_f H_m^\circ(UF_6,g) + \Delta_f H_m^\circ(SF_6,g) - \Delta_r H_m^\circ(4), \qquad (5)$$

where all the quantities in equation (5) are usually referred to the temperature T = 298.15 K. Thus, combination of $\Delta_f H_m^\circ$ of UF_6 and SF_6 with the experimentally determined $\Delta_r H_m^\circ(4)$ yields the desired $\Delta_f H_m^\circ(US)$. The auxiliary enthalpies of formation of the binary fluorides are derived from measurements of the energies of fluorination of the elements, for example:

$$S(cr) + 3F_2(g) = SF_6(g). \qquad (6)$$
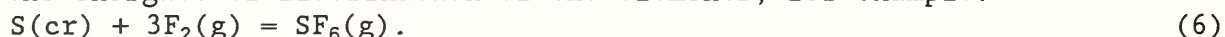
Because nitrogen forms as $N_2(g)$, the combustion reaction of nitrides is even more straightforward:

$$UN(cr) + 3F_2(g) = UF_6 + (1/2)N_2(g). \qquad (7)$$

Earlier studies revealed[1,2] only negligible quantities of NF_3 formed from Si_3N_4 and BN in flowing $F_2(g)$. In our experience, no detectable amounts of NF_3 are formed in the static atmosphere of a sealed combustion vessel. Thus, all the preliminary indications are that fluorine bomb calorimetry should be successful for the determination of enthalpies of formation of Si_3N_4 and BN.

Results for Si_3N_4

There are two crystalline forms of Si_3N_4; both are hexagonal, but the cell volume of the α modification is almost twice that of the β. When exposed to F_2 in the reaction vessel[3] of the calorimeter, Si_3N_4 did not react spontaneously; upon addition of a few mg of sulfur as fuse, rapid combustion ($S + 3F_2 = SF_6$) took place when the F_2 was introduced. Infra-red spectroscopic analyses showed that the Si_3N_4 reacted as follows:

$$Si_3N_4(\alpha,\beta) + 6F_2(g) = 3SiF_4(g) + 2N_2(g). \qquad (8)$$
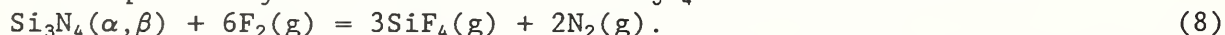
The massic energy of reaction (8) was found to be $-(28628 \pm 22)$ J·g^{-1} for α-Si_3N_4 and $-(28636 \pm 14)$ J·g^{-1} for β-Si_3N_4. These values include corrections for the impurities present in each sample; the uncertainty is twice the standard deviation of the mean. The corresponding molar enthalpies of combustion, $\Delta_c H_m^\circ = -(4018.5 \pm 3.1)$ kJ·mol^{-1} for α-Si_3N_4 and $-(4019.6 \pm 2.0)$ kJ·mol^{-1} for β-Si_3N_4, when combined with $\Delta_f H_m^\circ(SiF_4,g) = -(1615.8 \pm 0.5)$

kJ·mol^{-1},[4] yielded the desired standard molar enthalpies of formation: $\Delta_f H_m^\circ(Si_3N_4,\alpha,298.15\ K)$ = -(828.9 ± 3.4) kJ·mol^{-1} and $\Delta_f H_m^\circ(Si_3N_4,\beta,298.15\ K)$ = -(827.8 ± 2.5) kJ·mol^{-1}. These results indicate that the molar enthalpy change for the α-to-β transition is negligible within the uncertainties of the measurements. This conclusion, reported for the first time on the basis of experimental observations, is in harmony with an earlier modelling study[5] that predicted an enthalpy of transition of approximately 1 kJ·mol^{-1}. It also agrees with new, unpublished low-temperature calorimetric results of Boerio-Goates et al.,[6] who found the enthalpy difference between α- and β-Si_3N_4 to be negligible at T = 298.15 K.

Comparison with earlier studies

To the best of our knowledge, at least 12 previous thermodynamic investigations of Si_3N_4 have been described in the literature. It is surprising, therefore, that the fundamental thermodynamic properties are not yet known with any degree of certitude, and here we refer only to the most recent authoritative assessment of $\Delta_f H_m^\circ(Si_3N_4,cr,298.15\ K)$ by Gurvich et al.[7] (It is not always clear from the literature whether α- or β-Si_3N_4 was used in a particular investigation; however, our results quoted earlier suggest that it probably does not make any significant difference.)

The $\Delta_f H_m^\circ$s recalculated by Gurvich et al. are as follows {recall that our value is -(827.8 ± 2.5) kJ·mol^{-1}}: -733 kJ·mol^{-1} from Matignon;[8] -(827 ± 6) kJ·mol^{-1} from Hincke and Brantley;[9] -(835.6 ± 2.3) kJ·mol^{-1} from Pehlke and Eliott;[10] -840 kJ·mol^{-1} from McClaine et al.;[11] -(842.2 ± 3.3) kJ·mol^{-1} from Ryklis et al.;[12] -(809.4 ± 0.3) kJ·mol^{-1} from Colquhoun et al.;[13] -813 kJ·mol^{-1} from Wild et al.;[14] -(787.8 ± 3.0) kJ·mol^{-1} from Wood et al.;[15] and -(831 ± 2.5) kJ·mol^{-1} from Blegen.[16] On the basis of these values, Gurvich et al. proposed -(787.8 ± 3.0) kJ·mol^{-1} from Wood et al.[15] as the most reliable value. This was an unfortunate selection, because we found that the impurity correction made in the course of that work was greatly in error; when calculated properly, the correction causes $\Delta_f H_m^\circ$ to change to -761 kJ·mol^{-1}, almost 70 kJ·mol^{-1} more positive than our result.

Since the publication of the Gurvich et al. assessment, more recent results for $\Delta_f H_m^\circ(Si_3N_4,\ 298.15\ K)$ have been reported: between -763 kJ·mol^{-1} and -775 kJ·mol^{-1} by Andrievskii et al.;[17] -857 kJ·mol^{-1} by Pánek;[18] and in the region of -(777 ± 12) kJ·mol^{-1} to -(815 ± 12) kJ·mol^{-1} by Rocabois et al.[19]

Clearly, there is a great disparity between these values of $\Delta_f H_m^\circ$, and it was for that reason that the present study was undertaken. Our work, based on precise calorimetric measurements on carefully characterized specimens of Si_3N_4 and with a well-understood calorimetric reaction, is believed to have yielded the most reliable value for $\Delta_f H_m^\circ(Si_3N_4,\ 298.15\ K)$ obtained to date.

Application of the results

In connection with its use as a major technological material, the chemical stability of Si_3N_4 at high temperatures is of great importance. It has been well documented in the literature that, when heated, the pure nitride decomposes as follows:

$$Si_3N_4(s) = 3Si(cr,l) + 2N_2(g). \tag{9}$$

The equilibrium constant for reaction (9) is given by:

$$K(9) = \{p(N_2)/p^\circ\}^2, \tag{10}$$

where, at a given temperature, $p(N_2)$ denotes the equilibrium partial pressure of $N_2(g)$, p° is the standard reference pressure, and it is assumed that a = 1

for Si_3N_4 and Si, where a denotes activity. Since $\Delta_r G_m^\circ(9) = -R \cdot T \cdot \ln K(9)$, where $\Delta_r G_m^\circ(9)$ is the Gibbs free energy change for reaction (9), it follows that:

$$\Delta_r G_m^\circ(9) = -2 \cdot R \cdot T \cdot p(N_2)/p^\circ. \qquad (11)$$

Since $\Delta_f G_m^\circ(Si_3N_4, T)$ can now be calculated, the equilibrium pressure of N_2 as a function of temperature can be obtained. The combination of our result for $\Delta_f H_m^\circ(Si_3N_4)$ with the heat capacity values tabulated by Gurvich et al.[7] leads to the $\Delta_f G_m^\circ(T)$s shown in table 1. Those results are identical to $\Delta_r G_m^\circ(9)$; from them are calculated the equilibrium partial pressures $p(N_2)$ also given in table 1.

Thus, two vital thermodynamic quantities used in modelling studies of Si_3N_4 have now been derived reliably: the Gibbs free energy of formation and the equilibrium vapor pressure as functions of temperature. An example of the use of this information is given in table 2. Here, thermodynamic quantities are given for the reaction:

$$Si_3N_4(cr,\beta) + 3C(cr, \text{graphite}) = 3SiC(cr,\beta) + 2N_2(g). \qquad (12)$$
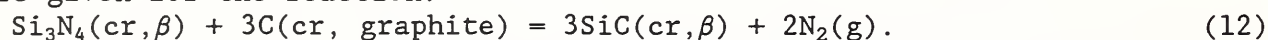
These results can be used to analyze the important stability fields of (silicon + carbon + nitrogen + oxygen).

Calorimetry of boron nitride: Preliminary observations

The second part of the investigation to be described here concerns boron nitride (BN), specifically, the three crystalline forms: hexagonal, cubic, and wurtzitic. To date, we have not procured a sample of the latter material, so our exploratory experiments have dealt with just h-BN and c-BN.

It was anticipated that BN would behave in the presence of F_2 in the calorimeter very much as Si_3N_4 had. As mentioned earlier, there is strong evidence in the literature[1,2] that BN reacts to completion with fluorine:

$$BN(s) + (3/2)F_2(g) = BF_3(g) + (1/2)N_2(g). \qquad (13)$$
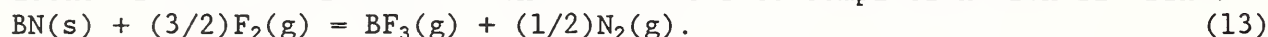
Our experiments have shown that to be true for h-BN, but not c-BN. Upon exposure to F_2, ignition of the hexagonal nitride takes place spontaneously and proceeds as depicted in equation (13). A small quantity (typically, ≈ 0.3 mg from an initial total of 400 mg) of apparently unreacted substance remains in the combustion crucible after experiments; as it can be readily weighed, a reliable correction to the original mass of BN can be made.

The calorimetric combustion of c-BN is quite different. At first, we attributed that to the physical form of the material. We have in our possession about six different specimens of c-BN, donated by various manufacturers throughout the world and provided to us by Professor Benno Lux (Technische Universität Wien). In our first series of studies, a hard abrasive material from Sumitomo (Japan) was investigated. It did not ignite spontaneously in F_2; even the use of a small amount of sulfur as kindler failed to induce a clean reaction as depicted in equation (13). In the most successful arrangement so far, the c-BN was supported in the calorimetric reaction vessel on a saucer of pure tungsten metal that rested on a nickel crucible. Approximately 3 mg of rhombohedral sulfur was used as a fuse. Upon admission of the F_2 to the reaction vessel, the sulfur ignited and that, in turn, initiated the reaction of the tungsten (to form WF_6) and the c-BN. The vigorous reaction of the tungsten was instrumental in improving the combustion of the c-BN. In experiments of this kind, the residue of c-BN, identical in appearance to the starting material, was usually about 0.7 mg (out of a total of about 300 mg).

We thought it likely that more finely divided c-BN would react more

readily with F_2, but trial runs with such a material (De Beers, South Africa) have shown that it probably will be necessary to use a tungsten auxiliary combustant for calorimetric measurements on c-BN. That conclusion was reached after experiments showed that approximately two per cent of the boron nitride remained unchanged after a typical combustion. It appears that the superior stability of c-BN vis-à-vis h-BN manifests itself even when exposed to an atmosphere of high-pressure fluorine.

REFERENCES
1. Glemser, O.; Haeseler, H. Z. Anorg. Allg. Chem. 1955, 279, 141.
2. Schumb, W. C.; O'Malley, R. F. Inorg. Chem. 1964, 3, 922.
3. O'Hare, P. A. G.; Tomaszkiewicz, I.; Seifert, H. J. J. Mater. Res., submitted.
4. Johnson, G. K. J. Chem. Thermodyn. 1986, 18, 801.
5. Durham, S. J. P.; Shanker, K.; Drew, R. A. Can. Metall. Quart. 1991, 30, 39.
6. Boerio-Goates, J. (Brigham Young University), personal communication (July, 1997).
7. Gurvich, L. V.; Iorish, V. S.; Chekhovskoi, D. V.; Yungman, V. S. IVTANTHERMO -A Thermodynamic Database and Software System for the Personal Computer. NIST Special Database 5. 1993.
8. Matignon, C. Bull. Soc. Chim. Fr. 1913, 13, 791.
9. Hincke, W. B.; Brantley, L. R. J. Am. Chem. Soc. 1930, 52, 48.
10. Pehlke, R. D.; Elliott, J. F. Trans. Met. Soc. AIME 1959, 215, 781.
11. McClaine, L. A.; Coppel, C. P.; Little, A. D. Rept. AFML-Tr-65-99. Pt. 11. 1966.
12. Ryklis, E. A.; Bolgar, A. S.; Fesenko, V. V. Porosh. Metall. 1969, (1), 92.
13. Colquhoun, I.; Wild, S.; Grieveson, P.; Jack, K. H. Proc. Brit. Ceram. Soc. 1972, 22, 207.
14. Wild, S.; Grieveson, P.; Jack, K. H. Spec. Ceram. 1972, 5, 271.
15. Wood, J. L.; Adams, G. P.; Mukerji, J.; Margrave, J. L. Third Int. Conf. Chem. Thermodynamics. Baden-bei-Wien, Austria. 1973. p.115.
16. Blegen, K. Spec. Ceram. 1975, 6, 223.
17. Andrievskii, R. A.; Khromov, Yu. F.; Lyutikov, R. A.; Zhmurov, S. A.; Galkin, E. A.; Yurkova, R. S. Zh. Fiz. Khim. 1994, 68, 5.
18. Pánek, Z. J. Am. Ceram. Soc. 1995, 78, 1087.
19. Rocabois, P.; Chatillon, C.; Bernard, C. J. Am. Ceram. Soc. 1996, 79, 1351.

TABLE 1. Gibbs free energy $\Delta_r G_m^\circ$ and equilibrium partial pressure $p(N_2)$ as a function of temperature for the reaction: $Si_3N_4(cr, \beta) = 3Si(cr,l) + 2N_2(g)$

T	$\Delta_r G_m^\circ$	$p(N_2)$	T	$\Delta_r G_m^\circ$	$p(N_2)$
K	$kJ \cdot mol^{-1}$	Pa	K	$kJ \cdot mol^{-1}$	Pa
298.15	-715.3	$2 \cdot 10^{-58}$	1200	-368.8	$9.5 \cdot 10^{-4}$
400	-676.5	$7 \cdot 10^{-40}$	1400	-292.7	0.35
600	-599.5	$8 \cdot 10^{-22}$	1600	-217.1	29
800	-522.3	$9 \cdot 10^{-13}$	1800	-132.3	$12.19 \cdot 10^2$
1000	-445.4	$2.4 \cdot 10^{-7}$	2000	-40.1	$30.34 \cdot 10^3$

TABLE 2. Thermodynamic values for the reaction: $Si_3N_4(cr,\beta)$ + $3C(cr,1;graphite) = 3SiC(cr,1) + 2N_2(g)$

$\dfrac{T}{K}$	$\dfrac{\Delta_r C_{p,m}^\circ}{J \cdot K^{-1} \cdot mol^{-1}}$	$\dfrac{\Delta_r S_m^\circ}{J \cdot K^{-1} \cdot mol^{-1}}$	$\dfrac{\Delta_r H_m^\circ}{kJ \cdot mol^{-1}}$	$\dfrac{\Delta_r G_m^\circ}{kJ \cdot mol^{-1}}$	lgK
298.15	19.960	353.198	612.100	506.794	-88.7871
400	13.543	358.172	613.808	470.539	-61.4453
500	7.599	360.543	614.859	434.588	-45.4005
600	2.450	361.460	615.355	398.479	-34.6902
700	-1.950	361.499	615.374	362.325	-27.0367
800	-5.745	360.985	614.985	326.197	-21.2982
900	-9.069	360.112	614.241	290.140	-16.8391
1000	-12.015	359.001	613.184	254.183	-13.2770
1100	-14.648	357.730	611.848	218.345	-10.3682
1200	-17.020	356.352	610.263	182.640	-7.9500
1300	-19.172	354.904	608.452	147.077	-5.9095
1400	-21.147	353.410	606.434	111.661	-4.1661
1500	-22.982	351.887	604.227	76.396	-2.6603
1600	-24.719	350.348	601.841	41.284	-1.3478
1700	-26.778	348.784	599.260	6.327	-.1944
1800	-28.693	347.199	596.486	-28.473	.8262
1900	-30.578	345.597	593.522	-63.112	1.7351
2000	-32.421	343.982	590.372	-97.592	2.5488

Session V. Stamping

An Overview of the Agile and Precision Sheet Metal Stamping ("Near Zero Stamping") Program
E.V. Vahala, Auto Body Consortium; X. Wu and J. Shi, University of Michigan

The Status of the Springback Predictability ATP Project
D. VandenBossche, Chrysler Advanced Manufacturing

Internal NIST Research Complementing the "Springback Predictability" and "Agile Precision Sheet Metal Stamping" Projects
L.E. Levine, R. Thomson, G.G. Long, D.R. Black, and R.J. Fields, NIST Materials Science and Engineering Laboratory

An Overview of the Agile and Precision Sheet Metal Stamping ("Near Zero Stamping") Program

Ernie V. Vahala [1], Xin Wu [2], and Jianjun Shi [3]

Near Zero Stamping Consortium [4]

ABSTRACT

This article provides an overview of the "Agile and Precision Sheet Metal Stamping - Near Zero Stamping" program, which is jointly sponsored by the NIST-ATP program and Near Zero Stamping Inc. The goal of the program is presented, and the major projects/tasks in the program are overviewed. In addition, a brief summary of the research and development progress are provided.

1. INTRODUCTION

Die and stamping manufacturing processes is one of the most critical manufacturing processes which has great impacts on the motor vehicle manufacturing. In this vital area, fundamental issues such as dimensional precision and agility of sheet metal stamping are grossly under-researched. The absence of academic research and industrial development is related to the intimidating complexity of these problems.

The Near Zero Stamping (NZS) Consortium was established with support from ATP and Consortium members. This program is to develop a new generation of sheet metal stamping technologies to achieve precision and agility in sheet metal stamping.

In the past year, the "NZS" program integrates suppliers and consumers of sheet metal stampings with researchers from universities and national lab, to attack these problems and ensure successful research, development, validation, and commercialization. Significant progress has been made since the program started in January 1996. This paper will first provide an overview

of the NZS, and then summarize the major progresses and accomplishments made in the past years.

2. OVERVIEW OF THE NZS PROGRAM

2.1. Background Of The Proposed Program

The overall goal of the program is to significantly improve the competitiveness of domestic die and stamping manufacturers. Two principal research and development objectives are: **(1) Precision:** To improve the dimensional integrity of stamped sheet metal parts to achieve sub-millimeter accuracy, and **(2) Agility:** to reduce the time required for sheet metal die design, try out, and production by thirty percent (30%).

The aforementioned overall goal represents the need to maintain and improve the competitiveness of domestic automotive body manufacturing industry. A typical cycle of the stamped sheet metal production process is composed of product design, die processing, die design, die construction, die tryout, production, and subassembly, as illustrated in Fig 1. We briefly outline some examples of how the proposed R&D will drastically advance the current state of the art.
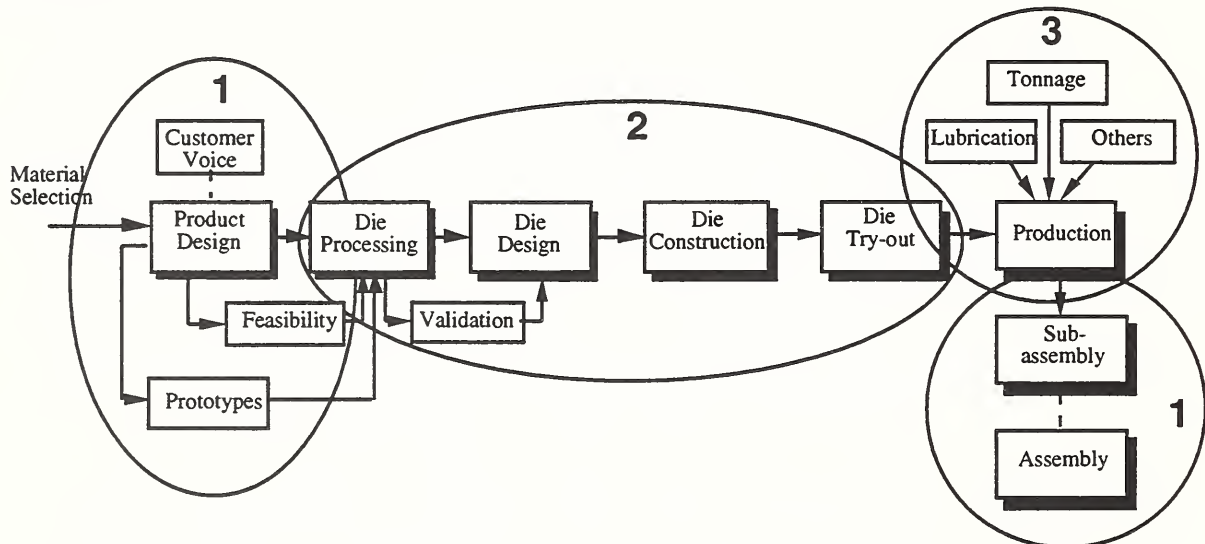
Fig 1. A typical sheet metal part development process

- Very few parameters of the stamping operation have been studied to determine their effects on the final product. Existing studies have primarily concentrated on avoiding catastrophic anomalies such as splits and wrinkles. In the proposed program, design of experiments and sensitivity analyses will be used to assess the effects of, and complex interactions between significant variables. In contrast to the current state of the art, this research will focus on controlling product variation instead of simply avoiding splits and wrinkles. By determining which process variables have significant effects on the dimensional integrity of stamped parts, quantifying these effects, and carefully controlling the variables, the variation can be reduced to near zero levels.

- Die tryout is a lengthy, costly, experience-based trial-and-error process. To make this process more efficient and less costly, a model-based, computer-aided die tryout system will be developed. This system will employ analytical and numerical models, along with a

database of current knowledge, to guide the die tryout process toward completion more quickly with less iteration and less reliance on human experience.

- A vast amount of stamping process information is widely available but not used to evaluate the stamping process. This program will consider strategies for making better use of these readily available measurements and for obtaining other process measurements related to product quality. This information will be statistically processed, both off-line and on-line, in a stamping process monitoring and diagnosis scheme.

To ensure the successful fulfillment of proposed research and development objectives, the NZS program deals with all facets of the sheet metal part development processes (Fig 1). After comprehensive study among the joint venture members, three major technical thrusts have been identified:

1) **Integrated Design for Stamping and Assembly** is to develop a product design procedure and evaluation system, which optimizes the integrated die development, stamping, and assembly processes. The optimization will minimize stamped-panel dimensional variation and reduce the development time for sheet metal. This is a primary objective of the "Near Zero Stamping" program.

2) **Agile Stamping System Design and Tryout** is to focus on improving the agility of the overall stamping system including die design, construction, try-out/modification, and material handling. This is the first attempt to address the agility at a system level, and will generate global impacts on the sheet metal stamping industry.

3) **Intelligent Stamping Process Monitoring and Control** is to concentrate on developing an in-process sensing and measurement strategy for monitoring, diagnosing, controlling, and maintaining stamping processes. These techniques and methodologies will be applicable to both existing production systems and future stamped sheet metal manufacturing systems.

2.2 Strategic Technical Plan

The technical plan of the NZS includes three research projects that address the most critical challenges to achieve the precision and agility goal of the program. These projects are:

Project 1.   *Integrated Design for Stamping and Assembly*

Project 2.   *Agile Stamping System Design and Try-out*

Project 3.   *Intelligent Stamping Process Monitoring and Control*

A brief discussion on each project is presented in this section. Each discussion provides the current industrial practice, proposed project objective, potential deliverables, and R&D tasks.

Project 1: Integrated Design for Stamping and Assembly

<u>Current Industrial Practice</u>:

(1) *Isolation of design expertise hinders competitiveness:* In the past several decades, manufacturing industries, including US auto parts stamping and assembly, have experienced some technical advancement. Concurrent with the technological developments, the design

methodology for stamping and assembly processes has changed some in recent years. However, within many companies these design changes are made in an isolated manner. This isolation is multiplied by the large number of independent auto parts suppliers who are also responsible for the stamping die design of the parts they supply. Although diversity in design expertise and experience can provide a driving force for competitiveness in stamping, cooperation could quickly optimize current design processes.

(2) *Lack of communication between part designers and manufacturers slows the production cycle: Sheet* metal parts are often designed by product designers who understand the functional requirements of the stamping parts, but have limited manufacturing experience. Consequently, difficulties often arise in the manufacturing stage, resulting in longer die manufacturing and tryout times, as well as poor formability and high dimensional variation. It takes an estimated average of 50 weeks for a single stamping part to move from the part design to the first acceptable, consistent batch. Out of these 50 weeks, 16 to 25 weeks are for die tryout, since ease of manufacture is not rigorously considered during part design. Obviously communication and knowledge sharing between part designers and manufacturers could significantly reduce the overall processing time and improve product quality. The concept of *Design for Manufacturing* is becoming accepted, but fulfilling this requirement is not an easy task. Stamping involves complicated deformation modes (including various bending, flanging, stretching, and deep drawing modes, and their combinations), and many processing variables and their effects are still not fully understood.

(3) *Inadequate tools make die design an art in practice:* Die shape cannot be determined by part geometry alone. Many complex factors, such as spring back, material variables and formability, lubrication, tonnage, etc. should be considered in die design. Thus, predicting the product dimensions from the die dimensions remains a technical challenge. On the long journey from art to science, simple analysis once prevailed over traditional experience and now modern numerical computing techniques are used to solve more complex geometry. To date more than 10 commercial finite element analysis (FEA) packages are available for stamping. However, due to the complexity of the stamping process and the many unknown factors such as material behavior and frictional boundary conditions, discrepancies between prediction and reality often occur.

(4) *Understanding key stamping process variables is imperative to achieving "near zero stamping":* Sheet metal stamping is a very complex process influenced by many variables. These variables can be grouped into six categories: blank, press, work material, die, lubrication & miscellaneous, and their interactions. The lack of understanding about stamping process variables and their relationships affect both sheet metal die designers and stamping manufacturers. At present, highly skilled and experienced "experts" design dies using an iterative process of development, tryout, and modification until a suitable part is made. These trial and error tests in die design and try-out compensate for the lack of knowledge, but they are tremendously expensive in terms of time, effort, and materials. In addition, lack of understanding about these variables and their relationships also impacts on stamping production. As an example, when a new roll of material is used in production, sudden splitting and buckling may occur. To continue producing good parts, the die setup may require hours of adjustment. A thorough study on the influence of these variables and their interactions has fundamental significance in stamping.

(5) *Hemming process and hemmer design is an under-researched area:* Hemming is a combined bending and joining process. As one of the last deformation processes in subassembly,

hemming can significantly influence the dimensions of the final component. However, the importance of hemming has been ignored, and the process has never been well characterized. A thorough study of hemming would improve the hemming process and give a scientific explanation for the process optimization.

Objective:

To develop design methodologies and evaluation methods for stamping processes and subassembly architecture to achieve sub-millimeter dimensional variation and eliminate engineering changes.

Deliverables:

(1) Detailed technical knowledge/documentation summarizing the analysis and modeling of critical parameters in sheet metal stamping.

(2) A comparative computer software package called Computer Aided Formability evaluation (CAFE) System which includes a knowledge-based expert system and analysis modules that can be linked with the existing CAD package to evaluate the formability of a designed part.

(3) A computer framework that simulates the hemming processes and directs the engineer toward optimal designs of flanges, hemming dies, and automation through a rule base.

Research & Development Plan

Task 1.1. Optimized Design Process for Stamped Sheet Metal Components and Assemblies. The objective of this task is to develop a standard optimized design methodology and architecture for sheet metal stamping and subassembly to achieve "near zero" dimensional variation. To achieve this, the task team will bench mark several existing product design processes, identify the major milestones in the body development, conduct the "buy-in" process that interacts with the big three companies. In addition, a quantitative study on optimal body structure design and joint design will be carried out by the task team, with focus on the dimensional accuracy and process stability. A computer-aided design supporting system will be developed based on the proposed study. The overall procedure in this research is summarized in Figs 2 and 3.
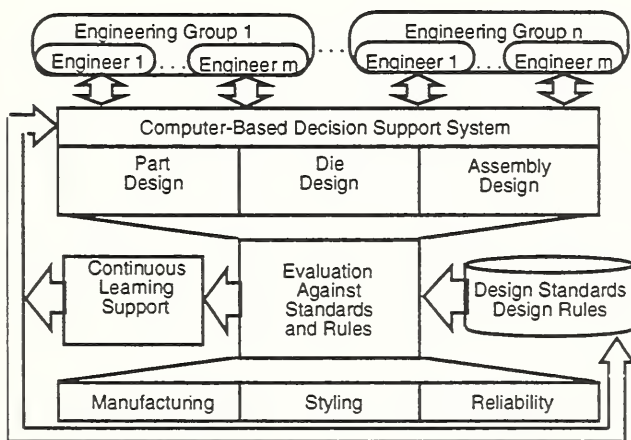
Fig 2. Implementation of New Design
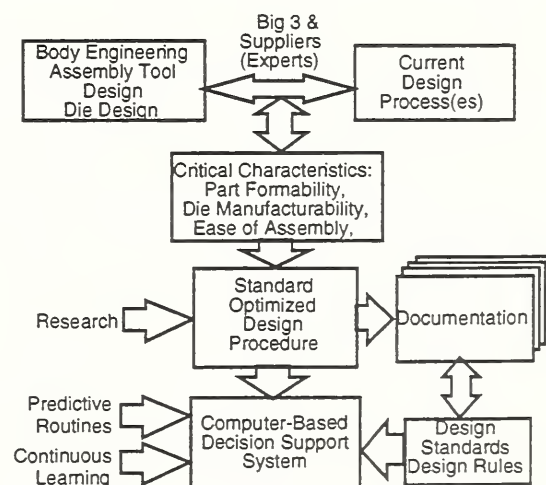
Fig 3. R&D Plan and Procedures

Task 1.2. Computerized Simulation and Evaluation System for Stamping Design. A strategy combining classical analytical method, FEA and knowledge-based system has been considered

233

as a solution for better accuracy and cost saving in computation. This method first provides an approximate analysis for global strain description, and then performs an FEA simulation for detailed deformation analysis. In combination, these methods provide a "tool kit" for modeling die design, and prepare the way for the continuous improvement of modeling. A overview of the task is shown in Figures 4 and 5.
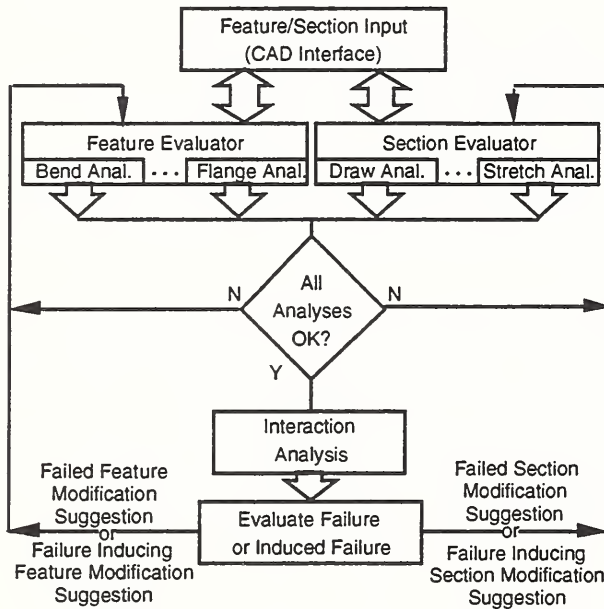
Fig 4. Computer Aided Formability System Structure
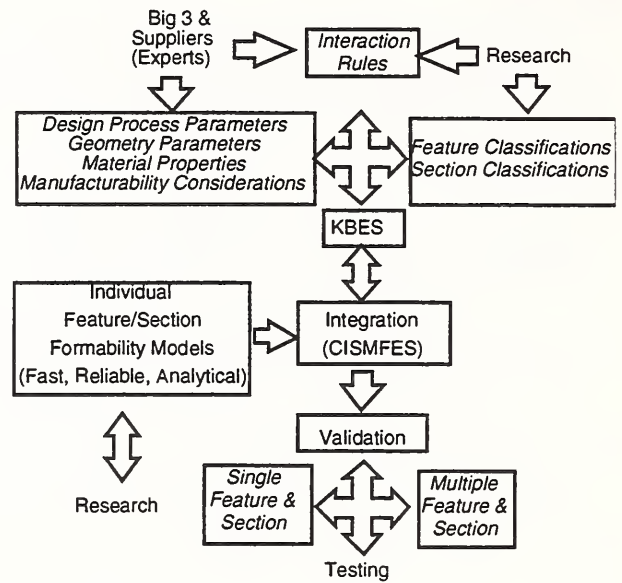
Fig 5. R&D Plan.

Task 1.3. Sensitivity Analysis of Critical Parameters. There are many parameters which have a significant impacts on the stamping process, and their relationship are normally complex. The objective of this task is to make the stamping process more predictable and controllable by identifying and prioritizing critical parameters responsible for stamping dimensional variation. In this task, typical parts (e.g., door inner panel, hinge pillar brackets) are selected for study. The research team is to identify and prioritize critical parameters in the process, to study the material behavior under various strain path and history, to conduct designed experiments in various stamping facilities. Some examples of those parameters and R&D strategy are shown in Figures 6 and 7.
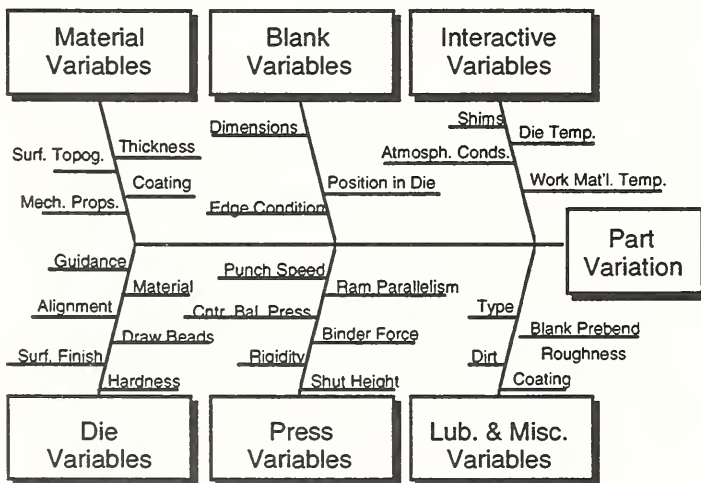
Fig 6. Fish-bone chart of process variables contributing to stamping part variation
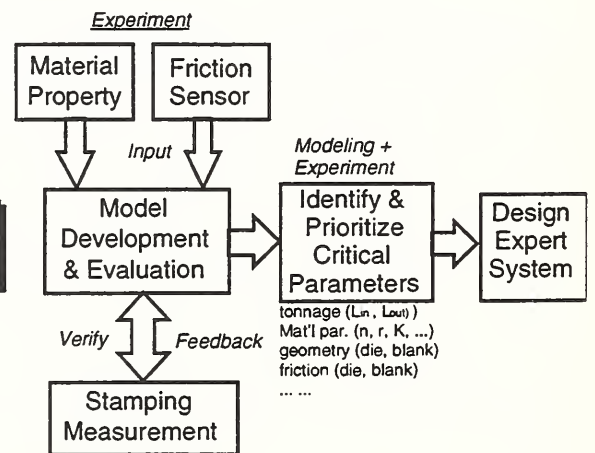
Fig 7. Task 1.2 R&D plan.

Task 1.4 Product/Process Optimization Considering Key Hemming Characteristics. Hemming is the last forming process in the major closure panel stamping/assembly. It has a significant effect on the panel quality and integrity. Compared with other forming processes, hemming has unique characteristics. For example, through sequential die motions and multiple flange bendings, two parts are joined together by hemming operations as shown in Fig 8. Since very high local deformation occurs along the curved edge, the rigidity of the joined component is significantly increased, and the final shape is confined by the dies after hemming. On the other hand, since most regions (except the edge) remain in elastic condition during hemming, high residual stress and spring back may occur whenever the incoming parts do not match the hemming die profile or hemming dies are designed improperly, causing shape distortion. In this task, a experimental hemming machine is built to conduct designed experiments to verify a hemming simulation model, which is developed by the research team. Based on the study and developed model, a hemming optimal design rule and process will be studied. Some related processes, such as gluing/bonding will also be studied.
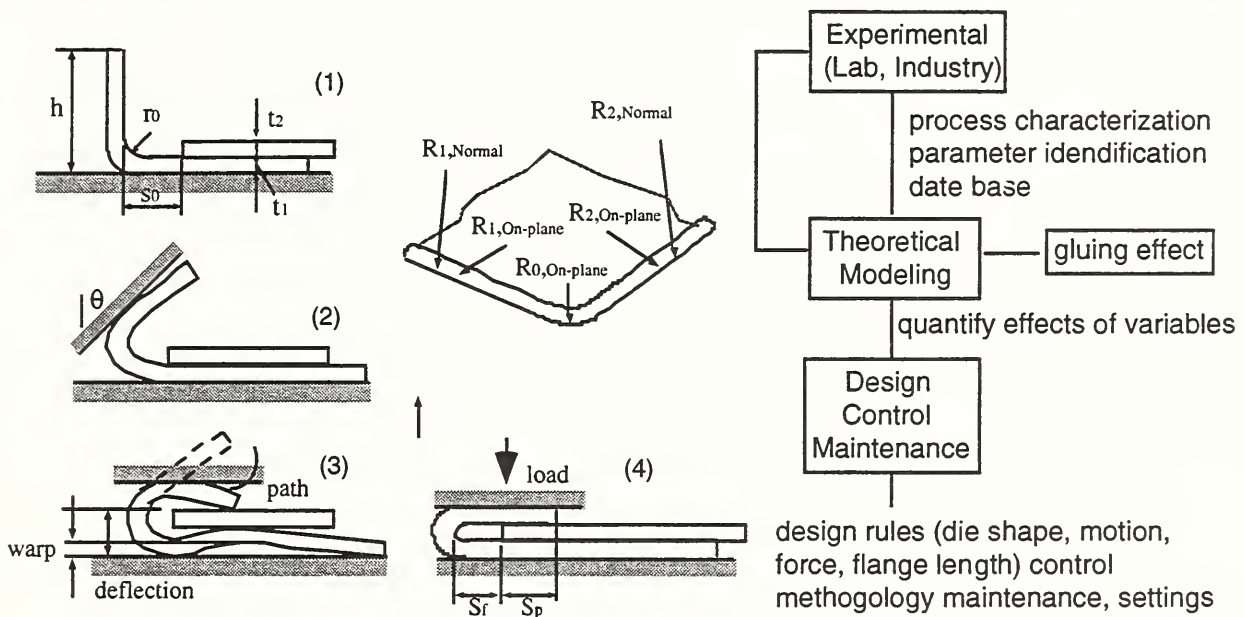
Fig 8. Schematic of hemming process and parameters

Project 2. Agile Stamping System Design and Tryout

Current Industrial Practice:

(1) *Die tryout is the major obstacle to achieving agile precision stamping:* In current practice, highly skilled and experienced "experts" design dies using an iterative process of development, tryout, and modification until a suitable part is made. Product, process, and tooling are designed based on precedent designs using the skills and judgment of these experts. The designs are then refined in die tryout through successive tests and modifications. Tryout is a non-value added activity which adds greatly to the cost and lead time of product introductions. Even when initial tryout is complete, there is no assurance that the part, the process, the material, or the setup has been optimized to provide a robust system. This often requires a secondary tryout. Furthermore, tryout information from various programs is not accumulated. The reasons behind successes or failures in part design, die design, and processing have not been successfully analyzed or documented.

(2) *Strain information collection and analysis is time consuming and incomplete:* In the current die tryout practice, a classical circle grid strain analysis may be used for trouble shooting, and to adjust the die setup under production conditions. The entire circle grid analysis includes chemically etching circle patterns on the blank surface, blank stamping, followed by circle measurement (using a flexible ruler and visual justification). This process takes several hours or days, thus its application is mainly for the first die tryout stage (by die makers). However, due to the lack of alternatives, this circle grid technique is also used in the secondary die tryout stage (in the stamping plant), slowing the response to stamping problems. In addition, due to the manual nature of the circle grid analysis technique, only limited regions with high strains are measured. This tactic yields incomplete information, which can only provide a comparison of the maximum strains with the metal formability limit for preventing the two extreme cases in stamping, i.e. split and wrinkle.

(3) *Separation of material handling, die design and try-out creates lost lead time, higher costs and lower quality:* Each stamped automotive part is produced at stamping plants in mass quantities. Parts must be moved between production lines and between stations within a line. During the design and tryout phase, the focus is on the intermediate product from each piercing or drawing stage of the stamping process. However, once the die and process have been satisfactorily designed, implementation on production automated transfer presses often results in a need for further die modification. Such modification is required to eliminate die-to-automation interference. These modifications result in die tryout being extended, often into the production stage. Thus, focusing on the material handling system in the early stages will significantly impact the agility of sheet metal stamping.

In addition, parts must be protected from damage during handling while minimizing the time and effort required to handle them. Currently, the major automakers spend $200 million on tooling systems for each new model. With an average of 10 new car models per year, an investment of 2 billion dollars is involved in the die and tolling systems. Material handling is still one of the top five causes of part dimensional variation [Shi, 1994]. There are two key shortcomings of the current material handling technology. The first relates to communication among presses and handling equipment, which is essentially nonexistent. As a result, material in a shop spends over ninety percent of the time waiting to be moved and processed (Sprow, 1991). Presses and assembly stations are not fully utilized because each acts as an independent system rather than being integrated with material handling equipment to coordinate their tasks. The second key shortcoming is related to material handling system design, which consists of setting parameters of and trajectory planning for the material handling hardware/device. Efforts to reduce the gap between the capability of presses (and assembly stations) and the equipment feeding them have been ill-fated since (a) material handling system designers have no feedback on how their systems affect part variation and (b) there exists no methodology for optimally designing these systems. Thus, agile precision stamping cannot be achieved without a thorough study of the material handling system.

Objective:

(1) Reduce the die tryout time by fifty percent by developing strain based stamping evaluation techniques integrated with formability prediction models and rule based die modification system.

(2) Define a systematic method for the quantification of physical structure and motion characteristics of transfer presses and automation equipment. Through motion simulation,

formulate constraints for use in the computer integrated die design and tryout system to completely eliminate automation related production tryout problems.

Deliverables:

(1) On-site global strain based decision-making system for sheet metal stamping;

(2) Integrated die try-out system;

(3) Motion simulation software for resolving problems of interference between the die, work material, press and automation equipment, and to automatically calculate die design constraints.

(4) An integrated system which facilitates optimization of press-automation interaction through computer aided design and evaluation of material handling design, installation, and set-up.

Research & Development Plan:

Task 2.1. Global strain based system for fast die evaluation. Next generation stamping die design is aimed at an optimal die design, which will not only eliminate splits and wrinkles in local regions, but also optimize overall strain conditions. When metal flow distributions are reasonably uniform over the entire die surface, the feasible forming window between split and wrinkle can be greatly widened and the dimensional stability can be improved. In order to provide a complete strain-stress picture and control-modification scheme, complete information on sheet metal flow is needed. This can be achieved by improving the surface measurement techniques and by refining the strain analysis algorithm to provide better precision and efficiency. A mathematical model is being developed, which will use the surface marker coordinates as input and formulate the solutions for strain distribution. Based on the previous formulation proposed by Sowerby et el (J. Strain Analysis, 1982, 17, 95) a refined formulation will be developed to take into account the curved surfaces and assure much higher accuracy and computational speed.
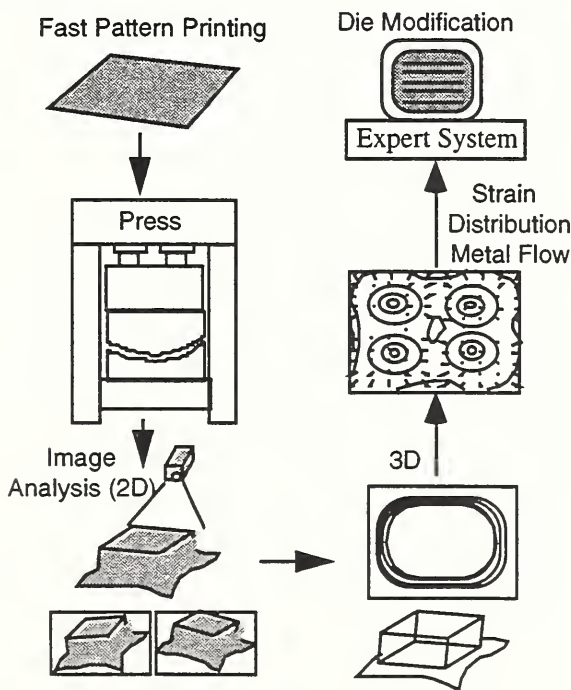
Fig 9. Strain analysis flow chart.
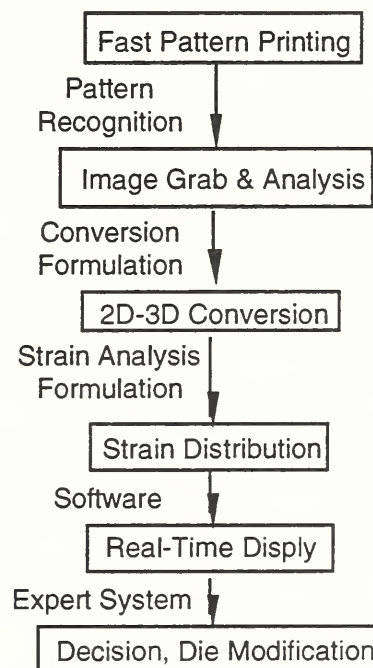
Fig 10. R&D plan for die tryout system

237

In this research, a fast marker printing technique will be developed for blank preparation. A fast global precision coordinate measurement and strain analysis system will be designed and tested. Based on the measurement, a signature analysis technique will be developed, and finally, the technique and system will be integrated in die-try out process and stamping production for die performance evaluation, design feedback, and maintenance. The principal and tasks in this study in shown in Figures 9 and 10.

Task 2.2. Agile die try-out and modification. Achieving die try-out agility (50% reduction in the die try-out time) by developing die modification rules is a high risk task because of the complexity of the die try-out problem, diversity of expertise in the domain, and inadequacy of conventional forming modeling techniques for complex sheet metal parts. However, the payoffs for die try-out agility are equally high since die try-out is the most time consuming and experience dependent task in stamping. In order to achieve this objective, a rule-based system with integration of forming modeling techniques, will be developed. One software die company, two die companies, and two OEM are involved in the task following the proposed R&D plan as shown in Figures 11 and 12.
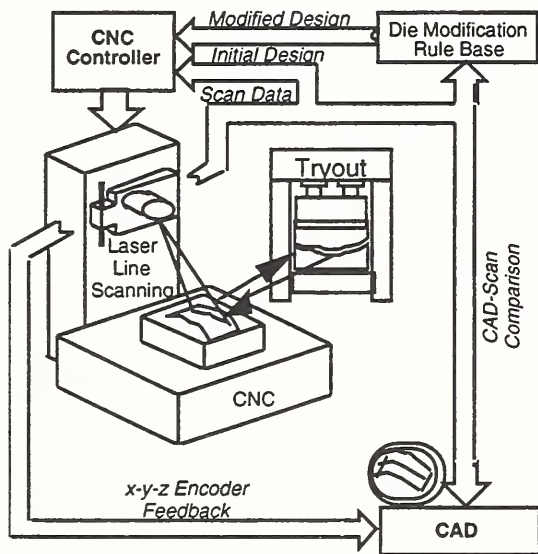
Fig 11. Computer-Aided Die Design and Tryout System Schematic.
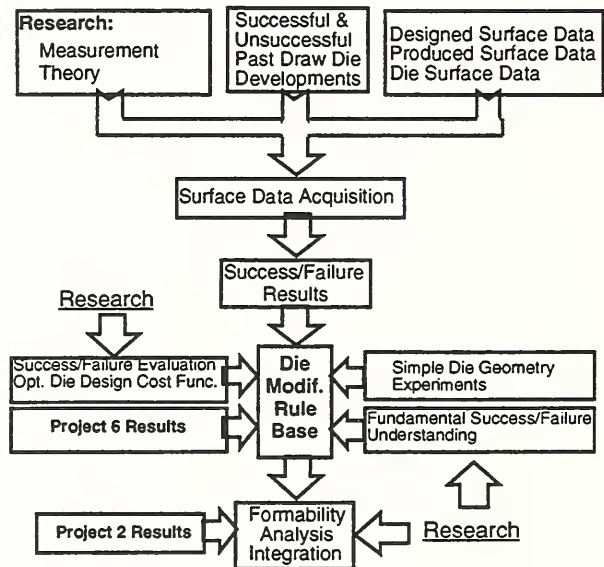
Fig 12. R&D Plan.

Task 2.3. Integration of material handling and press motion in die design. A material handling system would integrate part, die, press, and handling devices. The challenges to the material handling system design include zero deformation of the part with the highest handling efficiency and the lowest cost. According to the nature of the material handling, the task can be further divided into within tri-axis handling system, pick and place system (between presses), and between press lines (rack based material handling). In the study, some fundamental issues, such as optimization of gripper location on a part, path planning and optimization, and press/die/handling system characterization, are studied. Computer aided design and optimization toll is being developed. The framework for the research is shown in Figures 13 and 14.

238
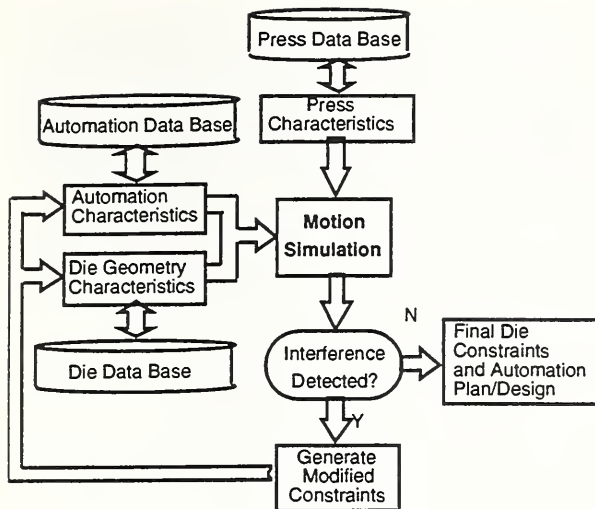
Fig 13. Motion Simulation and Die Design
Constraint Formulation System.
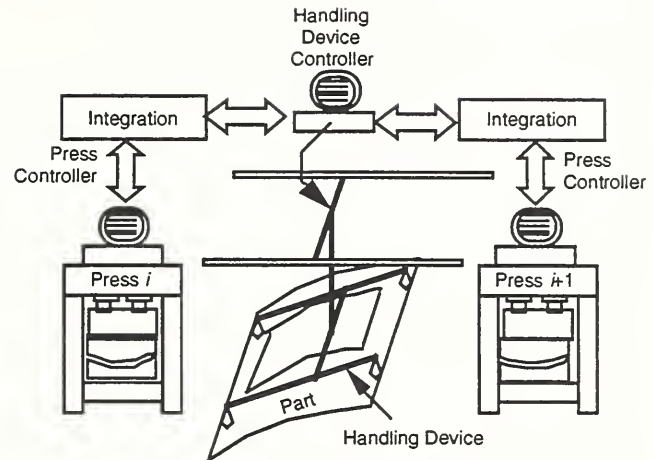
Fig 14. Optimized Material Handling System

Project 3. Intelligent Stamping Process Monitoring and Control

Current Industrial Practice:

During stamping production, many critical parameters affect the dimensional accuracy through complex interactions. These parameters can be classified as die wear/maintenance, press working status, die set up, material properties, and die-blank-press interaction. In general, these internal process parameters cannot be measured directly and their relationships are complex. Currently, stamping process control is realized by periodic dimensional inspections of the stamped parts using manually operated gauges or fixtures. This strategy cannot satisfy the need of modern stamping production and quality specifications, especially for high-speed presses. At a typical production rate of 75 pieces per minute, 4,500 bad parts would be made if a failure occurred right after an hourly inspection. Recently, on-line sensors measuring tonnages, shut heights, die temperatures, etc., have been applied to obtain inherent information about the process. Some companies (Helm; Signatures; Link Technology; Phoenix Consultancy, Inc.) have developed sensing techniques employing conventional SPC analysis integrated with signal analysis to monitor the stamping process. However, their practices primarily rely on a single type of sensor. Thus, the process control information may be incomplete which could bias the decision making process. In addition, this information is not fully utilized. For example, despite the wealth of tonnage data available, only the peak value and boundary of the signature profile are used. There is no further analysis of the detailed signature as a continuous signal. More importantly, no technology is available in either academia or industry, to correlate the stamping process signature with the process faults or stamping quality.

Similar situations can be found in the area of die maintenance. Automotive dies are very complex and their maintenance directly impacts productivity, product quality, and overall cost. Die maintenance schedules currently rely on worker experience and identification of defective part production. Neither a systematic documentation scheme for die predictive maintenance (PM) scheduling, nor a scientific method for predicting die wear currently exists. The result is that either the useful life of the die is shortened, or a broken/worn die produces bad parts and/or a production line shut down. The few existing PM software packages only perform record keeping and have no prediction capabilities. Locally catastrophic failures of dies and press parts which

239

are attributed to inadequate PM are responsible for down times on the order of twenty-five percent.

In general, current stamping process control has no in-process adjustment capabilities. Thus, the stamped part quality is very sensitive to die set up repeatability, material changes, lubrication changes etc. Errors due to worn out gib, tie rod tightness, thermal error from the press linkage and dies, etc., cannot be corrected during production. In the last few years, some research and development have been done to develop systems for controlling blank holding force in deep drawing [Nakagawa, 1994]. This work confirmed that forming defects such as fracture, wrinkles and surface distortion can be overcome through controlled forming conditions in laboratory experiments. In addition, it has been confirmed that an optimal parameter set up (e.g.. tonnage, punch speed) exists in terms of minimized dimensional variation. These parameters are much tighter than those to simply avoid split and wrinkle. Another advancement is the development of a punch tilt compensation system for a hybrid stamping press [Neugebauer, 1994]. The system can compensate for shut height errors due to various root causes and balance the stamping punch forces. Its kinematics behavior has been proven in both simulations and experiments. However, the development still focuses on the press mechanism and its controller alone, without integrating in-process sensing and sensor fusion. An intelligent, in-process, compensation system is needed to achieve high precision stamping.

Objective:

Develop an intelligent in-process monitoring, diagnosis and control system for stamping processes to control dimensional variation within sub-millimeter level during production.

Deliverables:

(1) Flexible, quantitative dimensional and surface measurement systems for sheet metal parts

(2) Stamping process signature analysis technique by integrating in-process, multiple sensors, advanced statistics, and pattern recognition to provide continuous process monitoring;

(2) Predictive die maintenance by integrating press, die, production information, signature analysis, die durability model and experience based system; and

(3) In-process error compensation by adjusting punch tilt, Nitrogen pressure, etc. based on the in-process sensing and signature analysis.

Research & Development Plan:

Task 3.1. Innovative Measurement Strategy. Measurement is the basis for the dimensional control. The current measurement strategy is not adequate for the purposes of efficient dimensional control of stamped parts. This task will focus on the dimensional and surface measurements of tools and parts. The scope of the work covers not only the development of a measurement gaging system and innovative fixturing ideas, but also a measurement strategy and specifications for dimensional and surface inspection. The detailed tasks include 1) evaluate Different Measurement Strategies for Process Control; 2) Define Specifications for Measurement Strategy of Choice; 3) Develop a High-Speed, Non-Contact Measurement System; and 4) Develop a Surface Quality Metric for Stamped Parts and an Associated High-Speed Non-Contact Measurement System. Some fundamental questions in measurement, such as where to measure, how to measure, how often to measure, and how to process data, will be investigated as shown in Figures 15 and 16.

240
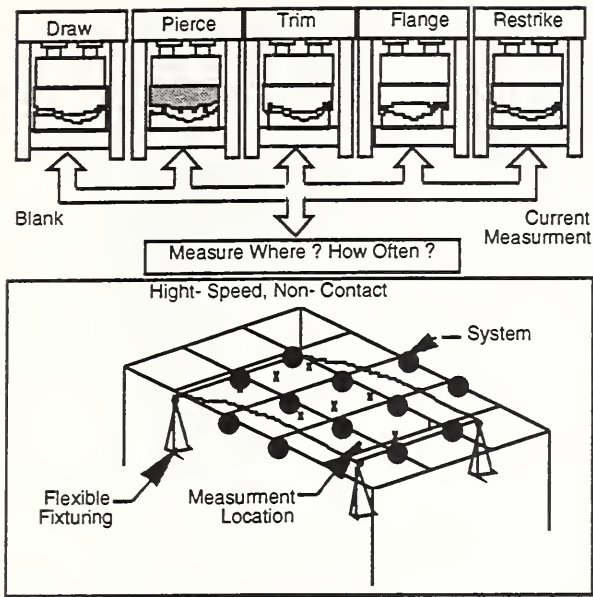
Fig 15. Non-Contact Measurement System.
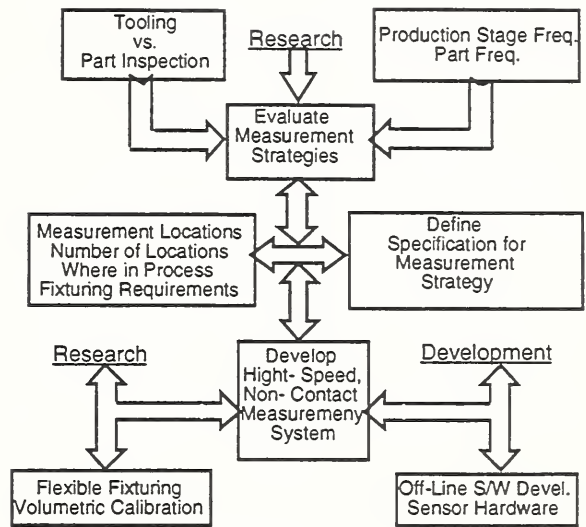
Fig 16. Task 3.1 R&D Plan.

Task 3.2. Signature analysis for sheet metal stamping. This task focuses on effectively using the information from in-process sensing (tonnage, strain gage, temperature, shut heights, vibration, etc.) for process monitoring and diagnosis. Sensor fusion or sensor synthesis techniques will be used to monitor multiple process measurements simultaneously, and integrate this information with knowledge about the stamping process, press, dies, etc. for process fault detection and diagnosis. This analysis scheme will (1) realize in-process quality assurance, (2) develop effective predictive maintenance plans, (3) monitor press and die working condition and faults, (3) achieve a repeatable die set up, thus eliminating run-to-run variation, and (4) provide information to integrate with a compensation system for in-process adjustment. Successful completion of the task will achieve intelligent stamping process control, thus significantly improving the understanding of the stamping process and approaching the goal of agile and precision stamping. The detailed tasks and principals are shown in Figures 17 and 18.
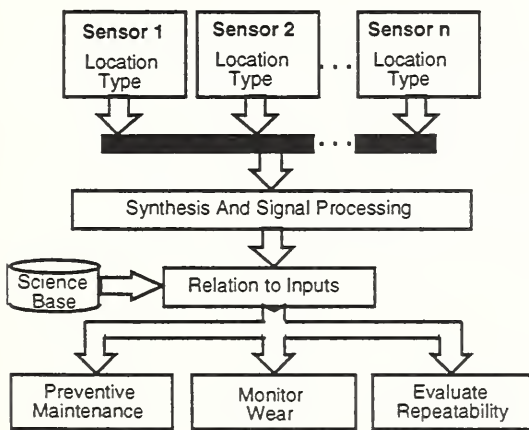
Fig 17. Stamping Process Monitoring and
Control System.
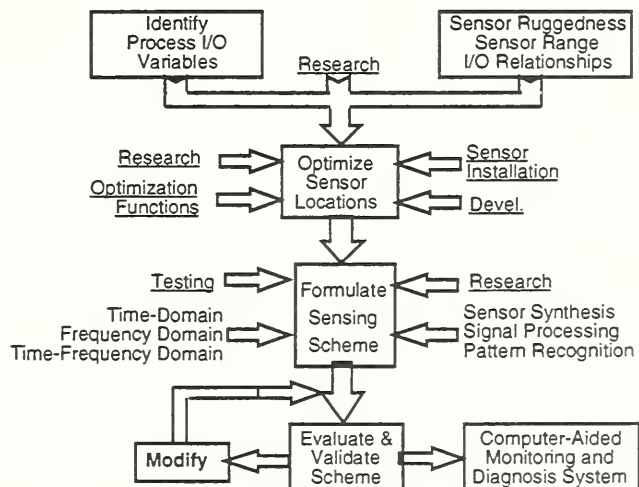
Fig 18. Task 3.2 R&D Plan.

241

Task 3.3. Multi-attribute Decision Supporting System for Die Predictive Maintenance.
Maintenance has a direct impact on the part quality and productivity. This task is proposed to achieve predictive die maintenance, rather than the reactive die maintenance used currently in production. In order to achieve this "predictive" objective, information from multiple sources: die, press, material, operation condition, maintenance, etc. will be considered in the decision making. Past experience and knowledge will be combined with statistical techniques and durability models to develop a computer-based system for PM decision support. Additionally, this system will help avoid unexpected early failure due to not pulling a die for PM since the system will be capable of modeling durability to account for the actual conditions rather than just the ideal design conditions. This challenging tasks include 1) Develop an Information Collection System and Information Base; 2) Formulate Durability Models; 3) Generate a Knowledge Base; and 4) Analysis, Prediction and Decision Making and Integrate the Complete System. Figures 19 & 20 indicate those tasks and their relationships.
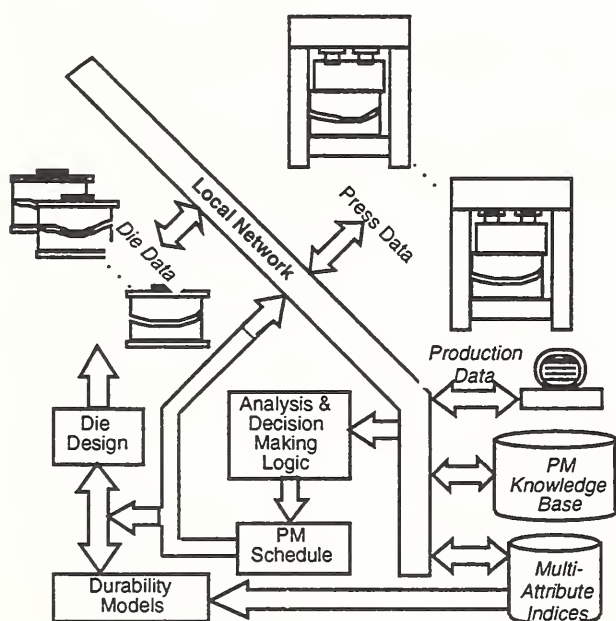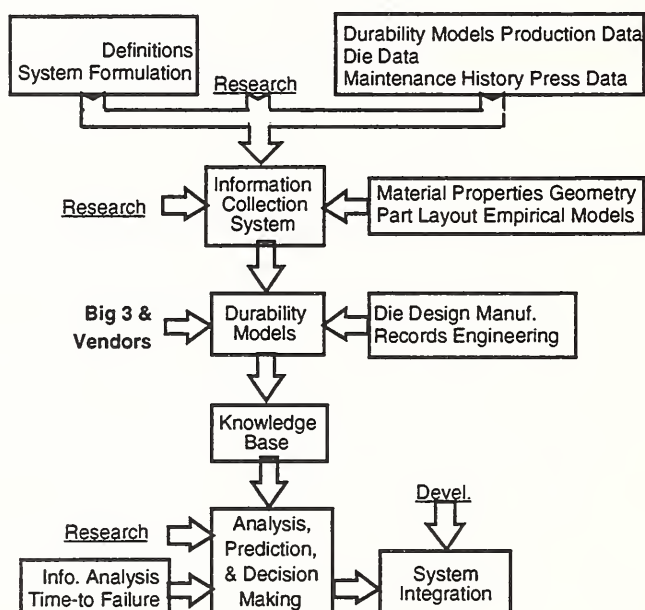
Fig 19. Die PM System.   *Fig 20. Task 3.3 R&D Plan.*

Task 3.4. In-process adjustment/compensation system. This task develops an innovative strategy to realize in-process adjustment capability for several key process variables. This capability essentially does not exist in current industrial stamping practice. The task relies on the in-process signature analysis results in conjunction with the off-line information (material properties, dimensional measurements, etc.). The in-process adjustment will not only avoid the conventional formability defects (split, wrinkle, surface distortion), but also achieve high precision with minimized dimensional variation. An in-process, batch adjustment method will be tested and implemented. The feasibility of realizing real-time adjustments on selected key parameters will be evaluated. In general, this task team will work closely with other NZS task teams, and integrates the results into a in-process compensation/control framework. The concept of the task is shown in Figures 21 & 22.
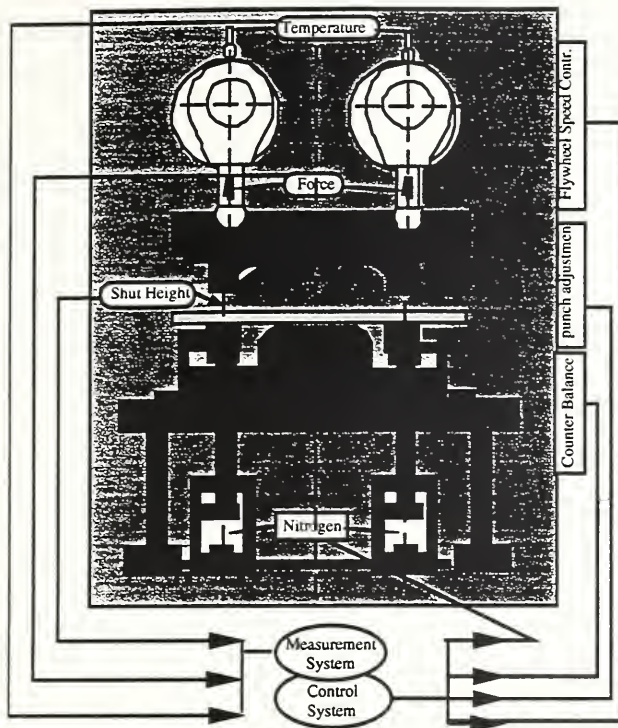
Fig 21 Examples of Potential in-process
sensing & adjustment variables
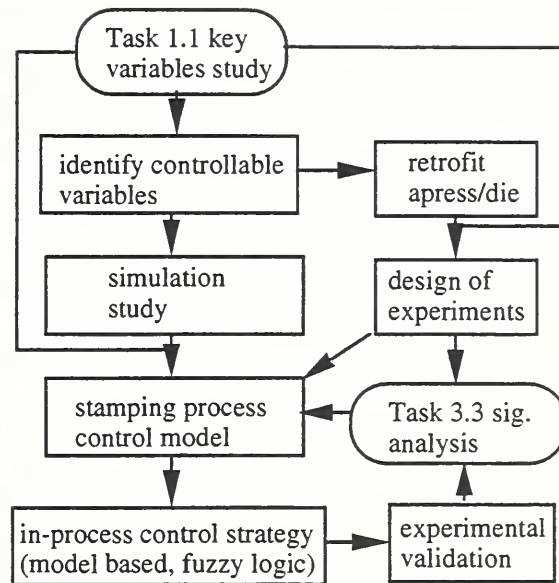
Fig 22 Proposed Task 3.4 R&D plan

3. NZS R&D PROGRESS SUMMARY

Significant progresses have been made by the joint venture companies and research team since the start of the NZS project one and a half years ago. In this section, a brief summary will be provided for each project to highlight the progress in the project. Due to the large scope of the project and the rapid development of the program within and between various task teams this summary can only provide a brief description of current R&D status. The detailed information can be found in quarterly technical reports submitted to NIST-ATP.

3.1. Project 1: Stamping Design

Stamping design includes product design and manufacturing process design, and the two should be closely related so that an optimal product function (such as stiffness, impact resistance, and strength to weight ratio etc.) and an optimal manufacturing process (both stamping and subassembly) can be fulfilled at the same time. Along this direction Task 1.1 is to deal with stamping design; Task 1.2 is to develop computer simulation tools for stamping design; Task 1.3 is to develop a fundamental understanding on stamping variables and their sensitivities to stamping formability and dimensional quality; and Task 1.4 is to develop understanding and design tools for hemming process.

Stamping Design (Task 1.1): In the past 6 quarters Task 1.1 has made great progress to develop an optimized product design methodology and architecture for sheet metal stampings and sub-assemblies that will consistently approach "near zero" dimensional variation. The achievements are listed as the following:

(1) Examined the design process through historical data and identified inhibitors to optimal performance (Sub-Task 1.1.1). The previous design practice and the lessons learned from it

243

are being collected and accumulated from Big-Three automakers. The design rules are going to be developed from both historical data and from the new understanding developed from this program.

(2) Identified the critical characteristics in sheet metal stamping (Sub-Task 1.1.2): During panel tryout and development the attributes and critical characteristics that relate to part design were identified from data analysis and discussion with key suppliers in the die and assembly plants. The function objective tree includes objectives that NZS projects are working to satisfy. Measured derived from function objective tree will be utilized as inputs in Quality Function Deployment (QFD) process to obtain the design, process and manufacturing requirements. Questions will now be included in above mentioned questionnaire in order to refine the function objective tree.

(3) Development of a Computer-Based Decision Support System (Sub-Task 1.1.4): A list of characteristics, requirements, and parameters, that NZS task teams were developed. The items on the list were classified into input, output or control. The structure of the stamping design evaluation process is described in Fig 23. The input, output or control elements of the design were classified, and were integrated into the Auto Body Product Realization Process model. In order to define the architecture of Decision Support System, (DSS), a survey questionnaire was developed.
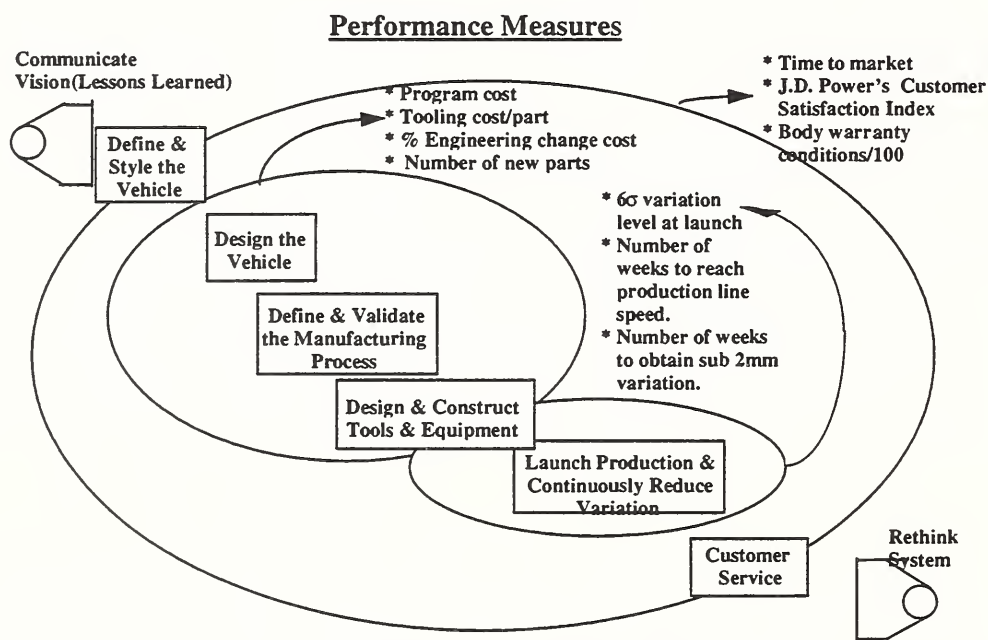
Fig 23. Generic process evaluate flow chart

In parallel to the design system development, a mechanistic modeling was conducted to analyze the subassembly process (Sub-Task 1.1.3 Quantitative Analysis of Current Subassembly Architectures). From the viewpoint of end product - Body-In-White this Sub-Task evaluates the stamping and subassembly process for achieving "Near - Zero" BIW variation. An improved variation simulation package based on mechanistic model is developed. It uses the in-coming stamping part variation and assembly fixture variation as the input for various subassembly architectures (i.e. the combination of series process and parallel process), and for various welding/joining process (lap join, butt join, etc.), and to obtain assembly variation (output). Thus, it can provide a feedback to stamping design.

As an example, for a new truck cab model, Figures 24 and 25 shows the results from the theoretical modeling on the sensitivity of assembly variation from the input stamping part variation and assembly tooling variation, for two classical joint techniques - lap join and butt join. The results indicate that, for lap joints, in Region 1 and 2 the part variation dominates the assembly variation, while in region 3 the tooling variation is dominating. For butt joints, in Region 1 and 2 the parts with heavier thickness has the dominant influence on the final assembly variation. As for Region 3, in 2D case the gun variation is dominating, while for 3D case both parts and weld gun variations have significant impact to assemblies.
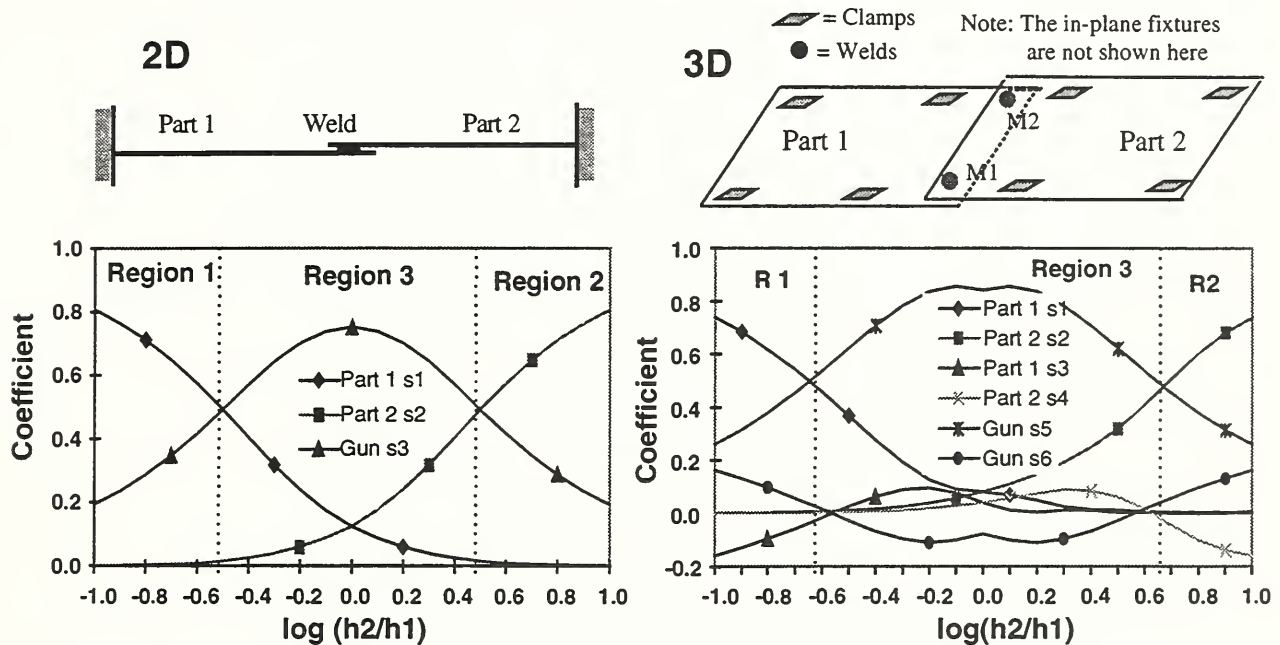
Fig 24. Sensitivity coefficients of sheet metal assembly with lap joints, where the sensitivity coefficients are:

s_1 (part 1) and s_2 (part 2) for the source variation at weld M_1 to the assembly variation at the same point M_1 (local sensitivity of the parts),

s_3 (part 1) and s_4 (part 2) for the source variation at M_2 to the assembly variation at M_1 ,(inter-sensitivity of the parts)

s_5 for the weld gun at M_1 and s_6 for we;d gun at M_2 , to the assembly variation at M_1 .(the intra- and inter-sensitivity of the gun).
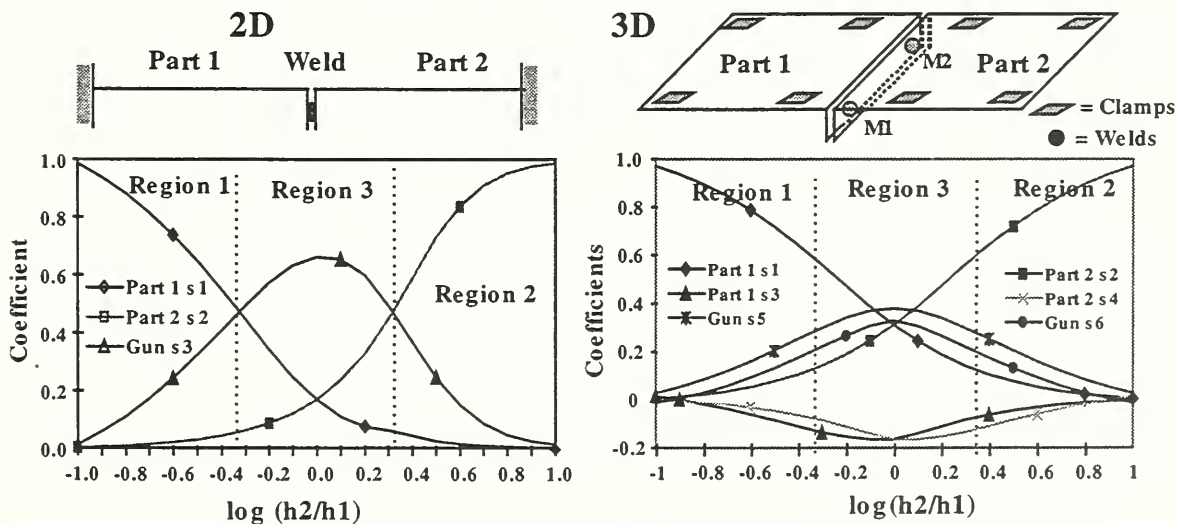
Fig 25. Sensitivity coefficients of sheet metal assembly with butt joints..

245

Computer Simulation: Computer simulation has been widely used in various tasks including stamping product-process design and subassembly design, stamping variable analysis, die development, measurement fixturing, material handling, and signature analysis. In Task 1.2 FEA tool was used extensively for process design, as seen the die development flow chart in Fig 26. FEA analysis has been used at various tool development stages and greatly reduces the tool development time and manufacturing cost. The draw processes of several major autobody panels have been simulated with different FEA packages including Pam-stamp and LS-Dyna3D, and one-step FEA package.
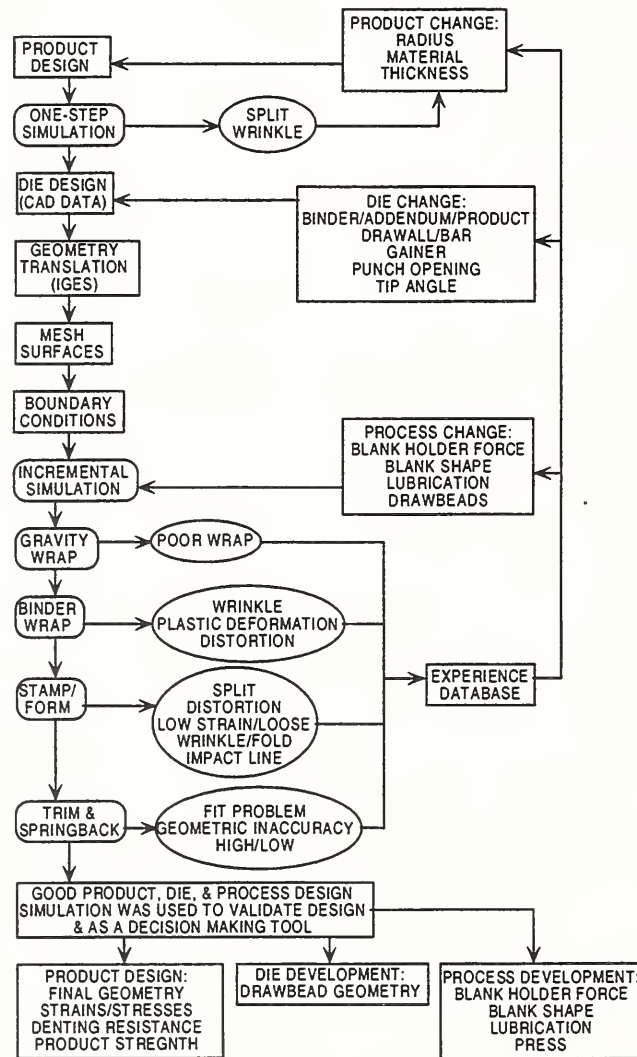
Fig 26. Application of FEA in die development (from Task 1.2 report by OSU)

Sensitivity of Critical Stamping Parameters: Using both experimental approach and science-based computer simulation the impact of stamping variables to stamping deformation and part quality have been studied extensively. The experiments were conducted in industrial stamping environment (mainly for easy-to-change parameters) as well as in laboratory (hard-to-change parameters). To fully utilize the experimental results an advanced DOE was developed which uses partial orthogonal matrix and separates the easy- and hard-to-change parameters based on statistical principle.

To fully understand the sensitivity of critical stamping parameter on stamping quality laboratory experimental study and analysis were carried on. Some results are show in Figs 27 and 28.

Figure 27(a) shows how main ram load, and normalized draw displacements are influenced by blank holding force (BHF), part depth (h), lubrication and speed. Ram load increases with increasing BHF. For shallow part depth (h = 10 mm), the dependence on BHF is weak. The dependence increases with increasing part depth, but the rise in ram load shows a decreasing slope with BHF. Lubrication has only a minor effect on ram load, improved lube shows a slight decrease in ram load. Figure 27(b) shows that when h is small, the normalized draw displacements have a weak dependence on BHF, but at larger h the decreasing trend of draw displacement with increasing BHF is clear. As expected, draw displacement is higher for larger part depth, but this effect diminishes with increasing BHF. Lubrication effect in this range of lubricants was found to be not critical. Figure 27(c) shows the effect of punch speed and lubrication on draw displacement. For small h these effects are small, but lubrication generally improves the draw displacement.
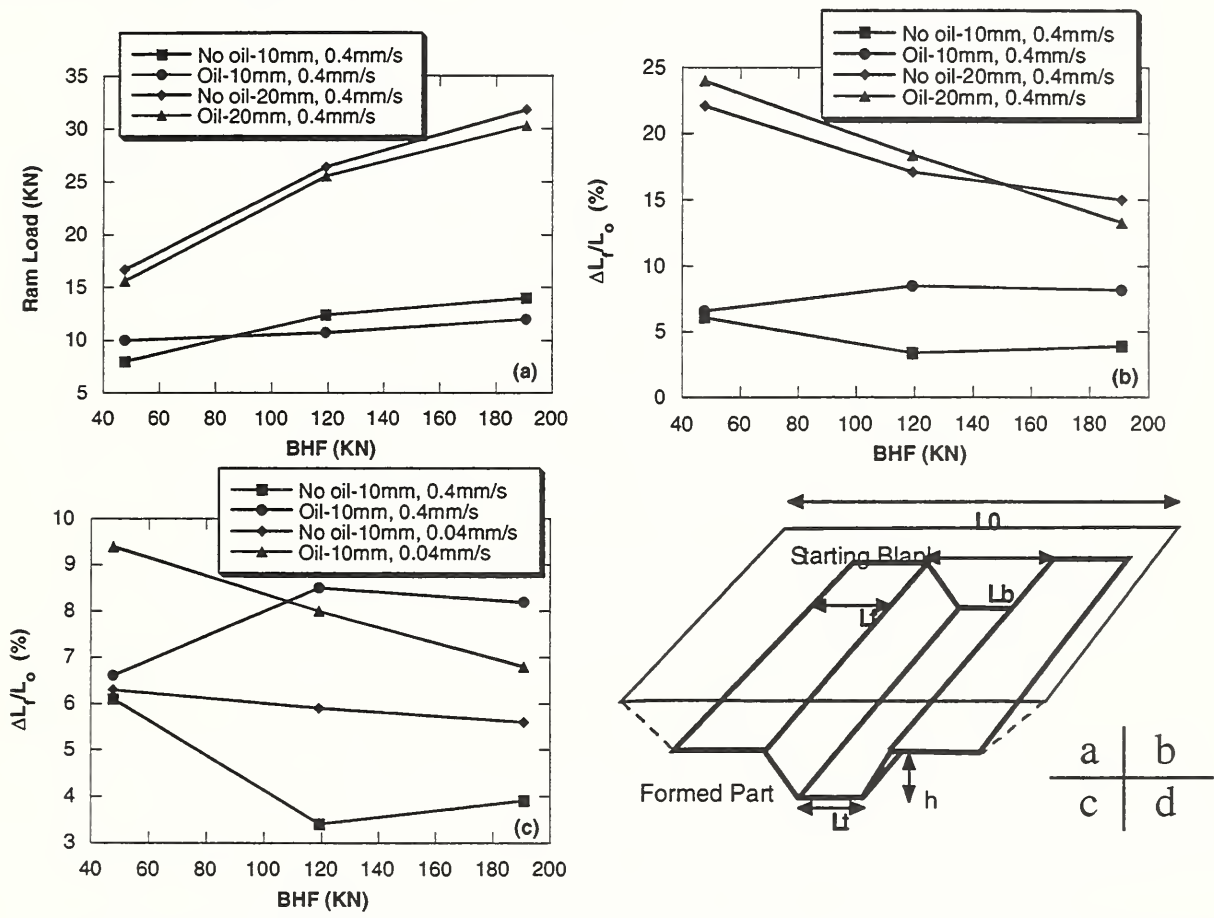
Fig 27. The effects of blank holding force (BHF), part depth, punch speed and lubrication on (a) ram load, (b) all draw displacements, (c) draw displacements for shallow parts. The plane strain forming process is show in (d)

Improved drawing displacements have the effect of minimizing stretching requirements on the part. The results on draw and stretch amounts are more clearly described in Figs.28. Draw efficiency, defined as the ratio of ΔL_f and h, shows a slight but steady decrease with increasing BHF for shallow part depths, but a rapid decrease for deeper parts, followed by a weak dependence above a moderate level of BHF. Mill oil lubrication on blank holding surface certainly increases draw efficiency for shallow parts. Fig. 28(b) shows similar results for deeper parts, where improved draw efficiency with draw speed is seen. Figure 28(c) shows that increasing BHF can increase %Stretch almost linearly for deeper parts. The BHF effect on

247

increasing %Stretch is weak for shallow parts. This suggests that the effects of n or biaxial n are also weaker for shallow parts.
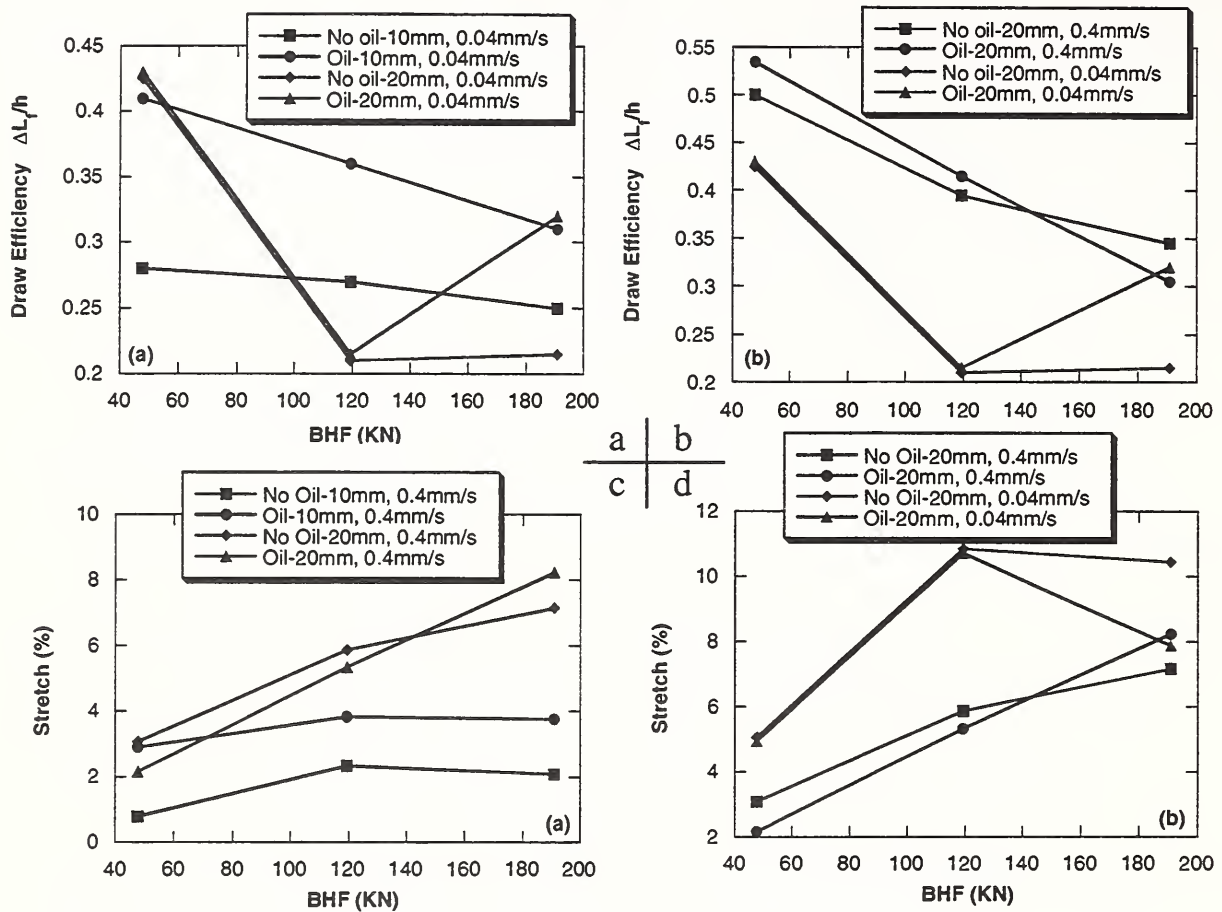
Fig 28. The effects of blank holding force (BHF), part depth, punch speed and lubrication on
(a) all draw efficiencies, (b) draw efficiencies for deep parts,
(b) and .(c) % Stretch at slow punch speeds, (d) % Stretch at all punch speeds, where
Drawing Efficiency = $\Delta Lf / h$, and $(\Delta Lf = Lo - Lb - 2 Lf)$
% Stretch = $Lt - \Delta Lf + 2$ Squrt $[0.25 (Lb - Lt)^2 + h^2] / Lb,$

Optimal Hemming Process Design: A complete hemming process and quality definition was developed, and current state-of-the-art review was conducted and completed based on the information from current industrial practice. It has been realized that hemming is a very critical process determining the final geometry of autobody closure panels (doors, hood, liftgate etc.), yet is a very complex process involving very localized deformation, spring back, and accumulated effect from flanging to pre-hemming and final hemming. At this stage the hemming process was focused on 2D straight hemming simulation and experimental validation. Three research institutions of this task teams used different FEA packages and the results were all reflecting the general trends of hemming process, though quantitatively there were some discrepancies among the different software packages, as shown in Fig 29.

2D modeling is going to continue for in-depth study. In next stage 3D modeling will be conducting for various combination of hemming geometry, and for various hemming stages including flanging, pre-hemming and final hemming.
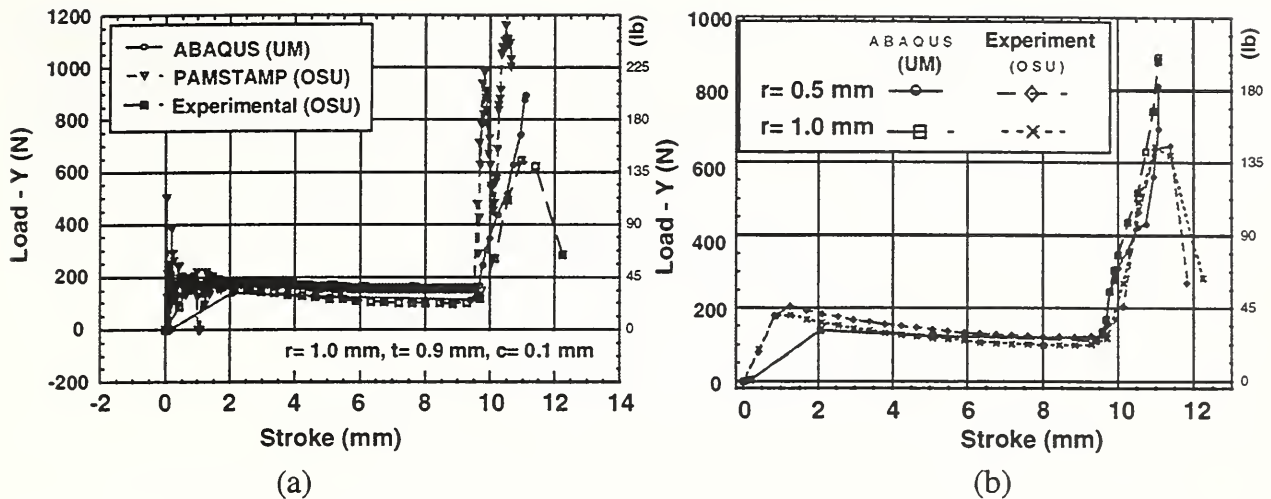
Fig 29. (a) Pre-hemming load; (b) final hemming load. From FEA simulations.

3.2. Project 2: Agile Stamping Tooling Development and Material Handling

Agile Die Tryout: For decades a long die tryout process exists from soft die tryout to hard die tryout and to production line secondary tryout, due to the lack of fully understanding of the complicated sheet metal flow and a science-based die development tool. The error in tool development has been realized from three aspects: (1) die machining and assembly error; (2) error from die misalignment and die deflection under stamping load; (3) design intent error due to lack of knowledge in sheet metal formability and spring back. Die machining capability and die deflection under load have been studied in a die manufacturing plant. For obtaining die deformation information during stamping loading tooling balls were mounted on the outer surface of die components, and the local die deformation is reflected through the tooling ball displacement, being measured by an optical system and portable contact probe, as show schematically in Fig 30. It was found that a significant difference exists between the stiffness of soft die and hard die. The results have received great attention on how to interpret the result from soft die surface geometry and tryout results and use them for hard die geometry.

To reduce this trial-and-error process a combined science-based approach (FEA) and experience-based approach (database from industrial practice) is used in this tryout study. The historical data as well as current tryout process were carefully collected and documented, and a computer tryout database software is in development, with the first version being tested-evaluated by three key die makers. This database allows various industry personnel to accumulate their results. It is easy to enter the current data and easy to retrieve previous data for reporting or for suggestions/aolutions to tryout problems. With the further development a tryout guideline or principle will be developed, and eventually a computer aided tryout system will be developed.

Global Strain Measurement System: For documentation of previous and current tryout process a quantitative description of all tryout procedure and results is needed. Combined global strain and geometry measurement tool is developed (Task 2.1). This system uses a vision based sensor equipped with a CCD camera and a laser structural light, and mounted in CMM machine for simultaneous measurement of surface geometry and surface strains (from pre-marked surface grids). Figure 31 shows the principle and structure of this vision-based system. Many hardware and software work (programming) are done to fabricate this integrated system. They include image processing, measurement path planning algorithm and programming, various calibration

249

steps, and the system integration. Now this system can be used in laboratory level, and will be further enhanced in its image processing speed and path planning algorithm for industrial application. This tool has the potential application for dimensional measurement of stamped part and die surface, and strain distribution for stamped part, and for validation of FEA results.
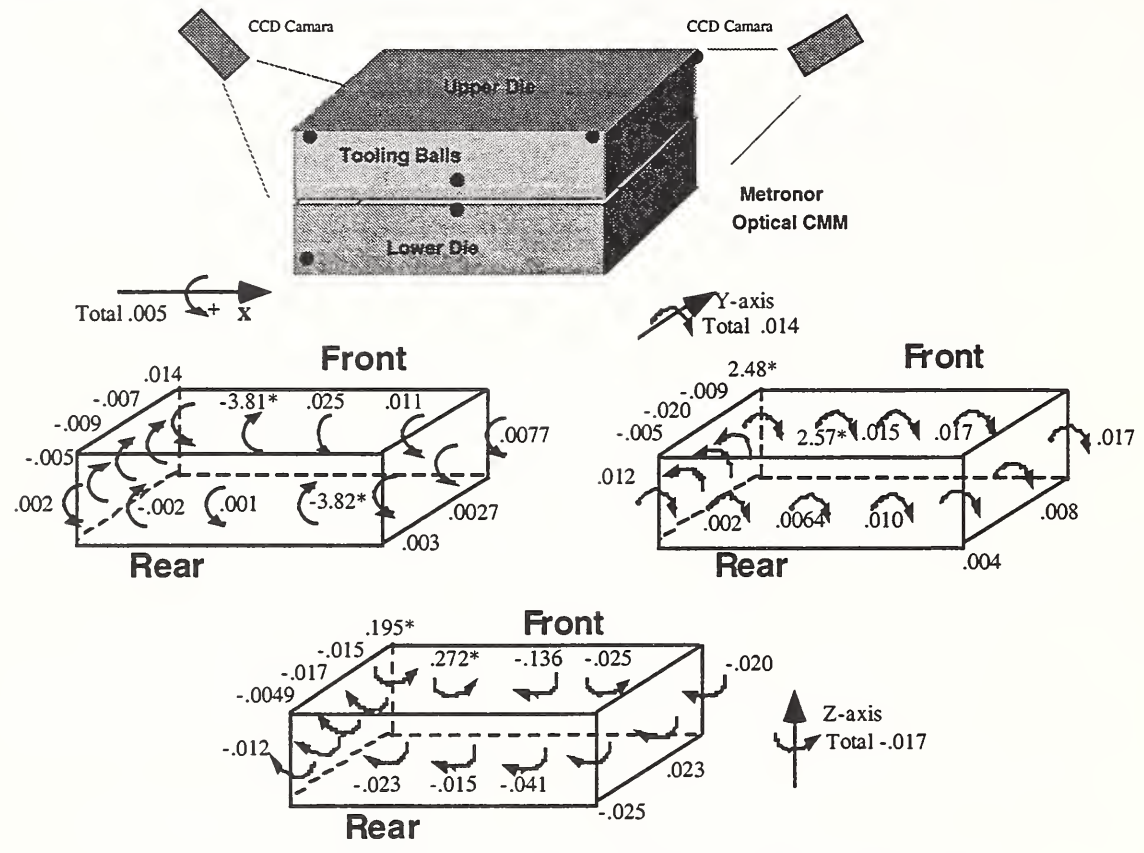
Fig 30. Tooling ball study on die deformation. Rotation around x, y, and z-axes were obtained from tooling ball position (in degree)
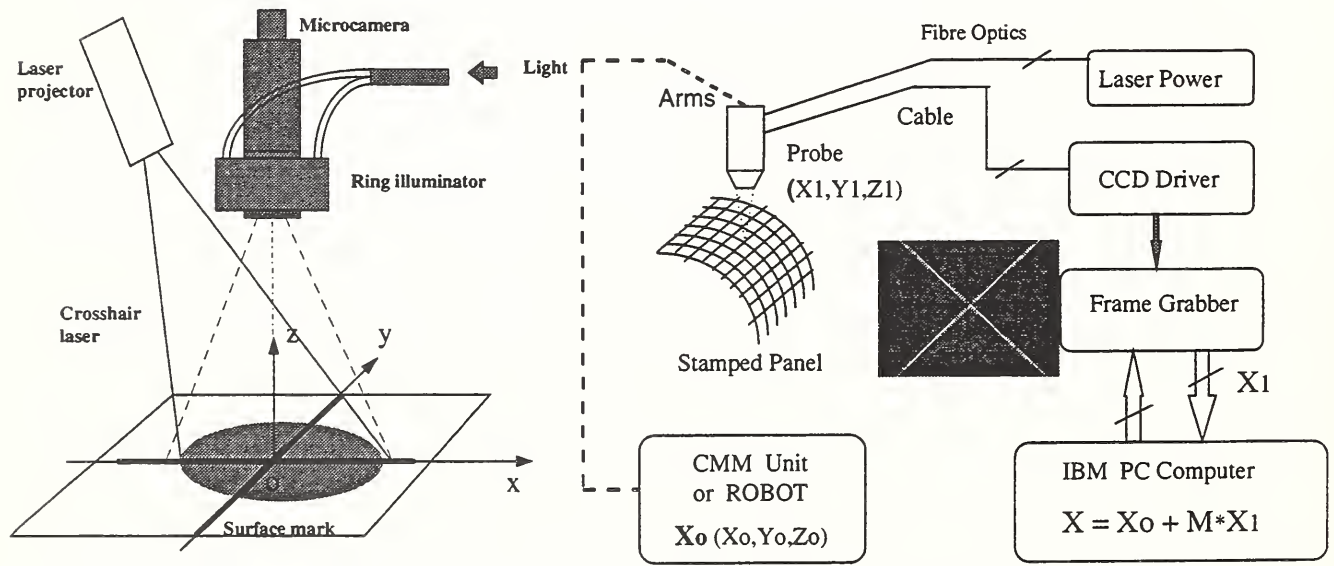
Fig 31. The configuration of the vision based geometry-strain measurement system

Material Handling: Research on material handling starts from the realization of its impact to sheet metal deformation caused by improper handling inside press (during stamping), between press (transfer), and between stamping line or between stamping to assembly (transportation and racking). To determine the optimal cup holding condition an FEA (ABAQUS implicit code) based automatic searching software is developed, which, based on the defined objective function (index) the software can allocate the optimal cup location. The panel gravity effect (dynamic or acceleration) and blank thickness effect are studied using this software program, and the results are shown in Fig 32.
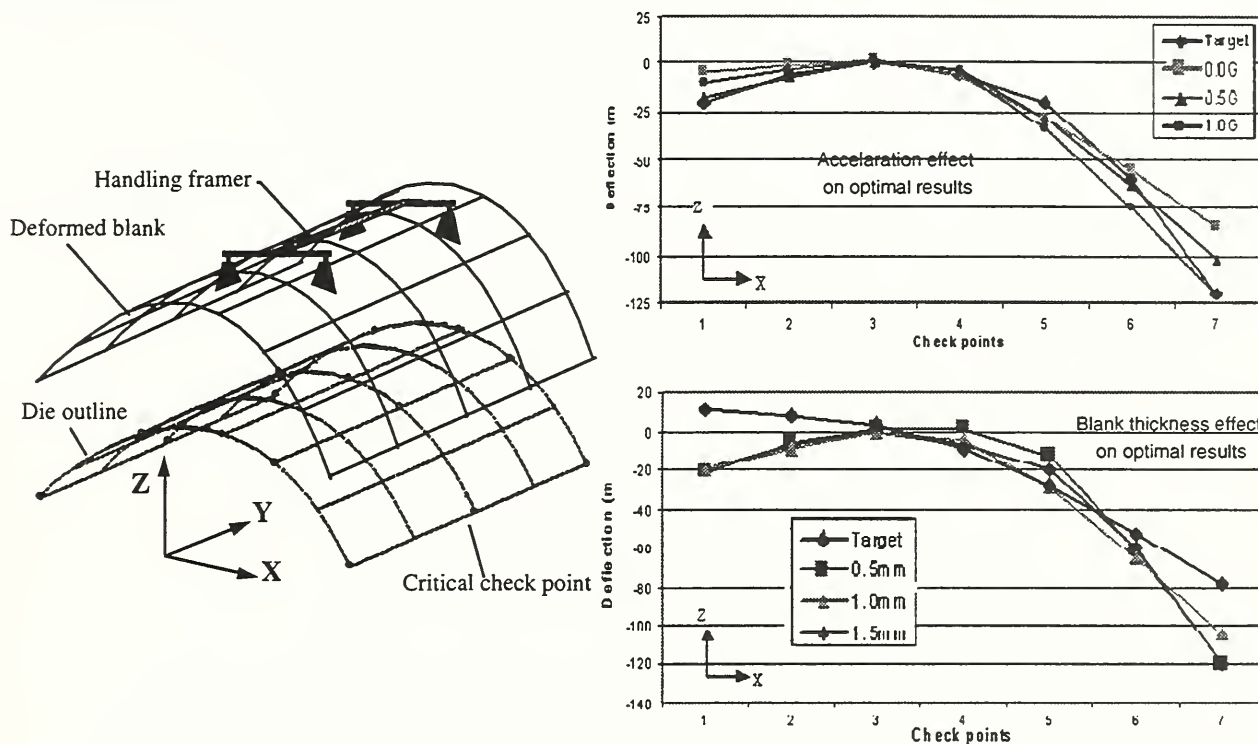
Fig 32. Examples of a blank handling problem within press, and FEA simulated optimal cup grabbing results for various acceleration and blank thickness conditions

3.3. Project 3: Stamping Process Monitoring And Control

Research on measurements of product (stamped part) is conducted in Task 3.1, and on monitoring production stamping process is conducted in Task 3.2. Extension of Task 3.2 leads to the preventive die-press maintenance (Task 3.3). The in-process control is studied in Task 3.4.

Stamping Measurement: Measurement of flexible stamping parts is not trivial. Over constrain fixture results in artifact of measured data, while 3-2-1 fxturing often fail in repeatability and reproducibility requirement (R&R test). A measurement strategy is in development to compromise the measurement speed and the number of measurement points for different stamping development phase, different characteristics of parts, and the suitable measurement tools-technologies. This task also provides a strong support for other tasks that need to have product dimension information as output or feedback.

The on-line production measurement on stamped parts is not available up to date. As an attempt to develop such a measurement technique a vision-based non-contact laser scanner was mounted

on a rigid robot, and if the system has good repeatability it will provide an opportunity for fast scanning the surface dimension without over constrain, to be used in production line.

Checking fixture is one of the areas being studied. With non-contact measurement technique the flexible stamped part can be located and supported in free state without over constrain. Note that the measurement system does not rely upon the accuracy of the physical location / support points. Instead, it uses a concept called "Visual Fixturing", which locates / measures a pre-determined number of features on a panel to establish its position and orientation. Subsequent measurements are then reported relative to the parts coordinate system. For improving the robot accuracy in production environment the thermal and geometrical error compensation algorithm is under development. The configuration of such a system is shown in Figures 33.
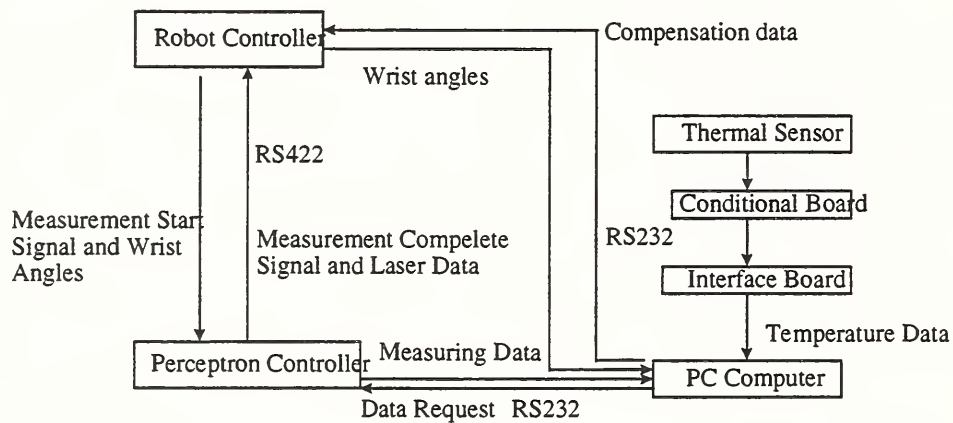
Fig 33. System Communication Diagram

Signature Analysis And Die Preventive Maintenance: Signature analysis is mainly used for production process monitoring, but in NZS program it has been extended to die tryout and preventive maintenance as well. Various press sensing techniques were studied, including tonnage signature from strain gage, shut height and ram travel displacement from LVDT or DC faster, temperature from thermal couple, thermal resistor and infrared thermal imaging, etc. Many of these techniques have been adopted by the member companies in this task. As an example, press thermal drifting was investigated for a high-speed progressive stamping process. During stamping production it was found that at the first two hours of operation the press experienced a warm up process: the temperatures at several locations were increased by several degrees, and the tonnage was increased accordingly. As a result for several times the peak tonnage reached a pre-set upper limit (safety value) and shut down the press operation. Such a correlated thermal effect is shown in Fig 34. Strong correlation between tonnage, temperature and displacement drifting indicated that the temperature was the root cause of tonnage over limit and down time.

To fully utilize the sensing technology and obtained process information for diagnosis, a database is needed. Since signature analysis is dealing with massive production data, instead of store raw production tonnage data, this database should store the key features for various signals which distinguishes different characteristics of stamping faults. These key features are being identified by design of experiment, principal component analysis, and other advanced statistics techniques within task members.

Temperature Rising

Shut Height Change

(a)

(b)

(c)

(d)
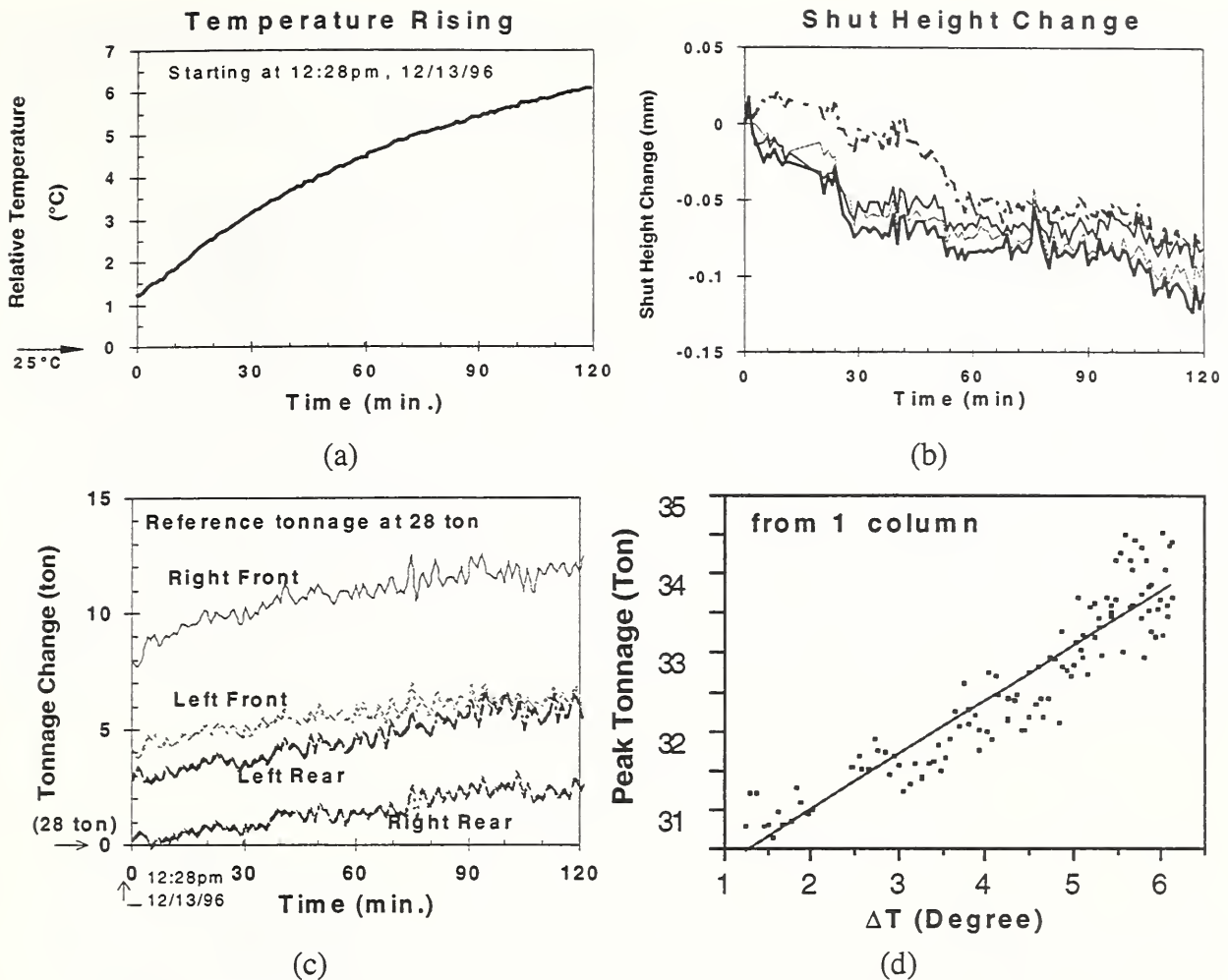
Fig 34. (a) temperature vs. operation time for one sensor; (b) ram displacement vs. time;
(c) tonnage change vs. Time; (d) correlation between peak tonnage and temperature changes

For maintain a normal production condition tooling maintenance is one of the major activities. For the best economical effect a preventive tool maintenance scheduling is critical. Task 3.3 team has collected die maintenance record from one stamping plant and conducted a detailed analysis. Different criteria were used to sort the top ten die faults, together with their occurrence frequency. The software of the maintenance database has been upgraded for ease data entry and retrieve/report. The analysis of the die service data has provided the focusing areas for preventive maintenance. Effort is made to develop a tooling damage model which will eventually lead to the optimization of die service activities.

In-Process Stamping Control: Task 3.4 team has developed a laboratory active binding force control system. With several independent cylinders and control loops this system can provide a pre-programmed binding force function or adaptive binding forces over different stroke position (or time) and over different locations. In addition, advanced sensing techniques are being used to provide a complete, accurate and reliable information of in-situ press condition and sheet metal forming condition. The tooling development for this study is under way. Two stamping presses, one for blanking-trimming study and another for drawing operation study, are installed in the new stamping laboratory at the University of Michigan (see Fig 35). The draw parts to be studied are shown in Fig 36. The part geometry simulates the two critical and difficult areas in most door inner panel draw operation: (A) window lower corner area; (B) door inner bottom

corner area. Both areas have a sharp change of draw depth. The experimental study can not only provide a means of study on process control, it can strongly support other tasks for experimental validation under well defined testing conditions. For control purpose it is needed to establish the relationship between input and output parameters. FEA is used for process simulation and for finding the correlation of input- output variables.
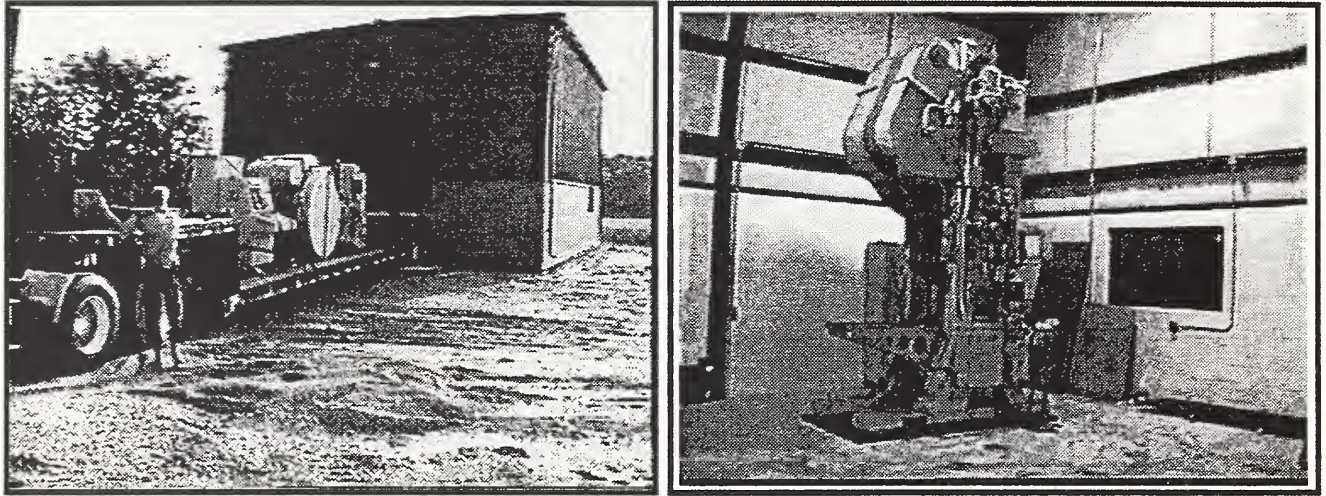
Fig 35. New stamping laboratory at the University of Michigan and a draw press(supported by Allen Bradley).

Designed Punch Head
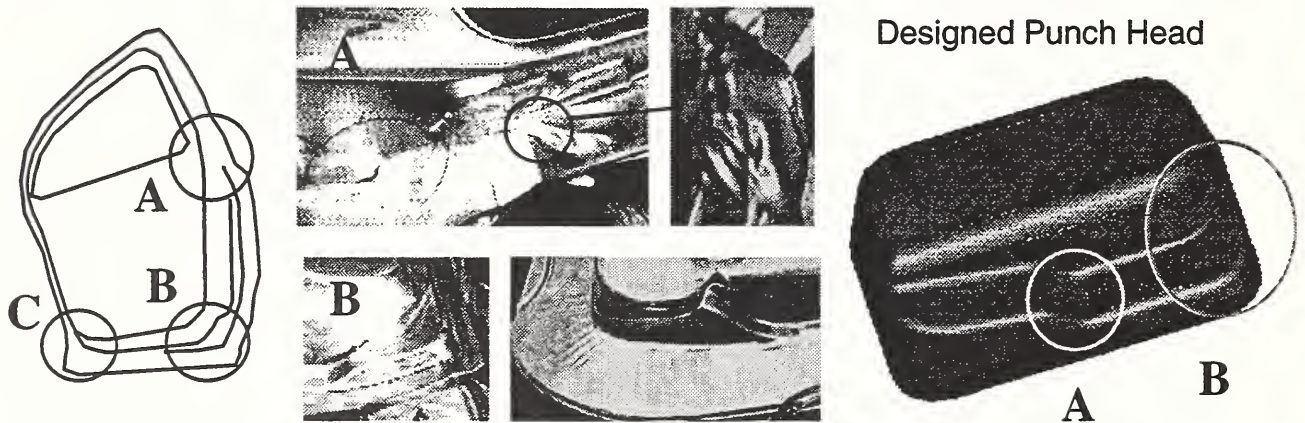
Fig 36. An example of tool design concept and designed punch head.

3.4. Project 4: System Integration

For NZS consortium it is very critical that all information can be shared among the member companies, and all research results can be integrated and commercialized or utilized by US auto industry. Internet with high security is being used to communicate between companies and researchers/engineers, and to share the knowledge gained from individual team members and tasks. Currently many technical developments show great potential for commercialization or significant knowledge/technology advancement. Following listed are some examples:

- Enhanced variation simulation software and mechanistic modeling tool kit (Task 1.1);
- Design-supporting and decision making system software for stamping design (Task 1.1);
- Understanding of stamping process and improved accuracy in computer FEA (Task 1.2);

- Understanding of critical stamping variables, and their impact/sensitively to dimensions (Task 1.3)
- Hemming design simulation tools/software and related knowledge (Task 1.4)
- Integrated vision-based systems for global strain measurement and dimensional measurement (Task 2.1);
- Software for image process, path planning, sensor calibration, results display (Task 2.1);
- Die tryout database and a computer aided tryout system (Task 2.2);
- Potentially elimination of soft die tryout process (Task 2.2)
- Location optimization and robot motion optimization software for material handling (Task 2.3)
- Kinematics simulation tools for press-end locator motion design (Task 2.3)
- A vision based laser non-contact scanner system installed on robot, for production quality measurement, to replace current checking fixture (Task 3.1)
- Sensing techniques for press monitoring (Task 3.2/3.3)
- Data reduction and key feature subtraction technique (Task 3.3)
- Signature analysis techniques for stamping production monitoring and diagnosis (Task 3.2)
- Die preventive maintenance model and software (Task 3.3)
- In-process control algorithm and control system (hardware) (Task 3.4)

4. SUMMARY

This article provides an overview of the Near Zero Stamping Program funded by the NIST MVMT/ATP program. Based on the first 1.5 years of program R&D practice great progress has been make in the advancement of all dimensions of stamping technology. Although due to the limitation of paper space only very limited results have been described in this article, the program development does demonstrate general trends and some unique features:

1) Concept of Stamping System: For the first time the stamping process has been studied at entire system level from product design, tool development all way to production and maintenance. It has been seen that with the participation of Big-Three automakers and key supplier companies the NZS program provides a unique opportunity to attach all critical stamping issues, which otherwise may not be able to realized by isolated research activities.

2) Competitor and R&D partners: For the first time the competing companies become R&D partners and work jointly to uncover common stamping problems. NZS organization contains 24 joint venture companies and 5 research institutions. Among them exist user-supplier relationship as well as competitor relationship, and at the beginning of the program there was a big concern on the possibility of collaboration between competing companies or universities. The past 6-quarter practice indicates that only collective effort can provide solutions to fundamental stamping problems and best benefit to each member; in this sense NZS program provides the vision to a unique opportunity for technical merging and integration. In order to achieve NZS objectives collective efforts have been made within entire consortium. The success of such large scope collaboration has demonstrated that by correctly setting the legal regulation on the intellectual property this barrier can be successfully removed.

3) Closed loop between stamping industry and university research: Since each technical task of NZS program is driven by industry members, it provides a guidance to research direction and a great funding opportunity for participating universities and national laboratory. With great enthusiasm scientists are working closely with industry engineers and jointly define-solve stamping practical problems. On the other hand with the rapid technology transfer from research to application the industry members can immediately share research results and knowledge generated from entire consortium. As can be seen even in the middle of the program many new technical developments have shown the sign of potential commercialization. Here government ATP fund pays a critical role to make this possible.

4) Bridge between art and scientific: The modern technology has been applied to classical stamping manufacturing area, e.g. large scale computer CAD/CAM and simulation, laser technology, precision engineering, advanced sensing technology and control techniques etc. Stamping is no more an isolated shop practice but has been realized as a very importance research area. Many trial-and-error processes, such as die tryout often seen in die shops, are now being simulated first in front of computer screen before start expensive real manufacturing and rework.

With the spirit of this collective effort within NZS team members it is no doubt that the momentum developed within the first period of program will be continued and further stimulated. We expect to see an accelerated progress and a revolutionary stamping technology advance to take place.

REFERENCES

1. Aritzur, B, Metal forming: Processes and Analysis, McGraw-Hill Book Company, 1968.
2. Buckley, D. H., "Surface Effects in Adhesion, Friction, Wear and Lubrication," 1981, Elsevier.
3. Cao, J. and M.C. Boyce, "Draw Bead Penetration as a Control Element of Material Flow", SAE Sheet Metal and Stamping Symposium, Detroit, MI, March 1993, SAE Paper No. 930517.
4. Ceglarek, D. and J. Shi, "Dimensional Variation Reduction for Automotive Body Assembly Manufacturing", Journal of Manufacturing Review, Vol. 8, pp139-154. June, 1995.
5. Ghosh, A. (to be published).
6. Jin, J and J. Shi, "Signature Analysis and Fault Classification Using Neural Network", NSF-I/U-CRC progress report, Jan 1996.
7. Keeler, S. P., Workshop presentation "New Analysis of Sheer Metal Formability" at Chrysler, May 1995.
8. Koh, H., J. Shi and W. Williams, "Tonnage Signature Analysis Using the Orthogonal (Harr) Transforms", C. K. NAMRI/SME Transactions, Vol. 23, pp229-234, 1995.
9. Koh, C. K. H., J. Shi, and J. Black, "Tonnage Signature Attribute Analysis for Stamping Process", ", accepted by NAMRI/SME Transactions, Vol. 23, 1996. (in-press).
10. Michler, J. R. and K. J. Weinmann, "Variation of Friction in a Strip Test Apparatus with Controllable Drawbead Penetration," SAE paper No. 950698, p65-73.
11. Michler J.R., A. R. Kashani, S. A. Majlessi, and K. J. Weinmann, "Feedback Control of Sheet Metal Strip Drawing, By Drawbead Penetration Adjustment", PED-Vol.64, Manufacturing Sci. and Eng., ASME 1993, p181-187.
12. Pashley, M. D. and D. Tabor, "Adhesion and Deformation Properties of Clean and Characterized Metal Micro-Contacts," Vacuum 31, (1981) 619-623.
13. Ustruck R., X. Wu, J. Zhang, et al, NIST-ATP "2mm program" Project 4 Quarterly Report No 10, July 1995.
14. Wagonar, R. et al "Recent Advances in Formability and Friction Studies," 1994 Int'l Conf. on Sheet Metal Forming Technology, Oct 3-5, 1994.
15. Wu, X., Proposal on Global Strain-Based Approach in Stamping Monitoring, NSF-I/U-CRC, 1995.

The Status of the Springback Predictability ATP Project

Daniel J. VandenBossche
Advance Manufacturing Engineering
Chrysler Corp.

INTRODUCTION

Springback is a major problem in the forming and stamping of metals, and in particular, the high strength steels and sheet aluminum. Therefore, springback predictability is very important to the producers of formed metal parts and to the suppliers of aluminum and steel used for sheet metal stamping. The springback problem must be solved to enhance the use of these materials in manufacturing applications. Suppliers such as Alcoa and U.S. Steel need to help solve the springback problem associated with sheet metal forming in order to maintain and grow their market share in the face of global competition and the future threat that plastic and composite materials will impact the market for aluminum and steel. Manufacturers such as The Budd Company, Chrysler, Ford, and General Motors also face global competition, high development costs, and high new product development/introduction cycle time. The solution of the springback problem will significantly reduce stamping die development costs, reduce the product development cycle time line, and enhance global competitiveness. Therefore, material suppliers and the U.S. automotive companies have agreed to form a joint venture to pool together large technical resources, including government and industry, to focus on the springback problem and to develop technology that will predict springback through modeling and simulation early in the die development process. Thus, Alcoa, The Budd Company, Chrysler, Ford, General Motors, and U.S. Steel are fully committed to help the sheet metal stamping industries resolve manufacturing problems inherent in the use of aluminum and steel. The cooperative Joint Venture allows the integration of the necessary technical resources to solve the springback problem.

Springback is a phenomenon that occurs in stamping flat sheet metal blanks into useful parts. It is a major problem in the design and fabrication of stamping dies used in sheet metal forming. Die tryout and redesign is a costly and time consuming step in the design of the dies that are critical to high quality metal stamping. It is in this area of the manufacturing process that springback and its deleterious effects stands out as a major problem for automotive body-parts manufacturers. The springback problem is further magnified with the use of high strength steel and light weight aluminum alloys. It is also a prevalent problem in sheet metal forming for a host of other products. Thus, the springback problem is a reasonable starting point in sheet metal forming research that will have great impact on U.S. global manufacturing competitiveness.

Technical teams associated with the Partnership for the Next Generation Vehicle (PNGV) have also identified springback associated with sheet metal stamping as a key problem area associated with the die tryout process in new vehicle manufacturing. Since PNGV's formation in 1993, technical teams comprised of government and industry technical experts have analyzed technical requirements

and developed research plans directed toward achieving the partnership goals. The PNGV technical teams recently selected sixteen projects from a list of 256 candidate projects that should be initiated to develop specific technologies that will advance progress toward meeting the objectives of PNGV. These technologies are basic to the design of common, non-competitive automotive products that, once translated into new products will provide the American automotive industries common parts for a wide range of vehicles. The end result, cost savings through economies of scale, will lead to enhanced competitiveness in the global economy. One of the sixteen projects selected is improved analysis of prediction methods, including springback prediction modeling for part dimensionally versus die design. The manufacturing of dimensionally precise stampings with weight-saving sheet material, such as thinner gauged high-strength steel and aluminum alloys will be critical to the new generation of automotive vehicles.

Currently, the process of automotive die development and construction requires a die tryout step in which the die is reworked to compensate for springback after a sheet metal stamping operation. The die tryout process is extensive, costly, and very stressful because the dies are urgently needed to initiate the production of new vehicles.

In recent years, advancements in computer technology, computation speed, and computer modeling and simulation associated with sheet metal forming have occurred and show significant promise to predict springback in a simulation mode. Springback prediction via computer modeling and simulation will enable construction of dies that compensate for springback, thus substantially reducing the costly die tryout process.

A concerted effort in material modeling to develop a more powerful constitutive theory of material behavior and development of a unified friction model is required to further advance progress in computer modeling and simulation in sheet metal forming. When proprietary and commercial code such as LS-NIKE3D and LS-DYNA3D are modified to interface with software modules to be developed during the Springback Predictability Project (SPP), the new FEM computer code will be able to quantify the parameters required for precision springback prediction. The new software modules to be developed in the SPP include robust solution control, variable formulation shell elements, material and friction models, and advanced equation solvers.

The SPP program is divided into six primary task groups: material modeling, friction modeling, FEM code development, laboratory validation, field validation, and commercialization. In the following, the status of each group of tasks is outlined and in addition the activities at NIST in support of SPP are summarized.

MATERIAL MODELING

The intent of the Material Modeling group is to provide advanced material data and constitutive models which will enable more accurate prediction of sheet metal response during the forming operation as well as in the subsequent springback. To meet this goal, we must gather more complete and more accurate information on the five grades of steel and the aluminum alloy used in the Springback Predictability Project than that commonly used in present-day formability analysis. This

is because the calculation of springback requires an accurate determination of stress throughout the formed part. This, in turn, will demand accurate data on plastic strain- hardening, anisotropy, and behavior during material unloading. The following describes briefly how we approach this task.

The work is being carried out in two phases. First, we establish baseline material properties typical of those used in forming analysis today for all the materials in this project. This data will be used in the initial modeling efforts of the Laboratory and Field Validation Groups. Participants in this phase include Alcoa, U. S. Steel, the University of Michigan and Ohio State University. This phase is complete, and the results have been provided to the Steering Committee. They include Young's Modulus, Poisson's ratio, true stress versus plastic strain data, yield strength, ultimate tensile strength, average coefficient of anisotropy, uniform elongation and total elongation. For consistency, all of the data reported to the Steering Committee came from Alcoa, the other labs serving as independent checks. Agreement among labs was very good.

The second phase represents the main contribution of this group. In it, we will provide more detailed data and more advanced constitutive models for the final version of the code. Participating labs are Alcoa, the University of Michigan and Ohio State. Alcoa has measured detailed data on the in-plane anisotropy of the materials. This includes uniaxial tension data at fifteen degree intervals from rolling to transverse directions in the sheet. Balanced biaxial tension test data and shear response were included, as was limited strain-rate information. Fully anisotropic yield conditions have been constructed. Mathematical models of hardening data have been found and compared to experiment. Although this information does not comprise new research, it establishes a comprehensive description of the materials, much superior to anything previously available, and will provide a framework for the work described next.

The University of Michigan will examine strain-hardening behavior and elastic unloading under three different stress states. First, very careful measurements are made of elastic properties (using strain gages) and plastic strain hardening (using extensometers) in uniaxial tension. Standard mathematical models are not capable of describing the hardening data, and new ones are constructed. In addition, elastic properties during unloading from various strain levels have been determined, and it is becoming clear that they depend on the level of plastic strain. For example, the average Young's Modulus decreases by 4.7% in going from zero to 16% strain. In addition, strain recovery upon unloading is nonlinear. The modulus continuously decreases as the stress tends to zero. This phenomenon directly affects the amount of springback and is not addressed in any standard material models. These measurements will also be made in balanced biaxial tension and in plane strain. Apparatus for these stress conditions have been built and work proceeds on hardening behavior and elastic unloading.

Ohio State University has performed tensile tests and has provided mathematical models describing their results. In addition, they have carried out careful low-strain (up to one per cent strain) tests in several sheet directions. Low strain regions tend to produce large amounts of springback, so low strain behavior is important. They have also designed and built an innovative three-point bend test apparatus which can be used in a bend/unload/reverse bend manner. This may reveal non-isotropic (Bauschinger) hardening effects, which could influence sheet response in material which is bent and unbent as it slides over small radii. They have also used an existing bend/stretch test at the Colorado

School of Mines, which can simulate springback at small radii as well as sidewall curling. Modeling efforts and data reduction for this test are now in progress.

When all of the data from physical tests has been generated, the main task of the Material Modeling Group will be to gather and rationalize the information. We will then need to put it into a form useful for the computer program. This will require construction of a constitutive model which can be framed as an algorithm for use in the code. Results will then be compared to experiments, and any final tuning of the model will be done.

FRICTION MODELING

In the final springback prediction, the accurate representation of friction is essential to obtain accurate predictions of forming stresses. It is the relaxation of the forming stresses that causes springback. The friction between sheet and die in sheet forming operations has a major influence on the stresses that develop, but actual values are not known. The ability to predict workpiece springback and hence the final shape of a formed sheet, is limited by uncertainty in friction forces. There is general agreement that the friction coefficient is not constant at all locations over the die face, but rather varies in response to local conditions. The actual friction force at any contact point will depend on factors such as the contact pressure, forming speeds, deformation history, lubrication, and the properties and surface topography of the sheet and die materials.

This milestone will be successfully completed when interface variables have been defined as delineated below:

1. For Each Sheet Surface and Tool Contact Point:
    - Local contact pressure
    - Relative sliding speed on tooling
    - Sheet strain rate
    - Tooling roughness
    - Boundary film characteristics
    - Storage of values at each step

2. For Lubricant
    - Local pressure
    - Local pressure gradients

3. For the System
    - Position of edge of contact of sheet with tool
    - Sheet and tooling relative tangential velocities
    - Sheet and tooling curvatures just outside of contact
    - Rates of change of curvatures
    - Lubricant distribution outside of contact
    - Temperatures of sheet and tooling

The approach being used in friction modeling considers two distinct but related categories: the Empirical Coefficient and the Stress Model approach.

The Empirical Coefficient approach is to measure forming forces and back calculate the friction coefficients that would explain the forces measured. Then, establish tables or curve fitted equations of the data. The instantaneous forming conditions encountered during a finite element simulation would be used to "look up" the appropriate friction coefficient and the calculations would then continue. This approach requires a lot of testing and tests must be repeated whenever a new situation is introduced. A new situation could be a new lubricant, sheet metal coating, or die surface treatment.

The Stress Model approach would depend on a theory of how the frictional stress would be generated using first principles and more generic information. For such an approach to work, there must be a model to explain how the surface topography of the forming sheet changes with straining and rubbing across the die face. It must also know the relationships between characteristics of the lubricant, such as viscosity, temperature, surface tension and pressure sensitivity and the lubricant performance. It is believed that once such a stress model was developed, the amount of testing would be reduced dramatically for each new situation.

The approach to be used in friction modeling will be to start on both Empirical Coefficient testing and the Stress Model approach immediately. There will be no redundancy since the development of the Stress Model will depend on the same data as the Empirical Coefficients.

Work during the first year on the empirical coefficients would be developed using one or more of the existing test methods. The coefficients would be developed for the conditions to be evaluated in the first "field" validation trial.

FEM CODE DEVELOPMENT

To predict springback which is controlled by the stresses at the end of the forming operation, the focus of the prediction accuracy must be directed toward the stress distribution in the entire stamping. For an accurate prediction of stresses improved material and friction models are required as well as better element formulation and enhanced numerical computation schemes. New approaches to solid elements, shell elements, and solid-shell interface are needed to model different tool radius-to-sheet thickness ratio and improved accuracy, robustness, and lower cost of numerical processing as discussed below:

Accuracy - New material and friction models are required for integration into the new FEM code to improve the accuracy for prediction of springback. The new code will use solid elements in regions of large curvature and the usual "plane stress" type of shell elements in regions of small curvature. A key development will be new transition elements that smoothly blend all aspects of the new FEM code.

Robustness - Improved accuracy and robustness will be realized by development of a predominantly implicit code with an option of switching seamlessly (forming stage

development) to explicit code with iteration for equilibrium.

Cost - One of the main barriers preventing the use of forming simulation on a broad scale is the high computing cost. A new solution solver based on iterative methods is expected to offer more than a five-fold speed-up compared to existing solver code.

LAB VALIDATION

Prior to the introduction of more complex and more expensive automotive stampings, laboratory validation of the development of 3-D finite element code must be conducted. It is necessary to build tooling and conduct experiments to evaluate the effect of material, design and process variables on shape distortion. Various part shapes (single, multiple, and plane strain bends, 3D shapes with compound curvature, corner radii, drawbead geometries, flat binders), materials (low carbon, dual phase, high strength steel and aluminum alloys) and different binder forces are required to provide the appropriate baseline for laboratory validation. The objective of the Lab Validation Task group is to fully integrate with the material, friction, and FEM code development work groups in developing baseline information for preliminary code evaluation and the large scale field validation with industrial stampings. Laboratory experiments are designed to physically simulate various forming operations such as bending, stretching, drawing as well as the complex forming of a rectangular pan. The tooling is designed, fabricated, and tested for part shape and geometries as established by all working groups on the SPP technical team.

The key deliverables of the Lab Validation Task are:

1. Complete build and tryout of complex channel.
2. Conduct testing and measurements for complex channel.
3. Complete testing and measurements for stretch/shrink flange and rectangular pan.
4. Complete baseline simulation of flanging operations, rectangular pan and complex channel.
5. Select test cases for simulation using new code.

The Lab Validation task group has made substantial progress toward the set objectives of the SPP project and the accomplishments can be summarized below:

Baseline Simulation:

- Material data has been incorporated into LS-DYNA3D and LS-NIKE3D codes. The 12 mm shrink flange die with the 1000 mm and 1500 mm plan view radii was selected as the common geometry for the analysis. The 0.75 mm thick DQSK steel was chosen as the common alloy for the baseline simulation.

- A number of runs have been make to evaluate the effects of speed, friction and interface damping. Friction and interface damping appear to have no effect on the results. Analyses indicated that predicted results appear to be dependent on the hardware and software version used in the simulation. John Hallquist indicated that due to the differences in architecture of

each machine, the precision, and the floating point being used for the calculation, the predicted results may vary somewhat. The SPP team also recognized that a 30% error in the experimental data can be observed in the measurements. Therefore, as long as the variation is in the order of the material thickness, the predicted results should be within a statistically acceptable limit.

Experimental:

- Alcoa Rectangular pans:
  - FUCHS DB4251 lubricant was applied to blanks using paint roller to establish a consistent coating weight of about 2 g/sq.ft. After water evaporation, the weight of the solid is about 200 mg/sq.ft. just prior to forming.
  - All of the pans have been formed and measured using CMM. The experiment indicated that in some instances high strength steel has about the same amount of variations in springback as aluminum alloy.

- GM Stretch and Shrink Flanges:
  - Testing of surface distortion for the stretch and shrink flanges experiment has been completed and simulation of the experiment is currently in progress. Springback angle measurements were made.

- Ford Straight and Curved Flanges:
  - All straight and stretch flanges have been tested and springback angles were measured.

- Budd Channel Die:
  - The channel die designed and fabricated by the Budd Company is used to generate additional experimental data for correlating with FEM simulation as part of the lab validation tests. The die was designed to accommodate a wide range of process, tooling, and material variations. The tool has been tested with different alloy sheets including aluminum and steels of various metal gauges. The parts produced has a wide range of springback conditions, from a few millimeters to over 20mm. A more comprehensive testing matrix is in progress.

FIELD VALIDATION

The approach to Field Validation is to assess the springback prediction technology by modifying production intent dies through simulation. This task will establish an initial benchmark based on the current technology and a new benchmark at the completion of the project.

Field Validation consists of four phases. Phase I will begin with the selection of several parts to be included in the validation testing. The strategy is to cover as broad as possible a range of shape/material/process combinations. The selected materials and die processes will be communicated to other task teams to start their coordinated research activities.

The objective of Phase II is to establish the initial benchmark by using current springback simulation capabilities. In this phase, several production intent dies will be constructed following the standard industrial procedure in die making. the major point of departure is that prior to die construction, the die shapes have been mathematically modified for springback compensation with the use of existing codes. In order to simulate closely the actual production environment, the dies will be made from production die material rather than soft tool. Once the dies are complete, draw or form panels will be produced in a fully instrumented press where the process conditions (such as binder tonnage and pad pressure) can be accurately controlled. The differences between the design intent and the actually stamped shapes as measured by laser scanning or equivalent constitute the benchmark of the current springback simulation technology.

The Phase III task will proceed immediately after the FEM Code team has completed their first release of the finite element code. With the improved prediction features in the code (e.g., new elements, new solver, enhanced material and friction inputs), the shapes of the existing dies will be re-modified. Comparison of the stamped shapes from the re-modified dies with the design intent defines a new benchmark of the springback simulation technology.

The task for Phase IV is identical to Phase III with the only exception that the second release of the finite element code form Task C will be used. In that release, new constitutive and friction models were already implemented.

COMMERCIALIZATION

The success of the Springback Predictability Project could significantly improve the global competitiveness of the U.S. automotive industry by eliminating the high costs associated with tryout and rework of dies used in sheet forming manufacturing operations. For the U.S. economy, from a broad perspective, the results of the Springback Predictability Project may be substantial.

Metal stampings are an indispensable, pervasive part of the contemporary industrialized society. An examination of most machines and products used in the U.S. manufacturing industries will disclose metal stampings in the assembly of a myriad of products. Currently, stampings are widely used in the manufacture of machines, tools, vehicles of all kinds, household appliances, hardware, office equipment, electrical and electronic equipment, containers, buildings, clothing, and most manufactured products. Broad application of the Springback predictability technology to be developed in is program may impact all sectors of manufacturing in which sheet metal stamping is the primary method of sheet metal forming.

The metrics listed below define measures of success that can be used to define project success and can serve as a description of the most important attribute for commercialization of the new FEM computer code to be used to model die performance before actual production.

1. Determine the savings expected (automotive and all durable goods manufacturing sectors) in die production costs and tryout as a result of the SPP program.

2. Definition of the reduction of cycle time associated with new product introduction.

3. Quantify improved quality in sheet metal stamping as a result of the SPP program.

Each of the possible attributes mentioned above are company-specific proprietary and business sensitive. Measures will have to be developed that provide generic information that can be used to provide required commercialization data.

ACTIVITIES AT NIST IN SUPPORT OF SPP

The work outlined below was initiated before awarding of the ATP project to SPP. However it is felt to be relevant as a starting point from which to address certain research topics carried out at NIST that complement the research undertaken by the SPP project.

The first accomplishment was to design, build and test a computer-controlled tensile stage for in situ x-ray measurements. The completed tensile stage is both small and light enough to be mounted on an x-ray goniometer. It can deform a specimen in situ while simultaneously measuring and recording the average longitudinal stress and the average linear strain. Since the sample stage must handle sensitive single-crystal samples, great care was taken in the design of the sample clamping mechanism.

In situ tensile experiments have already been conducted on beam line X23A3 at Brookhaven National Laboratory's National Synchrotron Light Source (NSLS); the tensile stage worked perfectly throughout these experiments. A paper has been written describing the tensile-stage system and a preprint is provided for the members of the SPP project.

Initially it was intended to obtain aluminum samples from commercial sources. Unfortunately, these samples were too highly defected to be usable. To overcome this unexpected difficulty, a facility to produce the needed samples was designed and built. Since this provided complete control over the entire sample production process, the system was designed to fabricate shaped samples that can be easily sliced into dog-bone shapes for the tensile stage.

Since any mechanical processing of the Al samples would introduce unacceptable numbers of defects, all processing is done chemically using an acid saw. The saw can slice high-quality individual samples as thin as 0.2 mm as required for the line-profiling experiments. After cutting, the samples are chemically polished and annealed in a vacuum furnace.

In situ tensile experiments on these samples have been conducted on the beam line at the NSLS. These experiments included both high-resolution diffraction imaging (x-ray topography) and small angle x-ray scattering (SAXS), which is line-profiling of the zero beam.

NIST has succeeded in developing two new techniques for measuring dislocation structure evolution in situ in bulk samples: in situ high-resolution diffraction imaging and in situ small angle x-ray

scattering. In addition, a completely new theoretical approach to the problem of x-ray scattering from dislocated materials has been developed. Using this approach, NIST has completed a rigorous and quantitative analysis of scattering from screw dislocations. Work on the edge dislocation problem is in progress. When completed, this theoretical work should allow quantitative physical parameters to be obtained from x-ray diffraction data; these parameters include the dislocation density, the dislocation ordering length, and the net Burgers vector of the sample. Qualitative information on the dislocation ordering mechanisms should also be obtainable.

These new experimental techniques and the new theoretical scattering analysis were developed for the purpose of studying dislocation structure evolution during uniaxial monotonic deformation, however, their applicability is far more general. The existing equipment could easily be used to examine dislocation structure evolution during fatigue, and dislocation structure interactions with grain boundaries. With some equipment modifications, experiments could also examine the effects of precipitate growth on work hardening.

When completed, the theoretical scattering analysis should find immediate broad applicability. X-ray line profiling is one of the most heavily used experimental probes of materials. The only quantitative analysis procedure currently available, however, is the Warren-Averbach approach; this technique extracts parameters that are not directly connected to physical quantities such as dislocation densities. This new analysis approach should provide this connection.

CONCLUSIONS

The main objectives of the proposed research are to develop a comprehensive simulation tool for analyzing sheet metal forming processes. The work has been focussed and, to a high degree, pragmatic. Each of the research and development tasks has a specific rationale and is simply and directly linked to the task of predicting sheet metal forming performance. These are delineated below:

1. We conduct extensive testing of sheet metal materials to determine the response of the selected materials in various strain paths including elastic relaxation from some states and including uniaxial load-unload-reverse load cycles.

   RATIONALE: These data will enable development of material models for use in predicting metal forming performance. These data will also be used to develop constitutive equations and will be implemented into a finite element computer code.

2. We conduct extensive testing of metal/die interaction to quantify the magnitude of the interfacial frictional stresses.

   RATIONALE: These data can be used immediately in existing finite element codes to improve the accuracy of forming simulation. More importantly, however, these data provide a basis for constructing a phenomenological theory of friction upon

which algorithms may be developed and interfaced with existing finite element codes.

3. With our code providers, we develop new finite element codes for more accurate and efficient simulation of sheet metal forming processes.

RATIONALE: Existing commercial finite element codes suffer from (1) lack of accuracy, (2) lack of robustness, and (3) high cost to run.

4. We validate the 3-D finite element code first in the laboratory.

RATIONALE: Validate the code on benchmark geometries and on scale model panels before it is applied to more complex automotive stampings.

5. We then validate the springback predicted by the finite element code on production intent dies.

RATIONALE: Validation of the technology is a necessary step for confidently performing modeling and simulation of springback early in the die design process thereby saving time and money as well as improving stamping quality.

6. NIST will complete the development of a theoretical foundation of x-ray diffraction and fabricate associated experimental apparatus to allow quantitative physical parameters to be obtained from x-ray diffraction data.

RATIONALE: NIST technical staff have developed a new theoretical approach to the problem of x-ray scattering from dislocated materials and new techniques for measuring dislocation structure evolution in situ in bulk material samples. These advances will enhance the research objectives of the SPP project in material modeling.

Success in achieving these objectives will result in the application of sheet metal forming simulation to assure a robust process prior to die construction, thus significantly reducing tryout time, the high costs associated with die tryout, and a reduction of the time required for the introduction of new vehicle products.

Acknowledgments

Edmund Herman - Task Leader, Friction Modeling
Group Leader , NAO Manufacturing Center
General Motors

Sing Tang - Task Leader, FEM Code Development
Senior Staff Scientist
Ford Motor Co.

Edmund Chu - Task Leader, Laboratory Validation
Senior Technical Specialist
Alcoa

Fang Cheng - Task Leader, Field Validation & SPP Technical Manager
Manager, Stamping Technologies
Chrysler Corp.

George Peace - SPP Project Administrator
Research Manager
ERIM

Richard Fields
NIST Technical Coordinator

Internal NIST Research Complementing the "Springback Predictability" and "Agile Precision Sheet Metal Stamping" Projects

L. E. Levine, Robb Thomson, G. G. Long, D. R. Black, and R. J. Fields

NIST, Gaithersburg, MD 20899

(August 25, 1997)

Abstract

Finite element modeling of metal forming operations has great potential for decreasing the time and cost of these operations as well as allowing increased precision of the resulting parts. To realize this potential, new constitutive laws must be developed that can accurately predict a material's behavior for large strains and complicated loading paths. The "Springback Predictability" and "Agile Precision Sheet Metal Stamping" Projects are working to solve this problem for a few specific metal alloys. In support of these projects, we have developed new measurement and analysis techniques for characterizing the microstructural changes that are directly responsible for the changes in material properties that occur during forming. These techniques include *in situ* high-resolution diffraction imaging, *in situ* ultra-small-angle X-ray scattering, and *in situ* profiling of Bragg peaks. To analyze the results of the scattering measurements, completely new analysis techniques were developed. Data from this program will be used to develop more general constitutive laws that are based upon measured physical parameters.

I. INTRODUCTION

An increasing percentage of metal forming operations are being designed using advanced finite element method (FEM) codes. This increase is being driven by the need to dramatically decrease the time and cost of new product development and to increase the precision of metal forming processes. Given sufficiently good models of material behavior during deformation, an FEM program could potentially be used to design dies of any required precision with no costly tryout and redesign steps. In practice, of course, several such steps are often required even for conventional steel stampings. When advanced materials such as high-strength steels and aluminum alloys are needed (as in future automobiles), the material behavior is less well understood; the resulting high prices for such parts are often prohibitive.

This problem is currently being addressed by the "Springback Predictability" and "Agile Precision Sheet Metal Stamping" Projects supported by the National Institute of Technology's (NIST's) Advanced Technology Program (ATP). Both projects will rely on FEM codes capable of accurately predicting the behavior of metal sheet during and after the forming and trimming process. These codes must handle large strain plasticity, non-proportional and reversed loading paths, and bending and unbending of thin sheets. The improved codes will use internal state variables in advanced constitutive relations between strain, strain rate, and stress.

The constitutive relations commonly used in FEM codes do not employ internal state variables. The use of these variables is generally unnecessary when the magnitude and complexity of the deformation are small and when high precision is not required. When high-precision parts with complicated shapes are needed, however, designers often introduce one or two internal state variables that keep track of the material's "state" better than just the strain; strain is usually considered a poor state variable, but is commonly used as such. While the internal state variable approach has been successful, there is practically no physical understanding or meaning attributed to existing internal state variables. The variables are measured by performing extensive mechanical testing of the material of interest. Due to time constraints, both of the above ATP projects are following this procedure and are working to develop workable solutions for just a few specific alloys. The general problem is not yet being directly addressed. Work at NIST is focused on this problem of developing general internal state variables for use in FEM codes. This work would expand the usefulness of both ATP projects by allowing their results to be applied to a much wider range of alloys.

Generally useful internal state variables must reflect the underlying physical processes that are directly responsible for the changes in material properties. These microstructural processes are very poorly understood. Since existing measurement techniques for measuring these changes was inadequate, NIST's first task was to develop new microstructural characterization techniques using unique NIST facilities such as the Cold Neutron Research Facility, the National Synchrotron Light Source at Brookhaven, and the Advanced Photon Source at Argonne. These techniques feature *in situ*, real-time capabilities using thicknesses of materials that are relevant to sheet metal forming.

This paper presents our progress in measuring those features of the microstructure most relevant to plastic deformation: dislocation structures. Dislocations are quanta of plastic deformation and their concentration, sign, character, and location are ultimately responsible for macroscopic plasticity. Dislocations interact with each other and form complex structures

that are in some sense more important to the understanding of plastic deformation than the dislocations themselves. Once an adequate understanding of these physical changes is reached, work on developing a suitable set of constitutive equations can proceed.

II. STUDIES OF DISLOCATION STRUCTURE EVOLUTION

Since the formation of dislocation structures in a deforming metal is a bulk process and since these structures usually extend over several micrometers in size, a minimum sample thickness of approximately 50 μm is required for *in situ* experiments. This eliminates the use of electrons as a possible probe leaving X-rays and neutrons as the most viable candidates. The high brightness and small source size of X-ray synchrotron radiation, along with the high spatial resolution and sensitivity of X-ray detectors, makes X-rays the preferred probe for these *in situ* measurements.

To conduct these *in situ* X-ray experiments, we have designed and built a computer-controlled low-profile tensile stage that is both small and light enough to be mounted on an X-ray goniometer.[1] It can deform a specimen *in situ* while simultaneously measuring the average longitudinal stress and the average linear strain. To avoid introducing undesirable defects into the sample, the tensile stage was carefully designed to apply tension along a well defined axis with minimal torques. This equipment allows *in situ* X-ray measurements to be conducted on single-crystal and large grain polycrystalline samples during monotonic deformation, fatigue, creep, and relaxation.

For all of the X-ray experiments, samples must satisfy two important criteria; they must be single crystals or large-grain polycrystals and they must take the form of flat dog-bones with a thickness of either ≈ 1.0 mm or ≈ 0.2 mm depending upon the specific experiment. Also, it is highly desirable to make multiple samples with the same crystallographic orientation with respect to the tensile axis. To accomplish this, we make starting blanks of high purity (99.99%) polycrystalline Al, shaped so that a series of straight parallel cuts will produce dog-bone samples with the required dimensions. The lower end of the blank is sharpened to a point which produces a single nucleation event during directional solidification. The blank is melted in a soft-mold of Al_2O_3 powder in the upper zone of a two-zone temperature controlled furnace. The upper and lower zones are set at 720° C and 620° C, respectively. The sample is lowered into the lower zone at a rate of 1.27 cm/hr, producing a single randomly oriented crystal with the shape of the original blank.

After subsequent cooling and removal from the furnace, the blank is sliced in a multiple-wire acid saw using a solution of 20% HNO_3, 40% HCl and 40% H_2O. The resulting single-crystal dog-bone samples are then chemically etched at room temperature in a bath of 10% HF, 49% HNO_3 and 41% H_2O. After washing, the samples are transferred onto a flat window of fused quartz in a vacuum oven where they are annealed at 450° C for 48 hrs. at a pressure of 2×10^{-5} Torr. Throughout the entire sample-making process, extreme care is required to avoid introducing defects into the samples. The resulting samples are generally of very high quality, with dislocation densities as low as $\approx 10^3/cm^2$.

Since slicing in the acid saw produces multiple samples with the same crystallographic orientation, we only need to determine the orientation of a single sample from such a set. The crystal orientation is determined using a conventional X-ray backscatter Laue camera.

In order to relate the dislocation structure evolution to the mechanical properties, it is necessary to obtain stress-strain data from the sample as it is deformed. The computer-controlled tensile stage system has four sensors that are used to determine the average longitudinal stress and the average linear strain. This system automatically acquires this stress-strain data every 0.5 seconds throughout the experiment. These data are stored on the computer for later comparison with the dislocation structure images. Figure 1 shows an engineering stress-strain curve taken during an *in situ* X-ray experiment. This particular sample has a relatively high number of pre-existing dislocations as indicated by the flow stress. Some samples we have produced have flow stresses below 4 MPa, but such very low dislocation density samples are not really necessary for these experiments.

A combination of several complementary *in situ* X-ray techniques is being used to fully characterize the dislocation structure evolution. These techniques include high-resolution diffraction imaging, ultra-small-angle X-ray scattering (USAXS), and Bragg peak profiling. At present, the X-ray experiments are being conducted on the NIST beam line at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Later experiments will be conducted on the new NIST beam line at the Advanced Photon Source (APS) at Argonne National Laboratory. Each of the three experimental techniques will be discussed in turn.

The early stages (up to $\approx 8\%$ strain) of the deformation process are being studied using monochromatic, high-resolution diffraction imaging on single-crystal samples deformed *in situ*. This technique allows *in situ* microscopic, stereoscopic imaging of the dislocation structures as the sample is strained.

At the X-ray synchrotron beam line, the sample is loaded into the tensile stage which is then mounted on an X-ray goniometer. All diffraction images have so far been obtained using a 002 reflection; under these conditions, all of the dislocations should be visible in the images. Future experiments could obtain images using several different reflections corresponding to dislocation Burgers vectors, thus allowing identification of the Burgers vectors of individual dislocations.[2]

The first step in a diffraction imaging tensile experiment is to locate the appropriate X-ray reflection and take stereographic high-resolution diffraction images from the unstrained sample. Nuclear emulsion film is used for these images giving a resolution better than 1 μm. Stereo images are taken by rotating the sample around the axis perpendicular to the diffracting planes in the crystal;[3] the usual stereo rotation angle is 10°. The sample is then strained *in situ* at a constant strain rate to a predetermined average longitudinal strain. In these experiments, the strain rate is usually 0.2%/min. When the desired strain is reached, the tensile stage is halted and the stress on the sample is maintained by the stage's compression springs. Stereographic high-resolution diffraction images are again taken from the strained sample. Exposure time for each image is approximately 6 min. After the images are acquired, the sample deformation resumes at the same strain rate and the above process is repeated until the end of the experiment.

In addition to the high-resolution diffraction images, an X-ray video camera with a resolution of 11 microns takes digitized diffraction images every 5 seconds throughout the straining procedure. These images are later combined into a low-resolution movie of the deformation process. This information greatly aids the analysis of the high-resolution image data.

Small relative rotations across cell walls occur as same-sign edge dislocations accumulate.[10] Since the image formation process is very sensitive to sample orientation, these relative rotations can be measured directly by rotating the sample in the beam and recording what regions are diffracting. These rotation sequences are performed when the tensile stage is halted for taking high-resolution diffraction images; the rotation sequences are recorded on video tape for later analysis.

Figure 2 shows a high-resolution diffraction image of a single-crystal Al sample deformed *in situ* by uniaxial tension along a [430] axis to a strain of 1.9%. The stress-strain curve in Fig. 1 was obtained during this experiment. The diagonal white linear features are planar dislocation structures aligned parallel to the primary {111} slip planes. The spacing between these "carpet" structures ranges from 15-50 micrometers. The large ragged right-angle features labeled A at the lower right hand edge of the image are caused by small rotations across the dislocation walls. We have measured these rotations to be approximately 0.005°. The linear white lines running nearly horizontally across the center of the image are grown-in dislocation structures. Amazingly, these grown-in structures and the carpet dislocations have had no measurable influence on one another. Thus, the mobile dislocations on the primary slip planes simply cut through the existing immobile dislocations.

High-resolution diffraction imaging provides important information on the local dislocation structure evolution and also aids in our analysis of X-ray scattering experiments. In contrast, X-ray and neutron scattering studies of deformed metals must rely on complicated scattering theories to interpret the results. The reason is that a scattering measurement does not yield a visual picture of the structure, but rather of its Fourier transform. Nevertheless, in principle, the Fourier transform contains a great deal of information about the structure of the sample, and this information takes the form of averages over the structure. In fact, it is precisely such averages which one can use to characterize those aspects of the structure which lead to the mechanical constitutive relations. In contrast, imaging techniques are much more difficult to relate to these global properties. Thus, diffraction tools are complementary to imaging techniques, and are very powerful in their own right.

A detailed reading of the literature revealed that the theory of X-ray and neutron scattering from dislocations is in a very primitive state.[5-10] Theoretical treatments of Bragg scattering were of limited usefulness and investigations of small-angle scattering from dislocation structures were non-existent. We therefore initiated a theoretical effort to develop usable theories for both USAXS and Bragg peak profiling.

The goal for our theoretical effort is to enable us to extract meaningful physical parameters from X-ray and neutron scattering experiments. The theories for USAXS and Bragg peak profiling require very different mathematical approximations so the two problems are quite distinct. Following from this, the information one can obtain from the two types of experiments is also quite different. In particular, Bragg peak analysis yields information on the gross densities of dislocations, and on the length scale over which dislocations order themselves to eliminate long-range strains. USAXS, on the other hand, provides important information on the spatial structures formed by the dislocations during deformation.

Since the required theoretical analysis was very complex, we validated the theoretical models by conducting extensive computer simulations of diffraction from dislocation structures. These simulations also allow us to simulate scattering results from dislocation arrangements which cannot be modeled analytically. This combination of analytical models and

direct computer simulation shows promise as an exceedingly powerful tool for interpreting dislocation structures in deformed metals.

In USAXS, a highly parallel mono-chromatic X-ray beam interacts with a thin specimen. The distribution of X-ray intensity about the main beam contains information on the scattering structures within the sample. We have shown that the small-angle scattering consists of a product of several independent structure factors. These include structure factors for the individual dislocations, and for the dislocation arrangements both along the length of a dislocation wall and perpendicular to the wall. Each of these structure factors depends critically upon such quantities as the dislocation density distribution across a wall, and the three dimensional orientations of the sample, the dislocations, the edge components of the Burgers vectors, and the scattering vector, q.

Figure 3 shows USAXS data obtained from a 0.18 mm thick sample cut from the same starting crystal that was used to make the diffraction imaging sample discussed above. As before, the sample was deformed *in situ* by uniaxial tension along a [430] axis. These USAXS results show excellent sensitivity to changes in the scattering as the samples are strained *in situ*. Due to the many difficult problems that needed to be overcome (such as multiple Bragg scattering from polycrystals, limits in the accessible q-range in conventional SAXS, and beam normalization difficulties), these experimental results are the first such results ever reported. The scattering intensity is plotted as a function of the reciprocal lattice vector, q. The data takes the form of two intersecting straight lines on a log-log plot. The regions on the sides of this intersection at about 0.001 Å^{-1} are referred to as the low-q and high-q regions.

To interpret these results, we examined the dislocation structures in the high-resolution diffraction images and used our theoretical models to make predictions of the USAXS scattering. This procedure allowed us to relate all of the features present in the data to various aspects of the evolving dislocation structures. Since our results are very new and not yet complete, they should be considered preliminary. In the low-q region, we are sensitive to the dislocations in the $\{111\}$ walls. In the high-q region, we see the contribution from all of the dislocations in the sample, with no effect from the dislocation structures.

Several systematic changes can be seen in these data as a function of the strain. Figure 4 displays the log-log slopes in the low-q region. The slopes steadily increase to a terminal slope of -3. Correcting for slit smearing gives a terminal slope of -4. This slope change is indicative of a change in the distribution of dislocations across the $\{111\}$ dislocation walls. Thus, the dislocation walls are very diffuse at small strains and become sharper with increasing strain. Additional work on the theory should allow us to extract the changing width of the diffuse boundary from the data. The corresponding terminal slope of -2 for the high-q data is reached very quickly and remains constant out to the largest strain of 15%. This behavior is also predicted by the theory.

Figure 5 shows a plot of the integrated scattering intensity within the high-q region. This value is directly proportional to the total edge dislocation content in the sample. Thus, the dislocation density is observed to be nearly constant up to a strain of approximately 8% and to grow rapidly after this point. The onset of this increase in dislocation density corresponds closely to the point at which the $\{111\}$ dislocation walls become sharp.

Combining all of the data from the diffraction imaging and USAXS experiments provides a preliminary, yet fully coherent, picture of what is happening to the dislocation structures

in this sample from the point at which the {111} dislocation carpets are formed. In the early stages, the {111} walls are very diffuse. As the strain, and thus the stress, is increased, the increased shear stress eliminates the more weakly bound dislocations from the walls, causing the diffuse boundaries of the walls to sharpen. This loss of dislocations is compensated for by either the incorporation of new dislocations into the walls or the formation of new walls. (This question could be answered with additional imaging data and comparison with the existing USAXS measurements of the total number of dislocations.) During this stage, the dislocation content remains nearly constant, presumably because the {111} channels between the walls provide ready paths for dislocations to leave the sample.

At a strain of approximately 8%, two changes occur: the dislocation walls become very sharp and the dislocation content of the sample begins to increase rapidly. At this same strain, X-ray video images of the samples lose all contrast, also indicating an abrupt increase in the number of dislocations. By applying our new scattering theory to the existing experimental conditions, we find that if additional walls were forming perpendicular to the {111} walls, then their signature would be indistinguishable from the background population of dislocations. The only observable change would be the increase in the number of dislocations. We therefore associate the changes that occur around 8% strain with the onset of cell wall formation through growth of new walls perpendicular to the {111} carpet walls. This closes off the channels and traps many of the mobile dislocations as they form.

Further theoretical work should allow us to extract several quantitative results from these data. We should be able to directly determine the total number of dislocations in the sample, the percentage of these dislocations present in the walls, and the changing width of the diffuse boundary of the walls. Additional data could provide many other important parameters such as the ratio of dislocation dipoles to monopoles, the average dislocation spacing in the walls, the width of the walls, and a measure of how disordered the dislocations are within the walls.

Comparison between the above description of the dislocation structure evolution and the stress-strain data in Fig. 1 leads to several important conclusions. First, it is often assumed in the literature that the square of the average dislocation density should be approximately proportional to the applied stress. This is clearly not true in our samples for strains up to 8%. Instead, the hardness change in this range is primarily caused by a change in the dislocation structure. Thus, the average dislocation density and the structures they form both play a role in determining the hardening of a plastically deformed sample. Second, the X-ray measurements identified a major change in both the average dislocation density and the dislocation structures at a strain of around 8%. Examination of Fig. 1 shows that the stress-strain curve remains quite smooth during this transition. Development of improved constitutive laws based upon the microstructure can only be accomplished through greater understanding of how the dislocation density and structure interact to produce such smooth stress-strain curves.

Experiments to measure the broadening of Bragg peaks have not yet been conducted since the necessary equipment is still under construction. These experiments will be conducted on a new beam line at the Advanced Photon Source (APS) at Argonne National Laboratory. Although experimental data is not yet available, work on developing a theory for interpreting these results is well under way.

As with existing SAXS theory, the theory of Bragg peak broadening by X-rays and neu-

trons is in a very primitive state. We have published a new approach to the problem, which we confirmed by conducting simulations on screw dislocation distributions.[11] A second paper has been partially completed extending the results to the more important edge dislocation case, which we expect to serve as a guide for more complex computer simulation studies. The screw dislocation results suggest that the Bragg peak structure is determined by two properties of the dislocations. One parameter is the total average dislocation density, and the second is the effective range of the dislocation stress field. This second parameter is important in interpreting the size and shape of dislocation structures, because it gives a measure of the dislocation self energy. In practice, Bragg peaks have been found to have an asymmetry which can be interpreted in terms of the relative densities of dislocations in walls and in the interiors of walls. So the Bragg peak analysis, in addition to giving direct information on the dislocation density, also gives at least indirect information on the dislocation structures.

III. CONCLUSIONS

We have demonstrated that the new experimental and analysis techniques discussed above allow us to directly characterize the evolution of dislocation structures in deforming metal specimens. Once the development work is complete, an extensive measurement program using large numbers of samples is planned for the nearly completed NIST beamline at the Advanced Photon Source at Argonne National Laboratory. These samples will not be restricted to high-purity single-crystal material. Instead, important factors such as precipitate hardening and the role of grain boundaries will also be investigated. The resulting data will comprise the first in-depth and *in situ* look at the evolution of dislocation structures in bulk-thickness samples. As discussed in the Introduction, these results will be used to guide and test the future development of advanced constitutive laws for FEM of metal forming.

REFERENCES

[1] L. E. Levine and R. J. Fields, submitted to Rev. Sci. Instrum.

[2] Bruce Steiner, L. E. Levine, Margaret Brown and David Larson, *J. Cryst. Growth*, **169**, 1 (1996)

[3] Kyoichi Haruta, *J. Appl. Phys.*, **76**, 1789 (1965)

[4] A. S. Argon and P. Haasen, *Acta metall. mater.*, **41**, 3289 (1993)

[5] M. A. Krivoglaz, *X-Ray and Neutron Diffraction in Nonideal Crystals*. Berlin: Springer. (1996)

[6] M. Wilkens, *Fundamental Aspects of Dislocation Theory*. ed. Simmons, J. A., de Wit, R. & Bullough, R., vol II. Washington: National Bureau of Standards, Spec. Publ. 317, 1191-1193, 1195-1221 (1970)

[7] M. Wilkens, *Phys. Stat. Sol.* (a), **2**, 359-370 (1970)

[8] M. Wilkens, *Microstructural Characterization of Materials by Non-Microscopic Techniques*. ed. Anderson, N. H., Eldrup, M., Hansen, N., Jensen, D. J., Leffers, T., Lifholt, H., Pedersen, O. B. & Singh, B. N. Denmark: Riso Nat. Lab. Roskilde, 153-168 (1984)

[9] A. Seeger and P. Brand., *Small Angle Scattering*, H. Brumberger, ed., Gordon and Breach, NY, (1965)

[10] H. H. Atkinson and P. B. Hirsch, *Phil. Mag.*, **3**, 476, (1958)

[11] L. E. Levine and Robb Thomson, *Acta Cryst. A*, in press

FIGURES

FIG. 1. Engineering stress-strain curve obtained during an *in situ* experiment. The initial flow stress of 18 MPa indicates that a significant population of dislocations existed in the unstrained sample.

FIG. 2. High-resolution diffraction image of single-crystal sample with an applied strain of 1.9%. The white features are dislocation structures.

FIG. 3. Ultra-small-angle X-ray scattering data obtained *in situ*. Dramatic changes in the spectra were observed with increasing deformation.

FIG. 4. Changes in the log-log slopes in the low-q region of the ultra-small-angle X-ray scattering data. These changes reflect the increasing sharpness of the dislocation walls.

FIG. 5. Integrated scattering intensity in the high-q region of the ultra-small-angle X-ray scattering data. This quantity is believed to be directly proportional to the total edge dislocation content of the sample.

Figure 1

Figure 2.

100μm

A

Session VI. Surface Treatment and Material Processing

Cubic Boron Nitride Coatings for Cutting and Specialty Tools
A. Inspektor, A.S. Gates, U. Schleinkofer, and E. Oles,
Kennametal; T. Lefkow, Northwestern University; W. Walch
and R.L. Resnick, Extrude Hone

Industrial Process Driven System Requirements for PSII
Applications
C.P. Munson, Los Alamos National Laboratory

Plasma-Based Ion Implantation: Surface Treatments for
Automobile Manufacturing
J.V. Mantese, GM Research Laboratories

The Commercialization of PSII into Multiple-User Markets
M.C. Dudzik, Environmental Research Institute of Michigan

Cubic Boron Nitride Coatings for Cutting and Specialty Tools

A. Inspektor, A. Gates, U. Schleinkofer and E. Oles
Kennametal Inc., Latrobe PA

T. Lefkow
Northwestern University, Evanston Il

W. Walch and R.L. Resnick
Extrudehone, Irwin PA

Introduction

Cubic boron nitride (cBN) coatings hold the potential to become the superior material of choice for cutting tools for ferrous materials and in a wide variety of specialty tools applications. Boron nitride is a III-V group material. Its system of crystalline phases is analogous to that of carbon with three corresponding crystallographic structures: (1) Hexagonal boron nitride (hBN) with a hexagonal graphitic structure, which is the stable phase at room temperature; (2) Cubic boron nitride (cBN) with a cubic, diamond structure, which is the high-pressure high-temperature phase; and (3) Wurtzitic boron nitride (wBN) with a wurtzite, hexagonal diamond structure, which is the high-pressure high-temperature phase. The cubic phase, cBN, is the second hardest material to diamond, up to 60 GPa (The hardness of diamond is 100GPa). Cubic boron nitride is thermally stable in air (up to $1300\,°C$ $-1400°C$) and does not react with metals such as iron, cobalt, and nickel at temperatures up to $1000\,°C$ $-1200\,°C$. These properties are in sharp contrast to that of diamond which, though harder, is easily oxidized in air at above $600\,°C$ and is rapidly dissolved by hot iron, cobalt, or nickel. Consequently, cBN is a superior tool material for machining difficult to cut ferrous workpiece materials. Likewise, diamond is the tool of choice for abrasive non-ferrous alloys and non-metallic materials.

The discovery in 1957 of the high-pressure high-temperature (HPHT) synthesis of the cubic phase paved the way for the production of cBN based superhard cutting tools, sintered polycrystalline cubic boron nitride (PCBN). The cutting edge of the PCBN tool is a superhard tip brazed onto a carbide insert. The tip is cut from a blank made of micrometer sized HPHT cBN crystals, intergrown with a suitable ceramic binder, and bonded onto a cemented carbide support. PCBN tools are extremely successful in machining gray cast iron, the most abundantly used workpiece material in the heavy and automotive industries. Some of the typical automotive applications of PCBN tools include machining of pumps, impellers, shafts, engine blocks, brake rotors and drums, and other parts made of cast iron. PCBN tools are also among the very few tools capable of economical machining powder metal (PM) parts, like the 3.11 V6 engine, pressure plates, sprockets, cam shafts, gears, etc. Other automotive applications of PCBN include machining of turbine parts and engine shafts made of superalloys as well as machining of pinion and side gears, transmission parts and bearings, all made of hardened steels. Hard turning is another expanding application of PCBN tools. Hard turning is being introduced to replace many of the present grinding processes, with the advantages of faster stock removal, one-step

287

processing of complex parts, and shorter set up times in small runs. An additional advantage of the PCBN tool is its compatibility with dry machining that will allow future conformity with stringent environmental regulations on coolants. The automotive industry is the largest PCBN tool user, followed by the heavy equipment industry. Both constitute 66% of the total market, with aerospace, pumps, and turning industries comprising the balance.

The PCBN tools are expensive, usually single tipped, and limited to a relatively few styles with a planar geometry. None the less, because of the superior combination of hardness and thermal and chemical stability these are often the only existing tools suitable for the economic machining of difficult to cut ferrous materials. As a result, PCBN tools are the highest growth rate segment among the available cutting tool materials. Yet, in analogy to similar polycrystalline diamond (PCD) tools vis-à-vis the recently introduced CVD diamond coated cutting tools /1,2/, a cBN coated tool, when developed, can open up new possibilities in tool manufacturing and in machining of advanced, hard to cut, ferrous materials. With a tailor-made-geometry, cBN coated tool should offer multiple cutting edges with chip control geometries while eliminating depth of cut (DOC) limitations. When developed, it may greatly enhance the application range and an economic machining process of superhard workpiece materials. Starting in the early 80's, experiments indicated the possibility of synthesizing cBN thin films by low-pressure techniques. These early results have been supported by recent advances in ion assisted technologies and by the successful introduction of diamond coated tools. As a result, the development of cBN coating technology has become one of the primary goals of the metalcutting tool industry and is the focus of this NIST-ATP program /3/. This paper discusses some of the key factors involved in the development of cBN films for metalcutting applications and presents the major experimental approaches in this project.

CBN Coating Technologies

In the recent years, thin films of boron nitride with a predominantly metastable cubic phase have been synthesized using different deposition techniques. Most of these are physical vapor deposition (PVD) techniques with different combinations of methods to deliver boron and ionized nitrogen to the reaction site. In most cases, elemental boron is obtained by electron beam evaporation of boron /e.g.: 4/. Excimer laser ablation /5/, ion beam assisted sputtering /6/, and rf diode sputtering of boron and boron carbide targets /7,8/ have also been reported as boron sources. Ionized nitrogen is supplied by low- or high-energy ion beams /4,9/, or created in rf bias to the substrate (with or without magnetic field) /10/, hollow cathode plasma and by mw discharge.

The common theme to all successful techniques is the need for intense ion bombardment of the film. It was found that energetic ion bombardment, high substrate temperature, and film stoichiometry are the necessary conditions for the deposition and stabilization of the cubic phase /4,11/. Accordingly, a number of mechanisms has been proposed for the stabilization of the metastable cubic phase. The present thoughts on the deposition of cBN films include: Thermal spikes that lead to a formation of local HPHT conditions /12/; Hydrostatic compressive stress that can be describe as a dynamic HPLT condition /13/; Subsurface stabilization of the sp^3

bonding with resultant kinetic stress at the sublayer and creation/annihilation of defects /14/; Subplantation process /15/; and Selective resputtering of the hexagonal phase /16/.

Other common issues are the small crystal size of the film and high compressive stress that leads to poor adhesion. The hexagonal phase is the stable phase at room temperature while cubic is the high-temperature, high-pressure phase. Because of this, a cBN film can only be grown as a coating by the imposition of compressive stress in the lattice. As the coating layer grows the compressive stress also tends to grow, until it becomes so high that it causes the coating to peel off. This is the critical problem with the deposition of cBN films. Accordingly, the key technical goals of this project are the development of commercially viable thick (at least 3-5 μm) cBN coated tool with sufficient adhesion to withstand the shearing forces of machining processes.

To meet these goals the project aims at two complementary tasks: (1) Methods to improve the adhesion and the coherence of thick cBN films, which will be the subject of a later publication and (2) Comparative study of several deposition techniques to identify a reliable path for the deposition of thick cBN films. Three deposition methods being investigated: (1) Ion beam assisted deposition, (2) Plasma assisted chemical vapor deposition and (3) Unbalanced magnetron reactive sputtering. The corresponding experimental approaches are schematically shown in Fig.1.
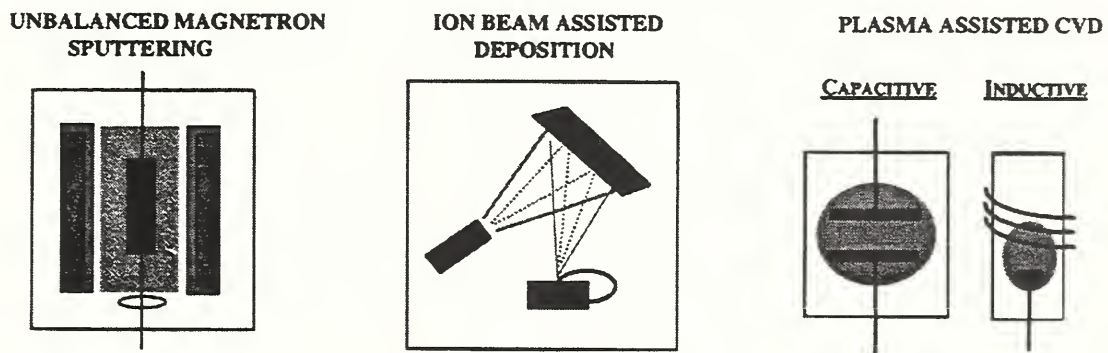
Fig. 1: Schematic representation of the three deposition techniques: (1) Unbalanced magnetron reactive sputtering, (2) Ion beam assisted deposition, (3)Plasma assisted deposition.

Ion Beam Assisted Deposition

The ion beam assisted deposition (IBAD) method involves the use of energetic ion bombardment (nitrogen and argon) of the substrate that is simultaneously being coated with boron. The advantage of this technique is its ability to control the ion current density, ion/atom ratio and the ion beam energy, independently over very large ranges.

Ion beam assisted deposition experiments were carried out to explore the effects of the evaporation rate of boron, ion beam energy and current, substrate temperatures and substrate material (silicon, SiAlON and WC-Co) on the formation of cBN. It was found that the formation of the cubic phase depends critically on the correct balance between the flux of nitrogen ions, the evaporation rate of boron, and the energy of the ion beam.

The coatings were analyzed on silicon and cemented carbide substrate materials. It was observed that boron nitride in the cubic phase was obtained with this deposition technique for both substrates. FTIR shows a sp^3 bonding signature in both reflection and transmission modes. The typical absorption peak for cubic boron nitride occurs around 1050 cm^{-1} while hexagonal boron nitride exhibits two absorption bands at 1373 and 816 cm^{-1} /17/. Figure 2 shows a FTIR spectrum of cBN coated silicon in transmission and cBN coated cemented carbide in reflection. The absorption peak around 1060 cm^{-1} indicates boron nitride in cubic modification.
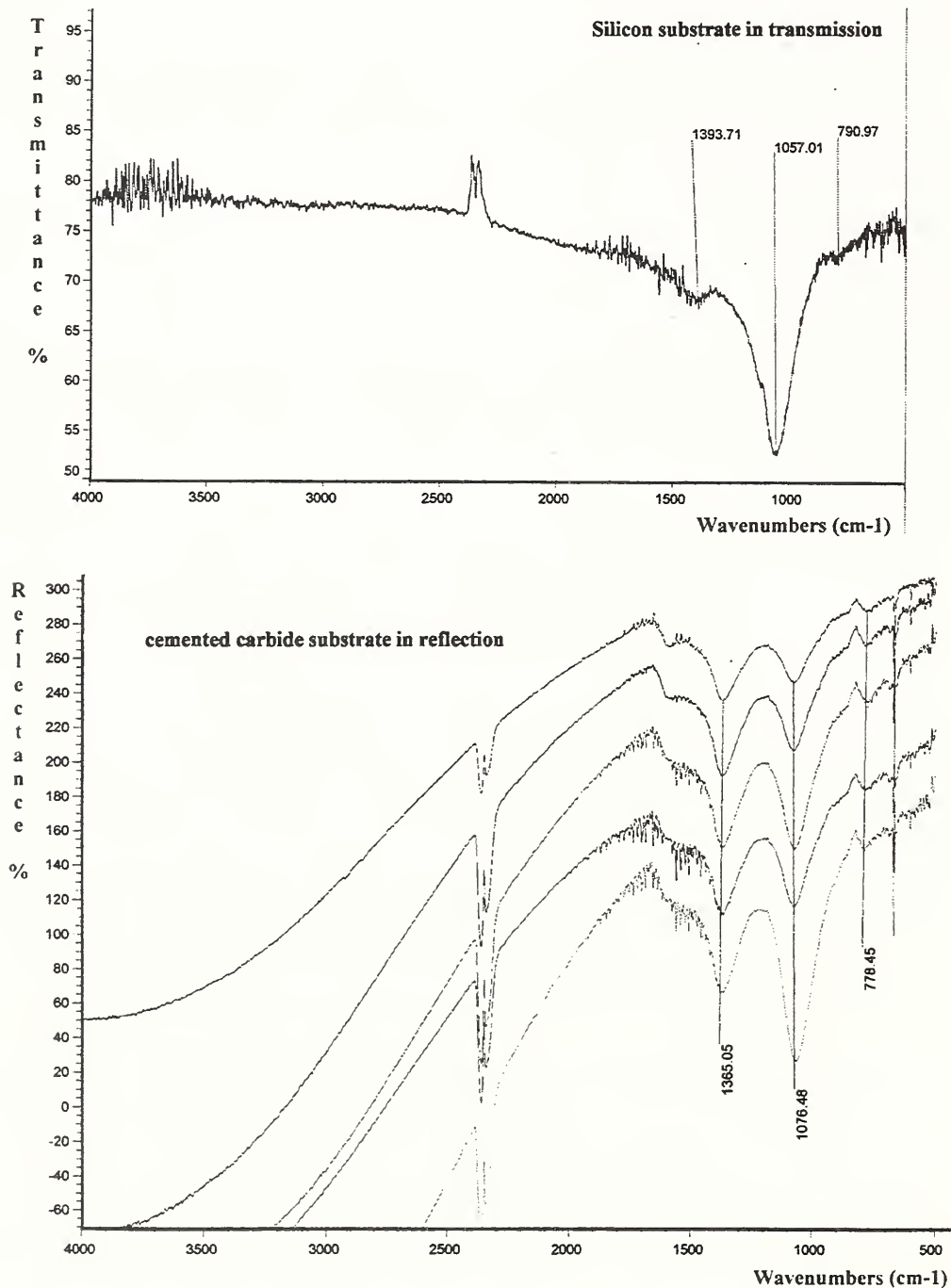
Fig. 2: FTIR spectra of IBAD boron nitride films on silicon and cemented carbide.

TEM-SAD technique confirms this result by showing sharp rings which corresponds to the lattice spacing close to cBN crystal structure (Fig. 3, silicon substrate). Similar SAD patterns on the carbide sample were also in good agreement with data for cBN and hBN.
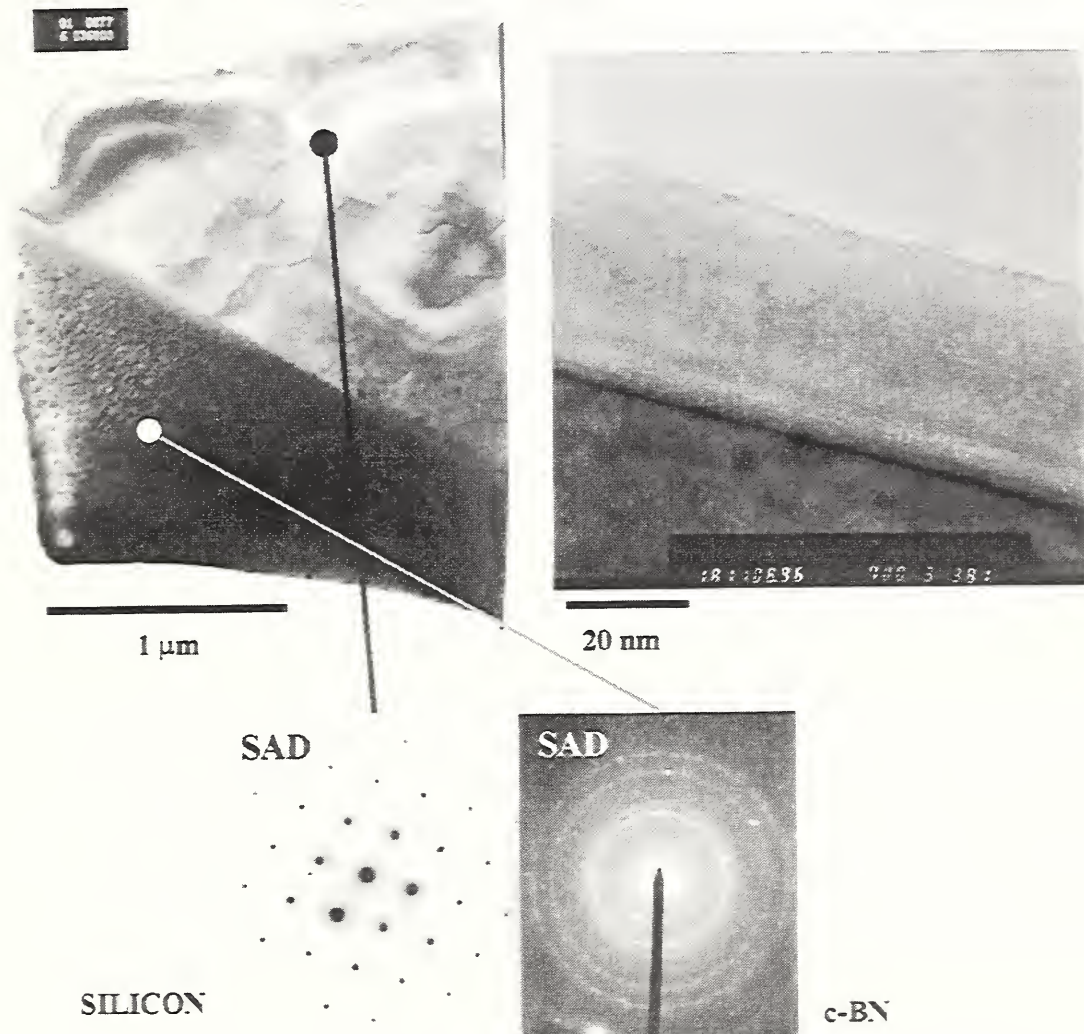
Fig 3: TEM micrograph with selective area diffraction (SAD) of IBAD cBN coating.

Plasma Assisted CVD

Plasma assisted chemical vapor deposition (PACVD) is a vapor deposition process in which plasma is used to activate both the chemical reactants and the treated surface. Plasma can be generated, directly, by a dc electrical breakdown between two electrodes or indirectly by a capacitive (C-PACVD) or inductive (I-PACVD) coupling of a high frequency electromagnetic field (Fig.1). The advantage of I-PACVD is that the intense plasma generated can be used to cause chemical reactions to take place to produce coatings at relatively low substrate temperatures in comparison with conventional chemical vapor deposition. When the substrate is rf biased, the plasma density can be controlled independently of the substrate bias.

M. Kuhr, S. Reinke, and W. Kulish have deposited cBN/hBN/xBN mixtures /18/ up to 75% cBN by plasma assisted deposition from trimethyl borazine, argon, and nitrogen gas mixtures onto silicon substrates at 550°C. The cBN coating mixtures had poor adhesion to the silicon substrate, probably due to the underlayer of hBN and amorphous BN beneath the cBN layer. Kulish et al. /19/ reported that cBN contents of 90% in cBN/hBN mixtures could be obtained at 700°C presumably due to lower hydrogen incorporation at higher temperatures and that improvements in adhesion could be produced using boron rich interlayers. Takanori Ichiki and Toyonobu Yoshida /20,21/ reported the deposition of cBN/hBN coating mixtures by IPACVD on silicon substrates using diborane, nitrogen, and argon gas mixtures at 900°C and showed that the deposition rate of cBN increased with decreasing pressure over the range of 2 to 0.5×10^{-3} torr. Finally, M. Z. Karim et al. /22/ report the formation of cBN/hBN coating mixtures on silicon and potassium bromide substrates from a BH_3NH_3 complex precursor at substrate temperatures of 48 to 180°C.

It was shown that BN films produced by I-PACVD without bias voltage are a mixture of amorphous and hBN. TEM indicate amorphous, turbostratic, and hexagonal layers, verified by SAD as diffuse diffraction ring pattern close to h-BN crystal structure. Plasma assisted deposition and characterization work continues and is a key element in the evaluation of the films under development.

Unbalanced Magnetron Sputtering

Unbalanced magnetron (UBM) sputtering /23,24/ is a common commercially available PVD coating technology that involves sputtering of a target material (in our case: boron carbide) in a controlled atmosphere of nitrogen and argon. The system (Fig. 1) includes strong magnets for strongly unbalancing the sputter cathodes and for plasma enhancement. The substrate is biased via an rf or a pulsed dc high frequency power supply. The combination of unbalanced magnet array with the biased substrate produces a high degree of controlled ion bombardment and an independent control of both the amount of ion bombardment and the deposition rate.

In a different way from the IBAD and the PACVD systems, in addition to experiments aimed at cBN films, the UBM unit was used also for the formation of structural templates. All the previous approaches rely on ion bombardment and on the resulting compressive stress to force the boron nitride into the cubic form. There is however another possible way to create a cubic structure: a structural template. In this method, a thin layer of normally hexagonal material can be forced into cubic structure when sandwiched between two structural template layers and a superlattice of alternating nanolayer films is formed /25-27/. The template material should be readily deposited and with a close match on some plane with the desired cubic structure. When alternating layers of the cubic template material and of the hexagonal material are deposited, the template can constrain the hexagonal layer to cubic form. Fig. 4 shows a typical nanolayer structure. The periodicity of the film depends on the substrate bias, on sputtering rate of each compound, and on the nitrogen partial pressure /25-27/. The utility of a nanolayer film in metalcutting applications will depend to great extent on the hot hardness of the coating and on the temperature threshold for inter-diffusion between the nanolayers.
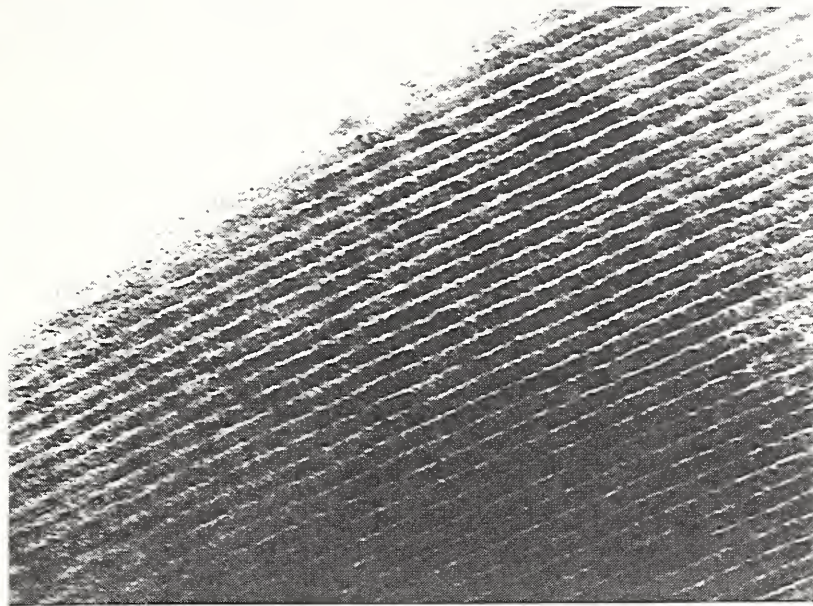
Fig. 4: TEM micrograph of a typical nanolayer film.

Analytical Methods To Determine The Structures Of Boron Nitride

Boron nitride deposited as thin films by different techniques tends to form in amorphous (aBN), turbostratic (tBN), hexagonal (hBN), wurzite (wBN), or cubic (cBN) structures. The bonding in these structures can be characterized as coplanar sp^2 or tetrahedal sp^3. Often, as discussed in the literature, the BN film is composed of several layers: amorphous basis, hexagonal transition, and the cBN layer. Fig. 3 (TEM-SAD on silicon substrate) is such an example: An amorphous underlayer of 10 nm, followed by 20 nm of hexagonal clusters and 20 nm of cubic clusters followed by a homogeneous crystalline cBN layer (confirmed by selected area diffraction SAD).

To determine the actual obtained structure of the deposited BN film two different approaches are possible. The first method relies on the macroscopic properties of boron nitride in different structures. For this approach hardness, micro indentation techniques, hot hardness, and color of the coating can be used to differentiate between the boron nitride modifications. The second approach to identify the structure of boron nitride coatings is more reliable and focuses on the microscopic properties such as atomic arrangement, bonding energies, and electronic properties. Here, different diffraction techniques like X-ray diffraction (XRD) or selected area diffraction (SAD) in a transmission electron microscope (TEM) can be used to determine the lattice properties, X-ray photoelectron spectroscopy (XPS) and infrared absorption technique (FTIR) are used to obtain information about the bonding energy in the structure, energy dispersive X-ray spectroscopy (EDX), Auger electron spectroscopy (AES), and secondary ion mass spectroscopy (SIMS) reveal information on the elements which form the structure, and optical microscopy (OM), scanning electron microscopy (SEM), transmission electron microscopy (TEM), and high resolution transmission electron microscopy (HRTEM) give

information about the morphology of the film. In this project, FTIR, XRD, TEM, SAD, and SEM/EDX were used to characterize the deposited BN films on silicon and cemented carbides. Some of the results are presented below.

As mentioned before, the FTIR technique shows the atomic bonding type in the coating with a typical cBN absorption peak around 1050 cm^{-1} (sp^3 bonding type) and two hBN absorption bands at 1367 cm^{-1} and 783 cm^{-1} (sp^2 bonding type) /17/. This difference reveals a good analytical tool to differentiate between the two structures of boron nitride. In practice, the substrate FTIR spectrum is always subtracted from the spectra of the coating for getting the deposit spectra solely. Figure 2 showed typical FTIR spectrum for an IBAD cBN coated silicon sample in transmission mode and for cBN coated cemented carbide in reflection mode. Because of the high infra-red absorption of a cemented carbide substrate, FTIR cannot be used in transmission mode on this substrate material. In addition, applicability of the reflection mode depends on the thickness of the checked film, as at thicker films interference fringes tends to form. The interference fringes distort the spectrum and lead to inconclusive results. To avoid this problem, thin films of the coating were produced by removing the cemented carbide substrate by appropriate grinding, polishing, and ion thinning techniques to obtain a sample for FTIR in transmission mode (Fig. 5).
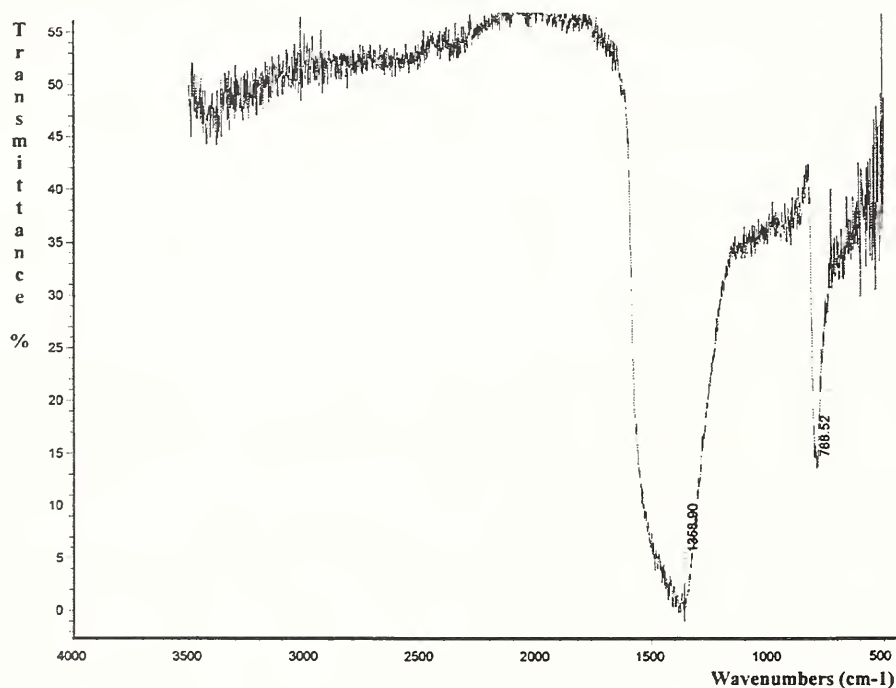
Fig. 5: FTIR spectrum, in transmission mode, of I-PACVD BN coating on a cemented carbide.

TEM sample preparation can be performed in cross section or perpendicular to the coating. In both cases it is essential to obtain an electron transparent area in the coating to perform the investigations. Samples prepared as cross section samples give information regarding the transition from substrate to coating and the morphology of the coating related to

coating thickness. Perpendicular preparation to the coating can be used to investigate a boron nitride coating on cemented carbide substrate material. An example for cross section preparation is given in Figure 3 for a cBN coatings morphology on a silicon substrate material with additional selected area diffraction experiments which reveals information regarding the crystal structure and lattice constant.

Summary

Functional cubic boron nitride coatings can offer the cutting tool industry a new paradigm for machining ferrous materials similar to the recent development in diamond technology: the tool of choice for hard to cut ferrous materials. Accordingly, three approaches to the deposition of functional cubic boron nitride coatings on cutting tools are being investigated:

- Ion beam assisted deposition which allows independent control of the ion current density, ion/atom ratio and the ion beam energy.
- Plasma assisted chemical vapor deposition that combines the necessary energetic ion bombardment with the advantages of CVD technologies and
- Unbalanced magnetron sputtering which in addition to experiments aimed at cBN films, has been used for the formation of structural templates and nanolayer superlattice.

A combination of analytical techniques, in particular FTIR, XRD, TEM-SAD and EDX have been used to determine the structure of the cubic boron nitride coatings on silicon and on cemented carbide. Further research is underway to understand and control the internal stresses in the film and in the film-substrate composite.

References

/1/ A. Inspektor, C.E. Bauer, E.J. Oles, Surf. & Coat. Technol. **68/69** (1994) 359.

/2/ E.J. Oles, A. Inspektor, C.E. Bauer, Diamond Relat. Mater. **5** (1996) 617.

/3/ National Institute of Standards and Technology -Advanced Technology Program # 70NANB5H1157.

/4/ D.J. Kester and R. Messier, J. Appl. Phys. **72** (1992) 505.

/5/ G.L. Doll, J.A. Shell, C.A. Taylor, R. Clarke, Phys. Rev. B **43** (1991) 6816.

/6/ F. Shimokawa, H. Kuwano, K. Nagai, Proc. 9th. Symp. on ISIAT, Tokyo, 1985.

/7/ B. Goranchev, K. Shmidt, K. Reichelt, Thin Solid Films **149** (1987) L77.

/8/ M. Mieno and T. Yoshida, Jpn. J. Appl. Phys. **29** (1990) L1175.

/9/ M. Murakawa, S. Watanabe, S. Miyake in R. Messier, J.T. Glass, J. Butler and R. Roy (eds.), New Diamond Science and Technologym MRS Conf. Proc., (1991) 1099.

/10/ T. Wada and N. Yamashita, J. Vac. Sci. Technol. **A10** (1992) 515.

/11/ S. Reinke, M. Kuhr, and W. Kulish, Surf. Coat. Tech. **74/75** (1995) 723.

/12/ C. Weissmantel, K. Bewilogua, D. Dietrich, H.-J. Erler, H.-J. Hinneberg, S. Klose, W. Nowick, and G. Reisse, Thin Solid Films **72** (1980) 19.

/13/ D.R. McKenzie, D. McFall, W.G. Sainty, C.A. Davis, and R.E. Collins, Diamond Relat. Mater. **2** (1993) 790.

/14/ D.L. Medlin, T.A. Friedmann, P.B. Mirkarimi, M.J. Mills, and K.F. McCarty, Phys. Rev. B **50** (1994) 7884.

/15/ Y. Lifshitz, S.R. Kasi, J.W. Rabalais, and W. Eckstein, Phys. Rev. B **41** (1990) 10468.

/16/ S. Reinke, M. Kuhr, and W. Kulisch, Diamond Relat. Mater. **3** (1994) 341.

/17/ R. Geick, C.H. Perry, Phys. Rev. **146** (1966) 543.

/18/ M. Kuhr, S. Reinke, and W. Kulish, Surf. Coat. Tech. **74/75** (1995) 806.

/19/ R. Freudenstein, S. Reinke, and W. Kulish, Diamond Relat. Mat. **6** (1997) 584.

/20/ Takanori Ichiki and Toyonobu Yoshida, Appl. Phys. Lett. **64** (1994) 851.

/21/ Takanori Ichiki and Toyonobu Yoshida, J. Appl. Phys. **33** (1994) 4385.

/22/ M. Z. Karim, D. C. Cameron, M. J. Murphy, and M. S. J. Hashmi, Surf. Coat. Tech. **49** (1991) 416.

/23/ B. Window and N. Savvides, J. Vac. Sci. Technol. A. **4(2)** 1986 196

/24/ B. Window and N. Savvides, J. Vac. Sci. Technol. A. **4(3)** (1986) 453.

/25/ W.D. Spourl, Science **273** (1996) 889.

/26/ X. Chu, A. Barnett, M.S. Wong, W.D. Sprael, Surf. Coat. Technol. **57** (1993) 13.

/27/ X. Chu, M.S. Wong, W.D. Sproul, S.L. Rohde, S.A. Barnett, J. Vac. Sci. Technol. A **10** (1992) 1604.

ROUGH DRAFT

Industrial Process Driven System Requirements for PSII Applications

Carter P. Munson
Los Alamos National Laboratory

Introduction

Plasma Source Ion Implantation (PSII) is a room temperature, plasma-based surface enhancement technology which is being commercialized through the efforts of a group of companies including General Motors, Asea Brown Boveri, Litton Electron Devices, Nano Instruments, Diversified Technologies, Ionex, PVI, Empire Hard Chrome, A.O. Smith, Harley-Davidson, Kwikset, Boeing, DuPont, and NorthStar Research, as well as Los Alamos National Laboratory and the University of Wisconsin, Madison. PSII makes use of the characteristics of ionized gas (a plasma - typically generated from N_2, NH_3, or carbon containing gases such as CH_4 or C_2H_2) surrounding a target or target assembly and high negative voltage, high current pulses, to accelerate ions into the target surface from essentially all directions, creating beneficial modifications (hardening, reduced coefficient of friction, enhanced resistance to corrosion, etc.) in the target surface. [1,2,3,4,5, 6,7,8,9,10] Beneficial surface modifications achieved can include increases in surface hardness and consequent wear life, improved corrosion resistance, and decreases in the surface coefficient of friction, and are similar to those obtained through conventional ion beam implantation [11]. The process, however, differs from conventional ion beam implantation in several important aspects (see Figure 1). Since the ions are being accelerated into the target by mechanisms within the plasma, which surrounds the target, the process is not "line-of-sight" - i.e. requiring an unobstructed path from a single ion source to the surface being treated. This allows the relatively easy treatment of multiple surfaces of a single target or simultaneous treatment of multiple targets, without the need for in-vacuum manipulation of the target assembly as would be required with conventional ion beam implantation. In addition, the average ion current to the target surface can be much larger (by an order of magnitude or more) that that typically utilized in conventional ion beam implantation (typically much less than one amp) [12,13,14], significantly reducing the required treatment time for large, complex target assemblies. This is possible because all target surfaces (up to many square meters area) are being treated simultaneously, which minimizes local surface heating effects.

A number of issues are critical to the successful design and operation of a commercial PSII system. These include overall vacuum system design, plasma source requirements and plasma-target interaction considerations, pulsed, high voltage sub-system (typically referred to as modulator) requirements, and target requirements and limitations. Critical system components are outlined in Figure 2, and overall system design issues will be briefly covered in the following sections of this paper.

ROUGH DRAFT

Vacuum System Requirements

Neutral gas working pressures for the PSII process are typically in the 0.2 to 1.0 milli-Torr range, with plasma densities in the mid 10^8 to mid 10^9 /cm^3 range, so that system pumping requirements driven directly by the implantation process are relatively minor. The basic pumping requirements are determined by the need to keep the overall plasma and neutral impurity concentrations relatively low (no more than a few percent of impurities in both the neutral gas and plasma ion concentrations). Impurities in the system can be generated by vacuum system leaks, outgassing from the chamber walls or from the implantation target, by sputtering of surface contaminants from the implantation target, or by impurities in the working gas introduced into the system. For reasonably clean targets with minimal target outgassing, even a relatively large target chamber (8-10 m^3) can be served by <1000 l/s pumping speed at the 0.2 to 1.0 milli-Torr pressure level. (Which can be provided by a single turbo-molecular or diffusion pump and roughing system.) The much more demanding system pumping requirements are driven by the need to keep initial pump down times relatively short, so that overall process cycle times will be as short as practical. One potential method of minimizing cycle time for batch processing in vacuum systems is to utilize multiple load-locks, in which targets are loaded through isolated chambers which are attached to the main vacuum chamber by a gate valve, or other isolation device. A target is placed in one load-lock, which is pumped to a pressure approaching the system base pressure before it is opened to the main chamber and the target transferred. While this first target is being processed, a second target is loaded into the second load-lock and pumped down. When processing of the first target has been completed, it is removed through its load-lock, and the second target is then inserted. (The system could also be configured for unidirectional motion, in which one load-lock is consistently used to load targets, while the second is used to unload processed targets.) This procedure works well for relatively small targets, but can become difficult and expensive to implement for large, heavy targets - which would require large isolation valves, and substantial in-vacuum manipulation capabilities for target transfer. Independent of the use of a load-lock, initial pump-down pumping speed requirements are determined by the target chamber size, effective outgassing rate of the chamber and target, and the desired total time to reach a given base pressure. In the absence of significant outgassing, the required pump-down time may be simply estimated as:

$t \approx (Vol/s1)\ln(P_{initial}/P_{rough}) + (Vol/s2)\ln(P_{rough}/P_{final})$, where t is the pumping time

required, Vol is the system volume, s1 is the high-pressure, or roughing system pumping speed, s2 is the high vacuum system pumping speed, and $P_{initial}$, P_{rough}, and P_{final} are the initial, intermediate roughing, and final desired system pressures. Since typical roughing pressures (at which the system can be switched to a high vacuum pump such as a turbomolecular or diffusion pump) are usually in the milli-Torr range, the total pump down time is often dominated by the size of the roughing pump system. (Outgassing of the system, or targets inserted into the system, can dramatically increase the required pumping time.) Shorter total pump down times may be achieved by the use of mechanical booster pumps of various designs. These are relatively high speed pumps which operate in an

intermediate pressure range (from 10 Torr or more down to the milli-Torr range). Typically, a system is pumped to a pressure well below the desired working pressure before the process gas is introduced, in order to minimize impurities.

An additional consideration for the vacuum system is the requirement for shielding the x-rays produced by secondary electrons impacting the chamber wall. For multi-cubic-meter systems running at no more than ~70 kV, the chamber walls will in general be thick enough (1/2 inch or so) to sufficiently shield the x-rays. Additional shielding may be required around ports and other thinner portions of the system, or if voltages above the 70 kV level are required for a particular application.

Plasma Sheath Issues and Plasma Source Requirements

The PSII process relies intrinsically on the characteristics of the plasma which surrounds the target, and the transient plasma sheath which is established during the process. The application of a negative high voltage pulse creates the transient plasma sheath around the target, which accelerates ions into the target surface. The transient plasma sheath has been described in great detail in numerous papers (/15ref sheath papers/), and the general sequence of sheath evolution is described as follows, and illustrated in Figure 3.

The dimensions of the sheath are determined primarily by the initial plasma density, the voltage applied to the target, and the duration of the voltage pulse. The initial configuration (Figure 3a) has a nearly uniform plasma surrounding the target. Due to the great difference in mass between the electrons and ions in the plasma, the electrons move rapidly away from the target during the early portion of the voltage pulse, exposing plasma ions (Fig. 3b). This initial "matrix sheath" dimension $D_{initial}$ is given by $D_{initial} = (2\varepsilon_0 V/en)^{1/2}$ for planar geometry, where ε_0 is the permittivity of free space ($= 8.9 \times 10^{-12}$ farad/meter), V is the negative voltage applied to the target, e is the ion charge ($= 1.6 \times 10^{-19}$ coulomb for singly charged ions), and n is the plasma density per cubic meter. The electric field now present in the sheath region accelerates ions into the target, as depicted in Fig. 3c. As ions are implanted into the target and lost from the sheath, the sheath edge recedes from the target, as shown in Fig. 3d. The sheath thickness increases during this time as $D(t) = D_{initial}((2/3)\omega_{pi} t + 1)^{1/3}$, where ω_{pi} is the ion plasma frequency ($\omega_{pi} = (ne^2/\varepsilon_0 M)^{1/2}$), n, e, and ε_0 are as defined above, and M is the ion mass in kg. Eventually, the sheath edge will extend to the vacuum chamber wall, or arcing will occur, limiting the useful pulse duration for implantation. Typical pulse widths range from several microseconds to almost 100 microseconds. A useful discussion of plasma sheath behavior in various geometries is given by Scheuer, et al. /15,16/. From the equations describing the sheath evolution, it is clear that higher initial plasma densities result in smaller sheaths. Sheath dimensions which are small compared to the scale size of important target features result in more uniform implantation of the critical target surfaces.

A number of possible mechanisms exist for generation of the initial plasma for the PSII process. These include both capacitive and inductive RF sources, electron emissive sources, microwave sources, electron cyclotron resonance (ECR) sources, magnetron sources, vacuum arc sources, etc. /17,18 ref plasma source review/. The optimal source for PSII applications would be capable of generating a uniform, low to moderate density plasma ($\sim 10^{14}$- 10^{17} /m^3) around the target, simple to implement and operate, and robust and reliable. Inductively coupled RF sources (both steady-state and pulsed) are capable of generating high densities within the source (up to the high 10^{17}/m^3 range) /ref tuszewski/, but generally do not produce this density level uniformly throughout a large chamber, as is required for PSII. Vacuum arc sources are capable of producing plasmas from metallic species which would otherwise be difficult to obtain from gaseous sources (organo-metallic compounds with high vapor pressures), but would require multiple sources to provide a reasonable level of plasma uniformity around a single target. Production of a uniform plasma throughout a complex target assembly is unlikely for a source of this type. ECR, magnetron, and other sources require ancillary magnetic fields for their operation, and are therefore more complicated to implement. For operation with gaseous implantation species (primarily nitrogen and carbon - from hydrocarbon gases such as methane) in relatively large systems (many cubic meters volume), one of the most useful sources is a capacatively coupled RF plasma source. For this type of source operating at the ~ 1000 Watt level, a single, simple antenna is capable of generating a relatively uniform plasma in the 10^{14}-10^{15}/m^3 range around targets in moderate volume systems (~ 5-10 m^3) from a neutral gas fill pressure of below 1 milli-Torr.

Modulator Requirements

The key component in a successful PSII system is the high voltage, high current switching sub-system - the "modulator". This system must be able to repetitively switch large currents (hundreds to several thousand Amperes), with switching times of a few microseconds, in order to allow implantation of large surface area targets with good quality and uniformity over the target surface. During the quasi-steady state portion of the plasma sheath evolution, the total current drawn by the target can be estimated from

$I_t = A(\gamma + 1)(4\varepsilon_0/9)(2e/M)^{1/2} V^{3/2}/D^2 (t)$, where V is the voltage applied to the target,

A is the target surface area, and γ is the surface secondary electron emission coefficient (which can be as large as 10 to 20), and M, e, and D(t) are as defined above. The modulator must supply at least this current in order to maintain the desired implantation voltage on the target (typically from 20 to 70 kV, or more). In order to keep the initial voltage rise period relatively short (no more than a few microseconds), the modulator must also be capable of transiently supplying an even larger current, since the target initially appears as a capacitive load during the time in which the ion matrix sheath is being established. The high voltage and high current levels required for large targets result in instantaneous switched power requirements which are in the megawatt range, and average power requirements from tens to hundreds of kilowatts - depending on the pulse width and pulse repetition rate of the system. Pulse widths of 5-30 microseconds are common in experimental systems, with repetition rates of up to several thousand pulses per second, and implantation times of several hours. Since it is desirable to keep the total

implantation time as short as possible, there is a very strong incentive to utilize very high plasma densities (high current demands on the modulator), and high repetition rates. Primary constraints on these parameters are the total power available for the modulator, and the thermal dissipation characteristics of the target and support assembly. The modulator must also be well controlled and protected in order to minimize the impact of arcing. Arcs can present a significantly lower impedance to the modulator than the normal target, and can lead to severe target surface damage if uncontrolled. Fast arc detection circuitry must be an integral part of the modulator, as well as controls which temporarily blank the modulator output in order to quench any arcs which do occur. Successful modulator designs have been based on high voltage, high current vacuum tube circuits of various types, solid state devices (Insulated Gate Bipolar Transistor packages), and pulse forming networks (PFNs).

Target and Mounting Considerations

The success of the PSII process depends not only on the characteristics of the system used in the process, but also on the details of the target(s) and the mounting structure utilized within the process chamber. The target surface material, geometry, and failure mechanisms must be suitable for significant improvement through the PSII process. Surface materials which have been successfully improved through the PSII process include chromium and high chromium alloy steels (with nitrogen implantation), tungsten and aluminum alloys (with carbon implantation), titanium alloys (with nitrogen implantation), more common iron alloys (with nitrogen implantation), and others. Given a suitable target surface material, the target failure mechanism and location must be evaluated. Mild abrasive wear failure is a prime candidate for PSII enhancement, if the surface material and failure location are appropriate. Failure areas must be accessible to the plasma ions, with accessibility determined by the location of the failure area and the relative scale size of the target surface and the plasma sheath. Areas of the target which are shielded from the implanting ions either by portions of that target, or by shielding from adjacent targets in a target assembly will not be suitably enhanced. Portions of a target surface which are good (and poor) candidates for PSII enhancement are illustrated in Figure 4. Another target consideration involves thermal management. Since the PSII process deposits energy into the target surface, targets must be cooled if they are to be maintained at essentially room temperature. High aspect ratio targets (e.g. long and thin) are therefore poor candidates for treatment unless they can be cooled through a central channel. Some targets with a relatively small surface to volume ratio (large industrial stamping dies, for example) may require no active cooling, as the thermal mass of the target is sufficient to keep the temperature excursion relatively small during the PSII processing.

Summary

PSII processing has been proven effective in enhancing the characteristics and service life of a number of industrial components. Based on these results and cost estimates for commercial scale processing of components, the PSII process is being commercialized through the efforts of a group of companies including General Motors, Asea Brown

ROUGH DRAFT

Boveri, Litton Electron Devices, Nano Instruments, Diversified Technologies, Ionex, PVI, Empire Hard Chrome, A.O. Smith, Harley-Davidson, Kwikset, Boeing, DuPont, and NorthStar Research, as well as Los Alamos National Laboratory and the University of Wisconsin, Madison. Commercialization has been achieved on a small scale through a facility being operated at Empire Hard Chrome, in Chicago Il., and additional small and large scale facilities are being planned. Successful implementation of the PSII techniques requires careful consideration of the basic plasma parameters, target characteristics and target support, modulator, and the interactions of the plasma-target-modulator system.

Figure Captions

Figure 1. Schematic comparison of PSII and conventional ion beam implantation. PSII utilizes the plasma sheath to accelerate ions into the target (or multiple targets) from all directions. Conventional, accelerator based ion implantation is a line-of-sight process, which requires in-vacuum manipulation of a target to implant complex surfaces.

Figure 2. Block diagram of a typical PSII system. Major system components include the vacuum chamber, pumping system, high voltage DC power supply, high voltage pulser or modulator, cooling system, target support assembly, plasma generation system, and working gas input system.

Figure 3. Plasma sheath temporal behavior for a planar target. 3(a) shows the initial configuration with nearly uniform plasma surrounding the target. 3(b) indicates the formation of the ion matrix after electrons have been excluded from the region close to the target. 3(c) indicates ions accelerating through the sheath region into the target surface. 3(d) shows the expanded sheath late in the voltage pulse. 3(e) indicates the relative times of a-d during the voltage pulse.

Figure 4. Schematic representation of portions of a complex target which are good (or poor) candidates for PSII enhancement with a given plasma sheath position.

1 R. Adler, Nucl. Instr. Method. B6 (1985) 123.

2 J. R. Conrad and K. Sridharan (eds.), *First International Workshop on Plasma-Based Ion Implantation*, Journal of Vacuum Science and Technology B, Vol. 12, No. 2, 1994.

3 G. A. Collins (ed.), *Proceedings of the Second International Workshop on Plasma-Based Ion Implantation*, Sydney, Australia, February 12-15, 1995, Surface Coatings and Technology

4 J.R. Conrad, *J. Appied Physics* **62**, (1987) 777.

5 J. R. Conrad, J.L. Radtke, R.A. Dodd. F.J. Worzla, N.C. Tran, *J. Applied Physics* **62**, (1987) 4591.

6 J.R. Conrad, U.S. Patent No. 4 764 394 (August 16, 1988).

7 J.R. Conrad, *Mater. Sci. Eng. A*, 116 (1989) 197.

8 D.J. Rej and R.B. Alexanander, J. *Vac. Sci. Technol.* B12 (1994) 2380.

9 D. J. Rej, in M. Nastasi and J. W. Mayer (eds.), <u>Ion-Solid</u> <u>Interactions:</u> <u>Fundamentals</u> <u>and</u> <u>Applications</u>, Cambridge University Press.

10 C.P. Munson, et al., *Surface and Coatings Technology* **84** (1996) 528.

11 R. B. Alexander - beam-PSII comparison

12 S.B.J. Charter, L.R. Thompson, and G. Dearnaley, *Thin Solid Films,* **84** (1981) 355.

13 F.A. Smidt and D.B. Sartwell, *Nuclear Instr.s and Meth. Phys. Res.* **B6** (1985) 70.

14 J.R. Treglio, A.J. Perry, and R.J. Stinner, *Proc. Of the 8th International Conf. On Surface Modification of Metals by Ion Beams* (Spetember 1993, Kanazawa, Japan)

15 J.T. Scheuer, M. Shamim, and J.R. Conrad, *J. Applied Physics* **67**, (1990) 1241.

16 Addl sheath refs

17 Lieberman and Lichtenberg, "Principles of Plasma Discharges and Materials Processing", John Wiley and Sons, Inc., 1994.

18 F.F. Chen, *IEEE Transactions on Plasma Science,* **23 (1)** (1995) 20.

plasma surrounding targets
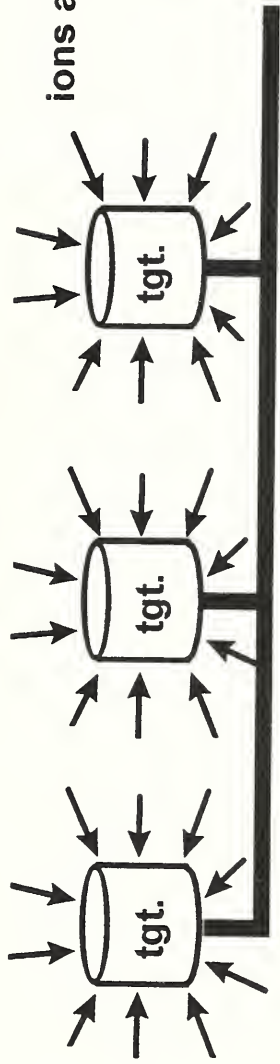
ions accelerated into targets

tgt.

tgt.

tgt.

HV pulses and cooling
connection to targets
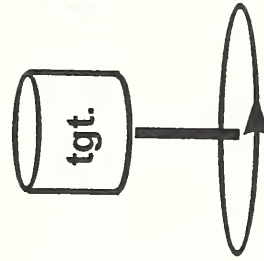
iob beam from accelerator
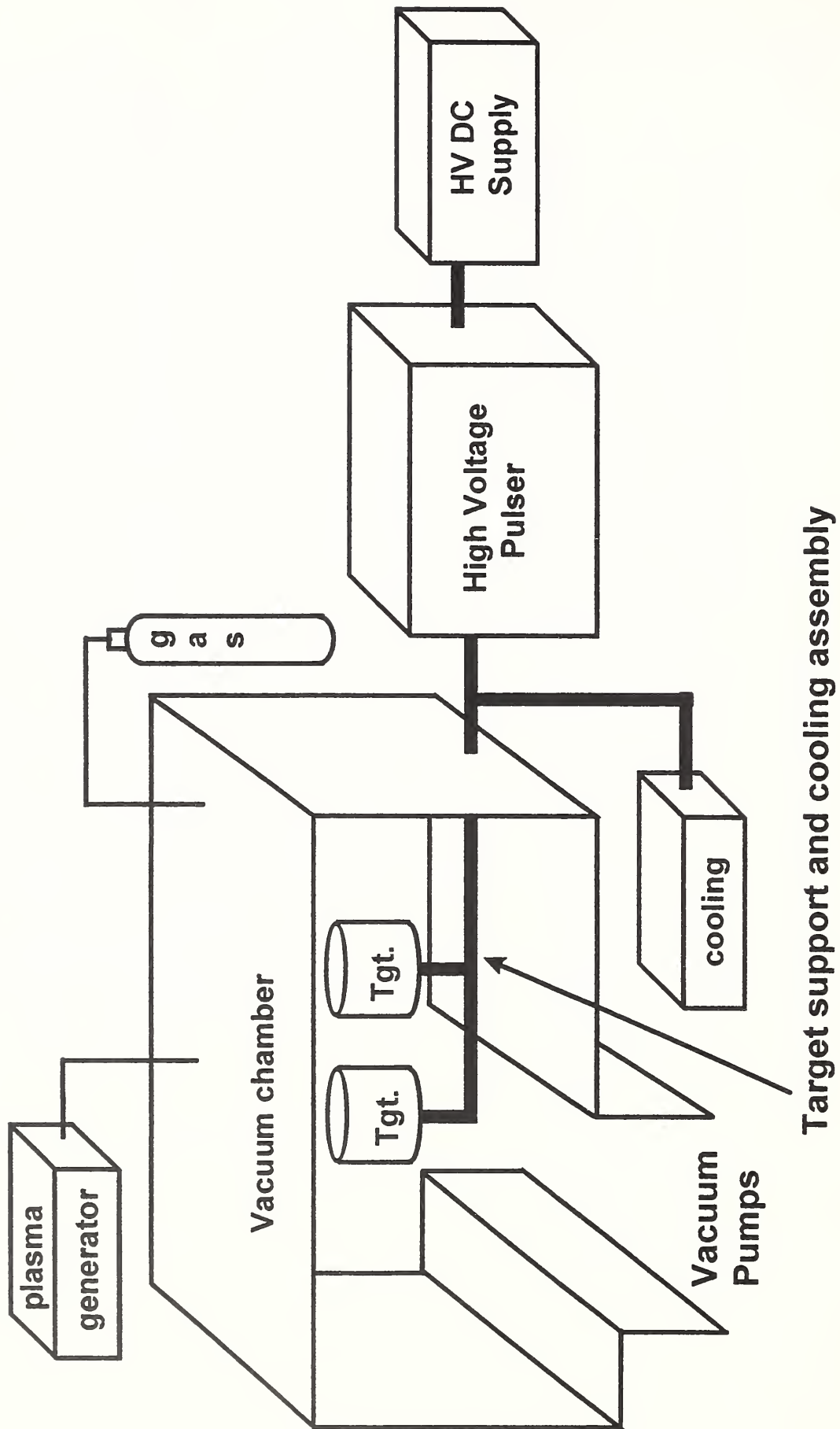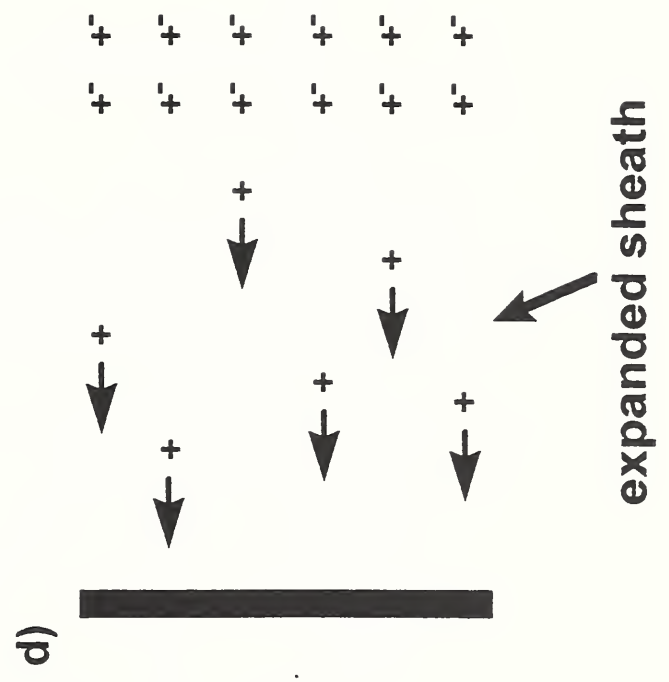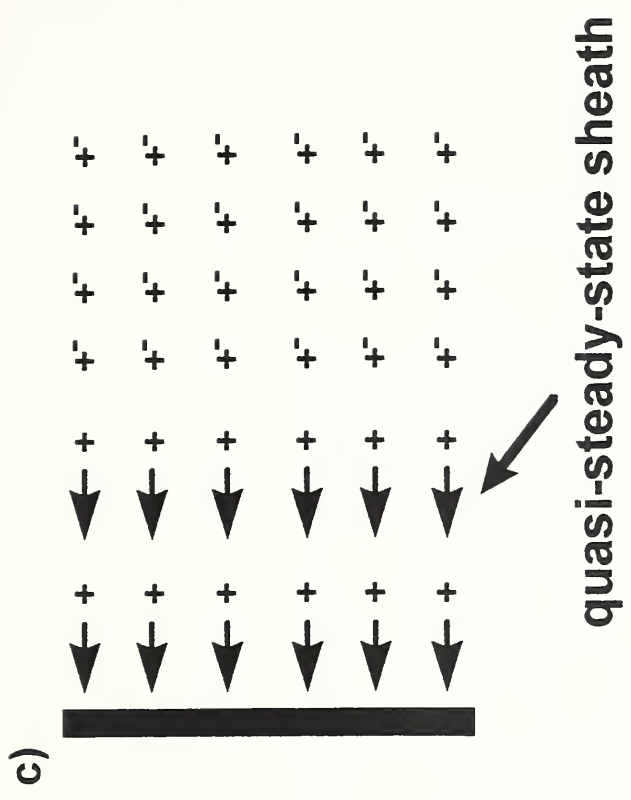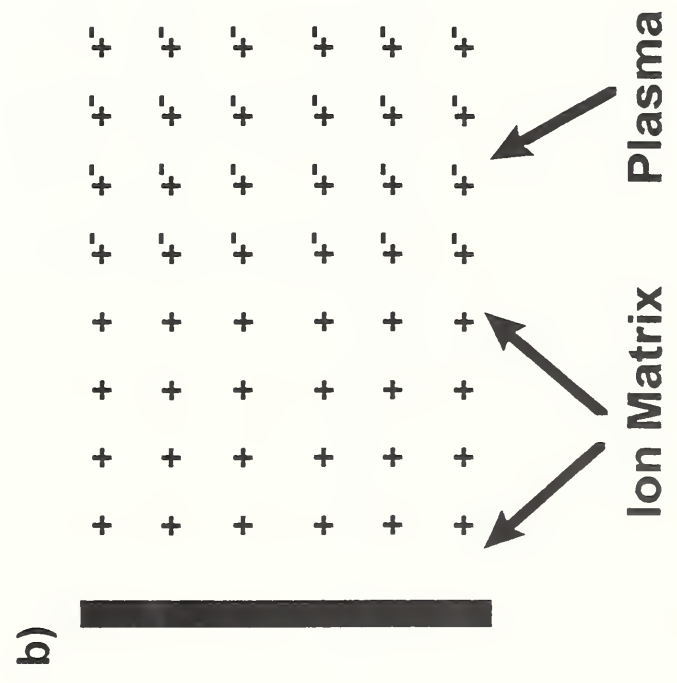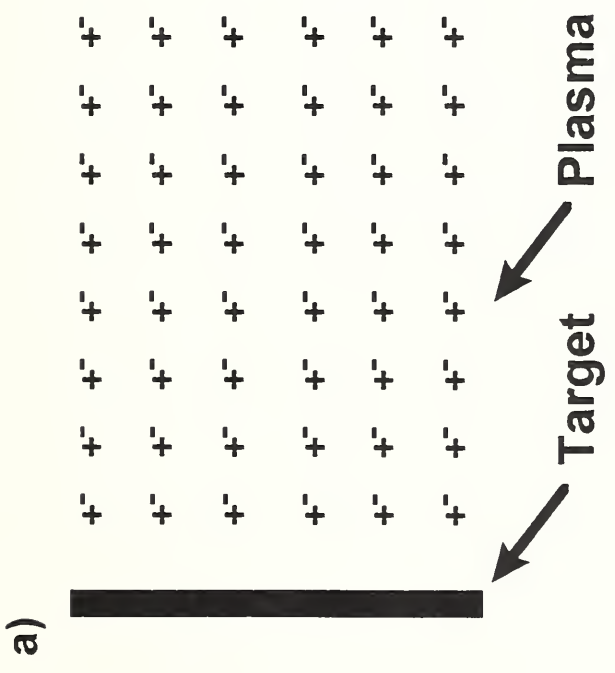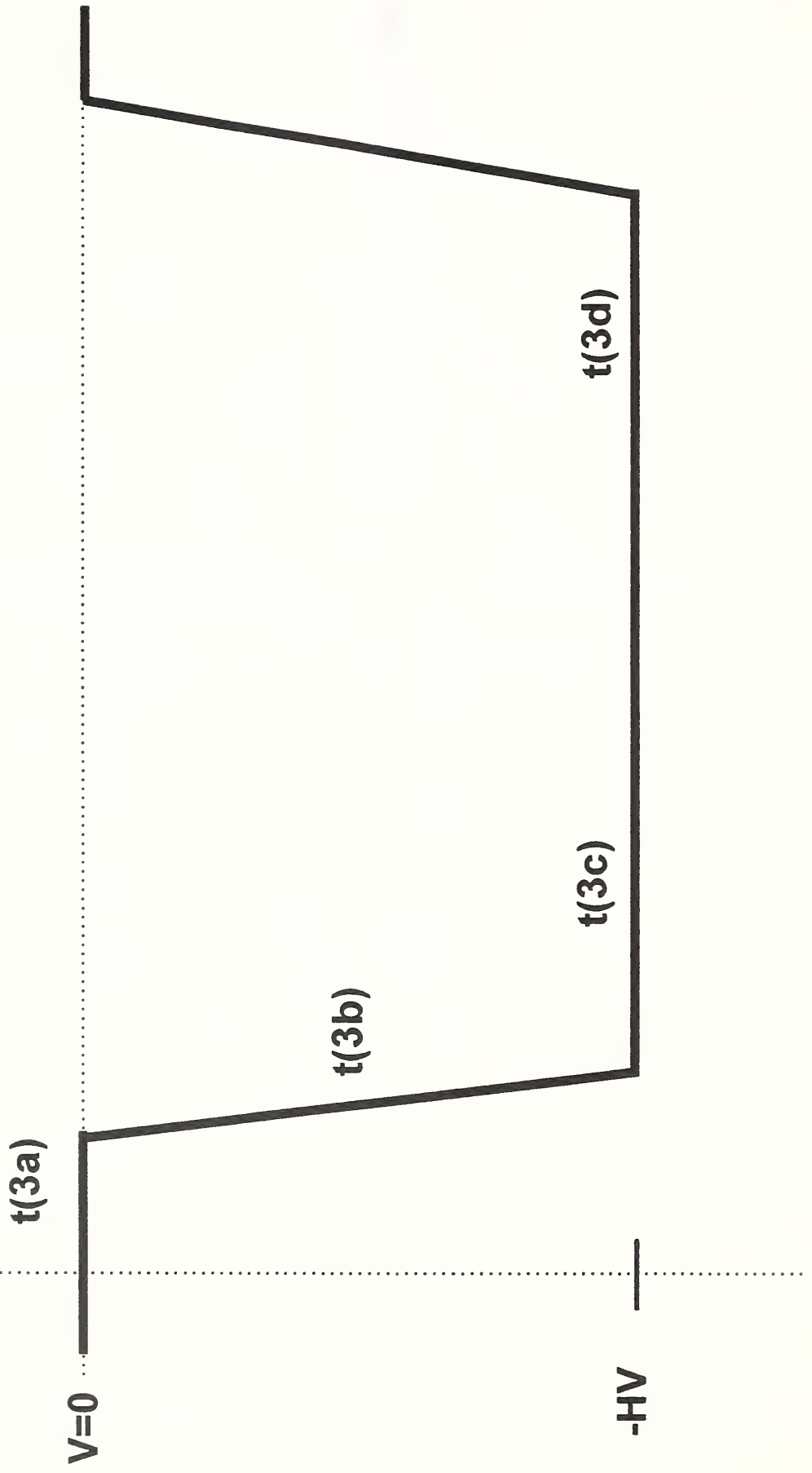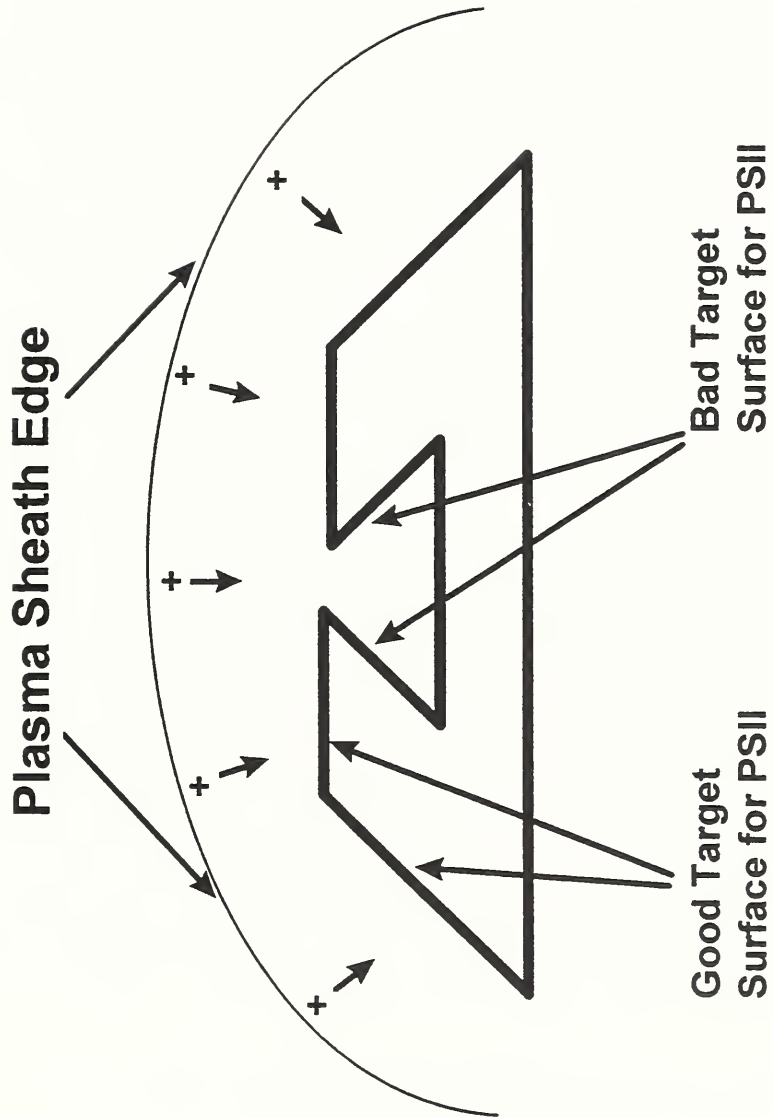
tgt.

in-vacuum manipulator

HV DC Supply

High Voltage Pulser

gas

Target support and cooling assembly

cooling

Tgt.

Tgt.

Vacuum chamber

Vacuum Pumps

plasma generator

306

a)

Plasma

Target

b)

Plasma

Ion Matrix

c)

quasi-steady-state sheath

d)

expanded sheath

307

e)

V=0

t(3a)

t(3b)

t(3c)

t(3d)

-HV

308

Plasma Sheath Edge

Bad Target Surface for PSII

Good Target Surface for PSII

Plasma-Based Ion Implantation:
Surface Treatments For Automobile Manufacturing

J.V. Mantese
General Motors Research and Development Center

Plasma-based ion implantation is an inherently batch processing technique for simultaneously treating large numbers of parts to enhance their chemical and tribological surface properties. This room temperature process is applicable to: ferrous materials (like steels and cast iron), lightweight metals (such as aluminum, magnesium and copper), and even polymeric targets (plastics).

In its simplest form, see Fig. 1, plasma-based ion implantation requires an enclosed chamber with a modest vacuum (~ 0.1 mtorr) similar to ones used in other common industrial processes such as plasma nitriding and vacuum brazing. In the chamber, electrons are stripped from atoms such as nitrogen or carbon, creating positive ions of these chemical species. By applying a high negative voltage to the target surface, the ions are accelerated and directed toward target materials, drawing the ions into the workpiece. This process is usually a pulsed implantation process, which permits additional positive ions to form and enables them to diffuse throughout the vacuum chamber. Surface-enhanced material surfaces are obtained primarily through the formation of nitrides, oxides, or carbides.

Plasma-based ion implantation is an outgrowth of beam-line implantation, with one important difference; it is inherently a batch process. Using plasma implantation, one can simultaneously process multiple parts that are electrically connected to the target stage and achieve the same performance as its line-of-sight counterpart. Plasma-based ion implantation is also often called by numerous other names, including: plasma source ion implantation (PSII), plasma immersion ion implantation (PIII or PI^3), or plasma ion implantation (PII). At General Motors, the technique is referred to as the IONCLAD process.

Three distinct implantation processes have emerged as commercially viable techniques for treating the surfaces of materials: (1) The implantation of nitrogen into chrome and ferrous materials. (2) The formation of highly adherent diamond like carbon (DLC). (3) Plasma immersion ion plating (PIIP processes) for the low temperature formation of highly adherent carbides, nitrides, and borides. We discuss each of these processes in detail.

The implantation of nitrogen into chrome and ferrous materials:

Along with plasma-based ion implantation, plasma and thermal nitriding are batch processes, however, plasma-based implantation offers several advantages. First, plasma implantation is a room-temperature process that nitrides material surfaces without the need for heating. Thus, this process can nitride without loss of steel tempering. In addition, plasma-based implantation avoids the brittle epsilon phase (also called the white phase) which sometimes forms at higher temperatures required by other processes.

Second, plasma implantation creates dispersed regions of ultra-hard nitrides that cannot delaminate from the surface and are highly resistant to oxidation.

Third, plasma implantation can be used to form carbides, simply by substituting a carbon-bearing gas such as methane for nitrogen as the working gas.

The surface hardened layer is typically about one micrometer thick. Although case depths of up to 100 micrometers or more can be achieved if thermally assisted processes are allowed (e.g., 200-300 °C) with the material.

Typically plasma processing has improved the fatigue and/or wear life from two to ten times for materials such as S7, D2, M2, S5, and chrome plate. Chrome plated materials are especially susceptible to improvement by implantation with 10 X improvement in life routinely observed in a variety of applications.

Plasma-based nitrogen ion implantation has been scaled for some applications such as the large chrome plated stamping die shown in Fig. 2. A variety of experimental and commercial units now exist which range in size from 1 m^3 to cylindrical chambers nearly 2 m in diameter and 5 m in length.

The Deposition of Highly Adherent Diamond-Like Carbon (DLC):

Diamond like carbon (DLC) is a highly ordered structure of carbon that has a combination of tetragonal (diamond) and graphitic bonds. By adjusting the diamond-to-graphite concentration, DLC can be made extremely abrasive, suitable for cutting tools, or highly lubricious and wear-resistant, useful as a durable solid lubricant.

An advanced form of plasma-based ion implantation, see Fig. 3, has recently been developed for producing DLC coatings with excellent adherence to everything from plastics and glass to steels and aluminum. In this process, carbon atoms are implanted into the surface of the base material, changing its chemistry and forming a surface to which DLC readily adheres. Because the plasma-based DLC deposition process occurs in the same chamber as the implantation process there is little added cost associated with the coating process.

This process, too, has been scaled as the 1000 piston batch shown is Fig. 4 demonstrates.

Plasma Immersion Ion Plating (PIIP):

The above DLC deposition process is actually a specific form of a plasma immersion ion plating (or PIIP) process. In PIIP, a working gas is chosen to form a specific carbide, nitride, or even boride. Like the DLC process, PIIP is a four step process, requiring: (1) A cleaning step to remove native oxides and contaminates. (2) An implantation step at 20 - 50 kV to chemically prepare the surface to accept the deposited layer and form a graded interface. (3) A short duration cleaning step to remove an undesirable graphitic layer. (4) The deposition phase where the carbide, nitride, or boride is formed at essentially room temperature.

The PIIP process offers tremendous potential in forming highly adherent coatings at near room temperature for production scale quantities, thus, eliminating concerns related to part warpage, and loss of temper associated with such techniques at the vanadium diffusion process.
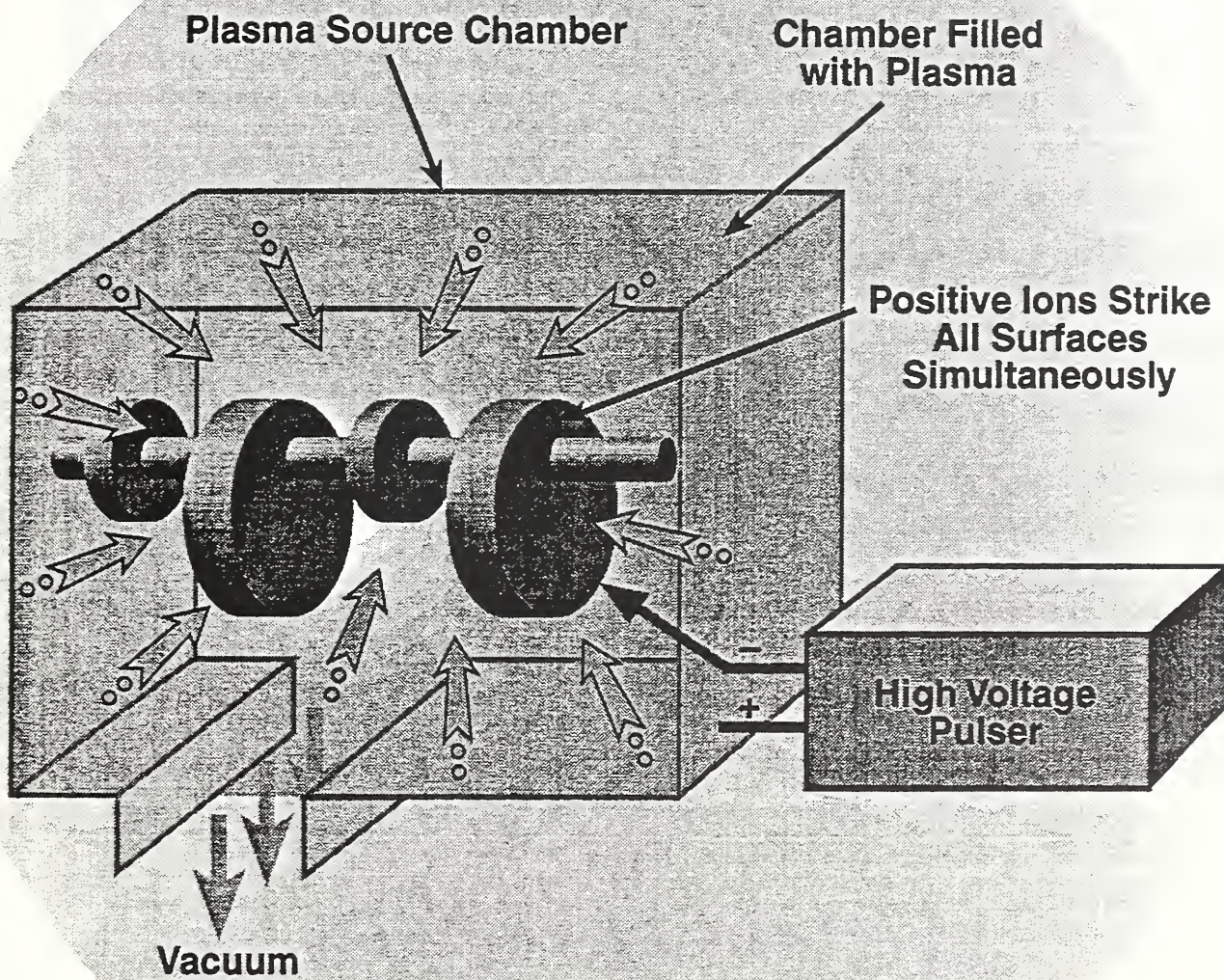
Figure 1: Schematic of the plasma-based ion implantation technology.
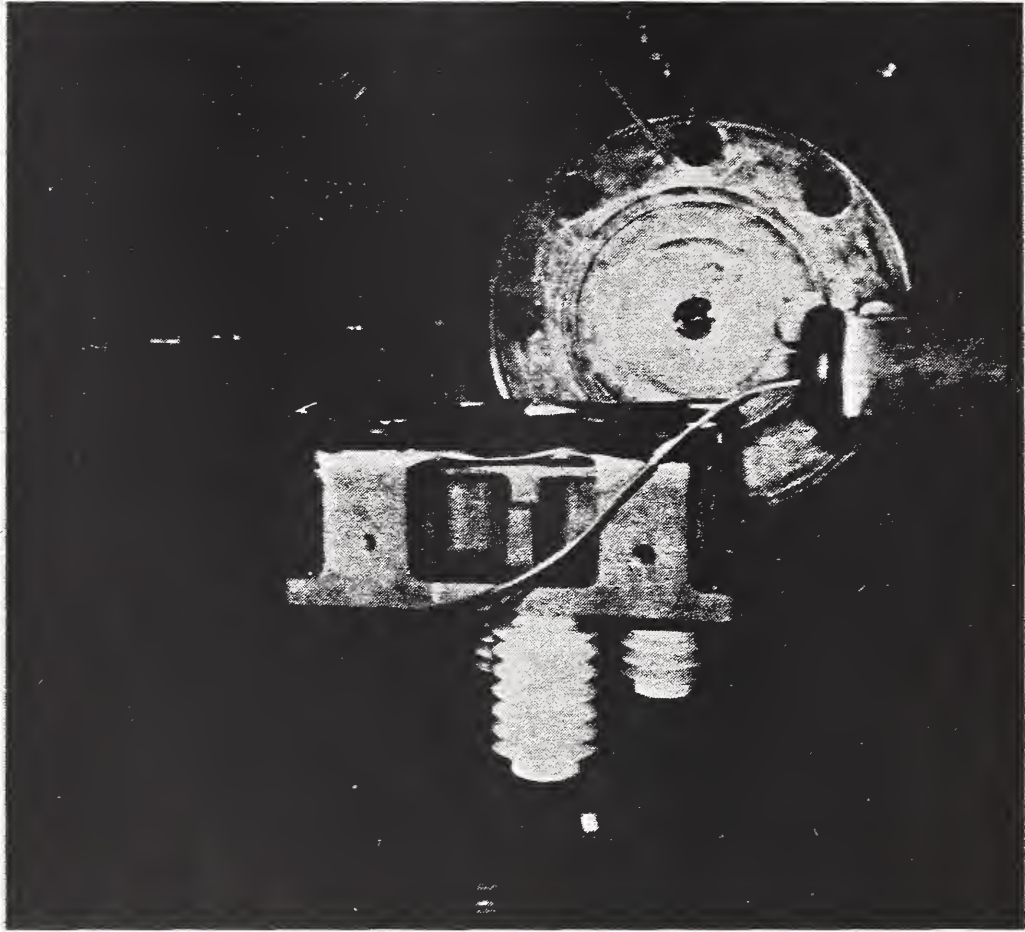
Figure 2: Photograph of a large chrome plated stamping die being readied for implantation.
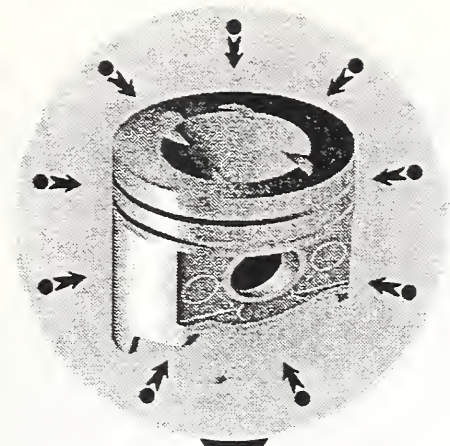
Sputter-clean native oxide
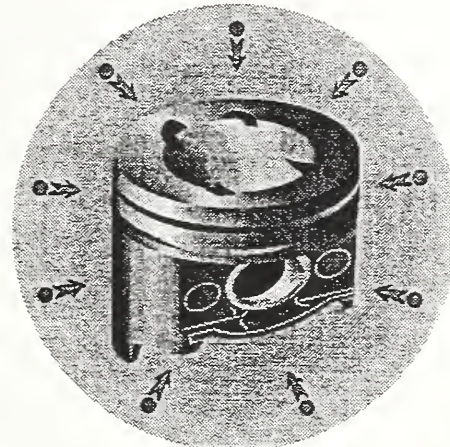
Formation of bonding layer
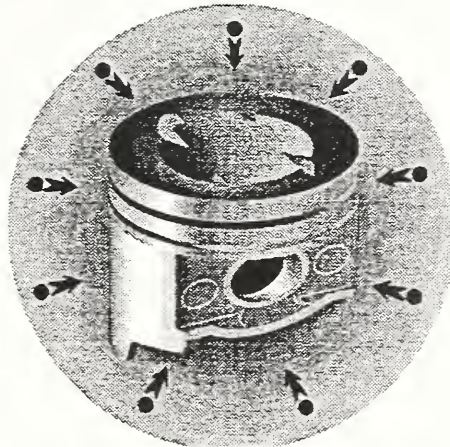
Removal of graphitic phase
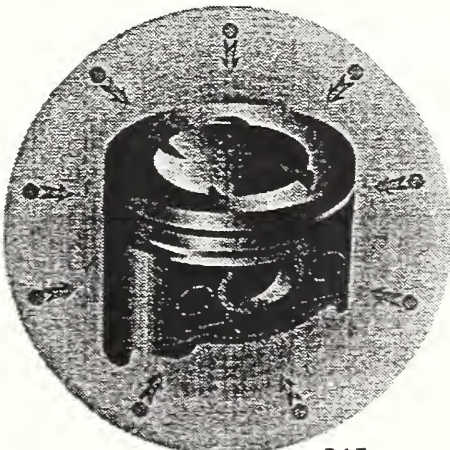
Deposition of DLC

Figure 3: Schematic of the diamond like carbon (DLC) deposition process.
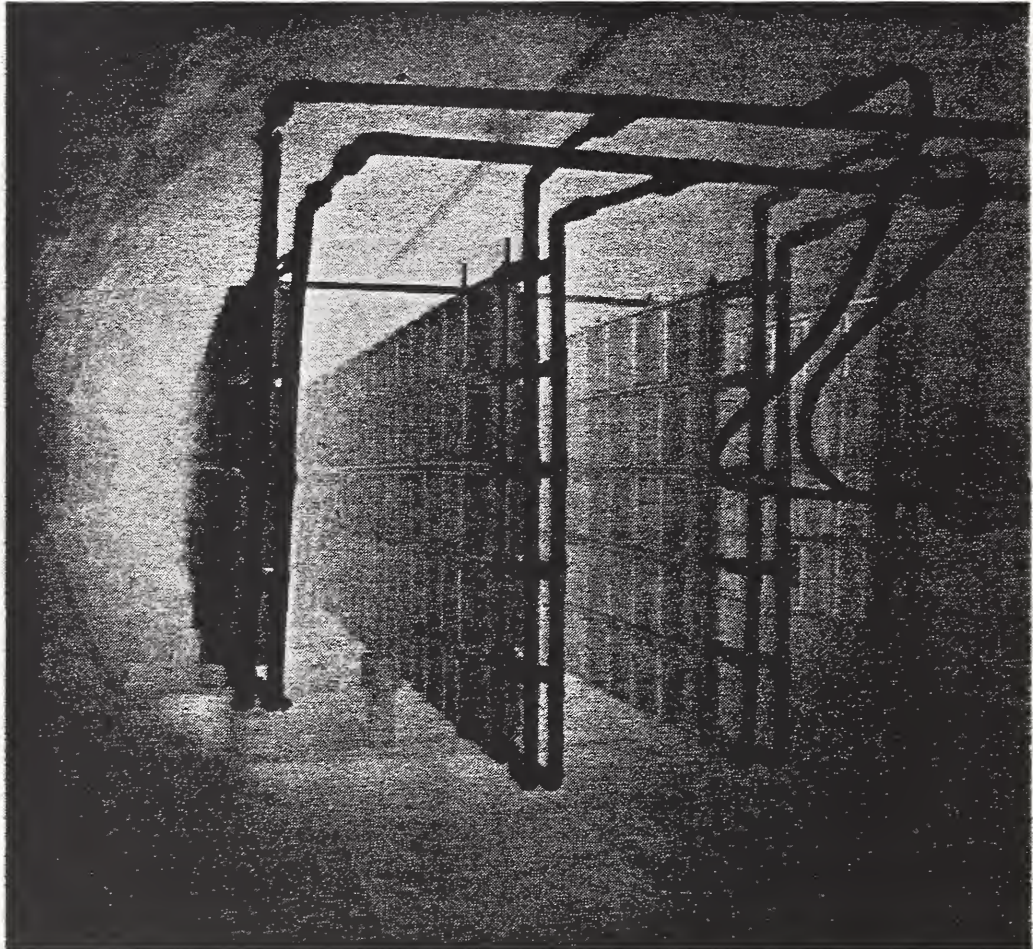
Figure 4: Photograph of the DLC process being run concurrently on 1000 pistons.

THE COMMERCIALIZATION OF PSII INTO MULTIPLE-USER MARKETS

M.C. Dudzik
Automotive & Transportation Programs
Environmental Research Institute of Michigan

1. INTRODUCTION

The commercialization of technology from laboratory to marketplace has long been recognized as a nontrivial process. The Plasma Source Ion Implantation (PSII) technology is a low-temperature method of producing hard, wear-resistant surfaces on metals, plastics, and thin films. PSII differs from conventional ion implantation processes in that it allows for the simultaneous implantation of large areas, without the need for complex part rotation. The movement of this technology from the R&D laboratory environment to multiple-user markets as a cost-effective process enabler is the focus of this paper.

The advantage of PSII is that it provides a demonstrated low-temperature means of producing hard wear-surfaces with minimal distortion or thermal degradation of the micro structure. Furthermore, the PSII process also allows for hybrid processing steps so that additional wear surface treatments, such as diamond-like coating (DLC) or ion beam assisted deposition (IBAD), can be allied in conjunction with the PSII process. The PSII use of gaseous implant species allows for surface enhancements through the formation of nitrides, carbides, and other amorphous surface layers. The implant species further provides environmentally friendly processing of materials without incurring the cost of disposal for caustic baths.

PSII test results have demonstrated 10 times improvement in wear life for tooling and stamping applications. These results translate to decreased costs of manufacturing by reducing downtime, improving production quality, and allowing greater throughput. Furthermore, PSII can allow the use of dissimilar materials in applications where wear surfaces in contact could produce soldering or galling of these materials. This can allow developers to use lightweight materials in new applications. Additionally, the modified surface in conjunction with hybrid coatings provides potential for anticorrosion applications and general surface hardening.

To move effectively from the laboratory to the marketplace, PSII developers have had to demonstrate three important parameters: (1) the performance of PSII wear surface treatments for product applications, (2) the cost effectiveness of the process for both single item and bulk material processing, and (3) the program had to demonstrate that PSII was scaleable to large manufacturing volume and component implantation machines.

2. ATP JV PROGRAM

Under the NIST ATP Motor Vehicle Manufacturing Technology Program, 95-02, the PSII Joint Venture (JV) began a three year $15.5 million joint-funded effort to advance the development of PSII beyond high-risk scientific barriers. The program goals include a 10 times increase in tool and die life, a 10 percent reduction in power train manufacturing costs through the use of lightweight materials and alloys, a 3 times improvement in corrosion resistant brass, and the elimination of 25 percent of the hazardous effluent generated by chrome-plating processes.

While the initial goal was to focus early PSII developments on automotive components and the automotive manufacturing process, it was clear that additional applications were also emerging in the manufacturing of consumer products, medical devices, military and aerospace components, and electronic displays.

The first key strategic decision in this program was to provide for true commercialization of PSII technology by building a **vertically integrated team** composed of R&D organizations, component developers, systems integrators, and end users. A vertically integrated team allowed critical barriers of the integrated product development process to be targeted fully, without incurring the derailing effects of internal competition among JV members. Furthermore, the use of a vertically integrated team allowed for the inclusion of multiple end users from different businesses. Multiple end users provided significant "voice of the customer" influence for recipe and machine development scenarios to ensure both technical merit and marketplace cost effectiveness. The JV also added new members to the team throughout the three-year program as new-end user applications were identified. In this situation, the legal flexibility of the Cooperative Agreement between the JV membership allowed for the addition of new members as necessary. The team is shown in Table 1.

Table 1. The PSII Team

General Motors	Diversified Technologies
DuPont	Ionex
Boeing	PVI
Harley-Davidson	Empire Hard Chrome
A.O. Smith	Kwikset
Asea Brown Boveri	Los Alamos National Laboratory
Litton	University of Wisconsin
NANO Instruments	ERIM
4th State	

The second key strategic decision in this program was to intrinsically focus attention in three areas: management, technical, and commercialization. To achieve this strategic goal, the program was structured to include **three oversight committees** as shown in Figure 1.

- The Management Committee is composed of senior management from commercial firms within the JV and meets quarterly. The purpose of this committee is to provide senior management oversight to ensure progress against technical, fiscal, and commercialization goals. In addition, this committee provides for an internal steering process to redirect activities or budgets as required.

- The Technical Review Committee is composed of senior technical staff from program member organization. The committee meets for two days each quarter to discuss and review progress by each member and to openly share and discuss developments. By this process, each member is allowed to develop expertise in all areas of the program in order to further drive commercialization.

- The Commercialization Committee is composed of marketing/business development or application engineers from each organization member. This committee meets quarterly to identify the commercialization issues to be integrated into technical and management decisions.

The use of the three oversight committees synergistically allows for the advancement of the commercialization process of PSII, since practical commercialization of the technology is the first and foremost goal of this program.

```
                           ┌──────────┐
                           │   NIST   │
                           └─────┬────┘
                                 │
                      ┌──────────┴───────────┐
                      │ PSII Joint Venture Team │
                      └──────────┬───────────┘
                                 │
                      ┌──────────┴───────────┐
                      │  Management Review   │
                      │     Committee        │
                      └──────────┬───────────┘
                                 │
  ┌──────────────────┐           │           ┌──────────────────┐
  │ Commercialization │──────────┼───────────│ Technical Review  │
  │    Committee      │          │           │ Steering Committee│
  └──────────────────┘          │           └──────────────────┘
                                 │
                      ┌──────────┴───────────┐
                      │ General Manager/Administrator │
                      └──────────┬───────────┘
                                 │
                ┌────────────────┴────────────────┐
    ┌───────────────────┐            ┌───────────────────┐
    │ Process Development │          │ System Development │
    │       Group         │          │       Group        │
    └───────────────────┘            └───────────────────┘
```
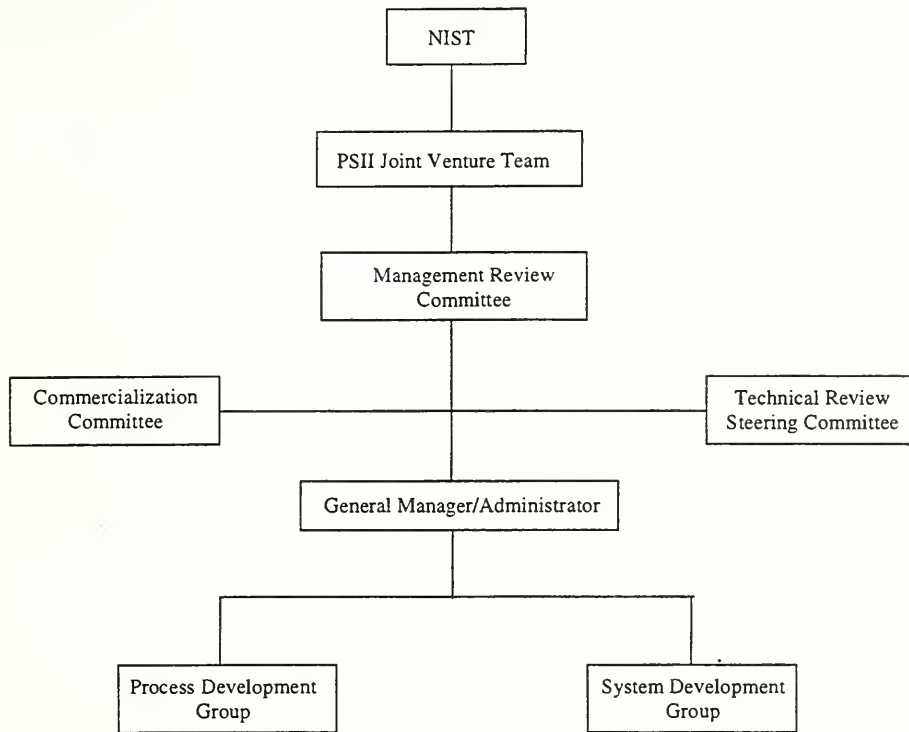
Figure 1. Program Management Structure

The third key strategic decision was to focus program resources on **demonstrating cost-effective solutions to technical barriers.** This decision allowed the JV to optimize for cost effectiveness rather than for complete technical fidelity. For example, the JV funded the development of two separate technical approaches for modulators, thereby developing a solid-state and electron tube based modulator so that the full range of operating regimes could be demonstrated rather than accept the risk of a single system solution. Ultimately, both technical approaches yielded acceptable devices for successful PSII implementation. These modulators led to new product sales in markets not available prior to the start of this program. Program resources were **not** focused on repeating conventional technologies common to PSII and other ion implantation techniques. However, adaptations or modifications were undertaken. In a separate example, a micro hardness test system was modified to specifically measure the unique properties of PSII-treated materials. The JV's philosophy was to maintain focus and resources only on cost-effective PSII commercial demonstrations, allowing private industry to backfill with fully optimized systems after commercial introduction.

The fourth key strategic decision of the JV was to **establish multiple PSII service providers as rapidly as possible.** To implement this decision, a two-part process is being implemented. First, three service providers, each servicing different markets, will be augmented with program-sponsored hardware to construct PSII machines. This decision allows for as many as three PSII machines to be available for commercial processing by late 1998. Second, as additional resources are made available, the JV will focus on the demonstration of a large volume (greater than 10 cubic meter) Beta Machine. The large volume Beta Machine is the final step in commercial demonstration of process scale up for batch processing of components (economic quantity driven) and/or large size components (dies, rollers, rams, etc.).

3. MULTIPLE-USER MARKET IDENTIFICATION

The original intent of this program was the active commercialization of PSII technology with the automotive marketplace as the primary driver. The early intent focused on tribological applications. However, within 12 months of JV formation, it became clear that applications of PSII recipes met requirements across multiple markets.

The ability of this team to reach out to additional markets was based in part upon the early systems engineering approach requiring each firm to conduct a market survey of internal and external customers. The market survey was the initial program phase for each firm. This survey established critical customer driven requirements, including:

- wear surface treatment recipe
- application component size
- batch processing size
- turnaround time
- customer acceptance price
- unique customer needs

This survey allowed the JV to identify of distinct machine sizes and to differentiate customer needs by group. For example, PSII application to stamping dies is driven by processing turnaround time. Furthermore, die costs are secondary to the cost of production downtime. Therefore, PSII application to this segment focuses on rapid turnaround time and the ability to handle varying die sizes and weights. In another example, batch component processing, high throughput without lot-to-lot process variation is critical.

Customer price acceptance was determined by two methodologies. First, the JV used computer models to develop a bottom up analysis of application costs. This allowed the JV to focus on target pricing based upon a realistic understanding of applied costs. Second, end users began to query customers directly as to their willingness to pay additional for enhanced wear surface treatments, especially after sample parts had been treated and tested at the customer's facility.

Armed with an understanding of customer needs and requirements, as well as with initial pricing data, the Commercialization Committee developed a strategic planning process for multiple-market introduction. Initial commercialization plans recognized market rollout opportunities. These opportunities were considered in a time-phased plan is shown in Figure 2 and were based upon the Standard Industry Code (SIC) for application type. Using this strategy, new JV members were identified and accepted in order to rapidly move PSII into cost-effective applications and develop a pathway for obtaining a return on investment (ROI) stream for invested program capital. This strategic planning process allowed the leveraging of market-pull and technology-push to achieve multiple-market introduction.

Figure 2. Market Opportunity Matrix

Processes	Team Leader	SIC 3544 Dies & Fixtures	SIC 35431 Tool Steels	SIC 34651 3592 Powertrain Components	SIC 363 Consumer Products	SIC 3841 3842 3843 Medical	SIC 3761 Military	SIC 3721 3724 3764 Aerospace	Competitive Technology
Cr	EHC	X		X	X	X	X		Plasma and Thermal Nitriding, Titanium Nitride
Tool Steel	A.O. Smith		X	X	X				PVD, CVD PECVD Coating
Zn/Brass	Kwikset			X	X			X	Plating, Titanium Nitride
Al/Mg	GM			X	X	X	X	X	Plasma Spray, Plating metal matrix composites
CR-Environmental Process Improvement	EHC	X	X	X	X	X	X	X	PVD, CVD, PECVD Coatings
System & Components	PVI/Ionex/DT	X	X	X	X	X	X	X	Beam Line Systems plating bath, Vacuum deposition equipment
Market Penetration	1996 1998 2000 2002 2004 2006 2008 2010								
Intellectual Property Strategy	Low [gradient] High	• Patent Technology • Provide Products Through Service Providers	• Patent Technology • Provide Services • Provide Products	• Patent Technology • License Suppliers • Sell Components Section 3.3.5	• Patent Technology • License Suppliers • Sell Components Section 3.3.5	• Patent • Sub License Section 3.3.5	• Patent • Sub License	• Patent • Sub License	

321

4. SUMMARY AND CONCLUSIONS

The commercialization of PSII into multiple-user markets by the PSII JV team has been based upon a dedicated process. To move this technology into the marketplace the program has identified several critical attributes:

- Develop a management process that focuses primarily on commercialization as the end result.

- Seek the voice of the customer through the addition of product end users to drive the commercialization requirements, including market price.

- Overcome technical barriers with cost-effective solutions that may not be the best technical solutions, and focus on getting to the market, not in building the absolutely best component.

- Use a vertically integrated team of research, engineering, component suppliers, and end users. Vertical integration reduces internal competitive frictions and allows idea sharing in a nonthreatening environment.

- Seek new members in noncompetitive market areas that require similar processes.

- Protect and provide full disclosure of all program-funded developments to all members throughout the program's life; at the same time inform the general public as appropriate through internal publications, seminars, and meetings.

The ability of this program to introduce PSII into the motor vehicle manufacturing marketplace, as well as into multiple commercial markets, within three-years is a result of the advantages of government-industry-research partnerships to accelerate the development of emerging and enabling technologies. These technologies lead to revolutionary products, processes, and services for the world's markets. The team attributes and sound management practices identified in this paper are necessary to successfully transition technology from the laboratory to the marketplace in a cost-effective manner.

Session VII. Welding and Compaction

An Overview of "Intelligent Resistance Welding (IRW)"
Program
W.F. Faitel, Lamb Technicon; S.J. Hu and H. Zhang,
University of Michigan

Precursory Study for Manufacturing Bi-Metallic, Light Weight
Automotive Components
J.A. Duggan, R.A. Marando, M.F. Miller, G.L. Meyers, and
B.A. Yablochnikov, Dana Corporation

Sensing of Gas Metal Arc Welds
T.A. Siewert, R.B. Madigan, and T.P. Quinn, NIST Materials
Science and Engineering Laboratory

Dynamic Magnetic Compaction (DMC) Technology for High
Density/Net Shape P/M Parts
J. Barber, D. Bauer, B. Chelluri, IAP Research, W. Jandeska,
GM Powertrain, T. Grady, D. Score, Delphi Automotive, T
Cadle, and J. Mandel, Zenith Sintered Products

An Overview of

"Intelligent Resistance Welding (IRW)" Program

William F. Faitel[1], Principal Investigator

S. Jack Hu[2], Technical Director

Hongyan Zhang[3], Deputy Technical Director

Intelligent Resistance Welding Consortium[4]

Abstract

The Intelligent Resistance Welding (IRW) Consortium was established with support from ATP and member companies to develop advanced technology and the knowledge base required to ensure quality and consistency of resistance spot welds for both steels and aluminum alloys. The objective of the program is to develop intelligent resistance welding process monitoring and control algorithms and systems.

Significant progress has been made since the program started in January 1996. This paper will provide an overview of the program and highlight achievements in the following five areas:

1) Quality definition of resistance spot welds

2) Finite element modeling and simulation of resistance welding process

3) Sensors and data acquisition system for process monitoring

4) Establishment of a fault tree for analyzing resistance welding machines

5) Integration and commercialization activities.

---

[1] Mr. Faitel is Director of R&D, Lamb Technicon - Body Assembly Systems. Phone: (810) 497-6518.

[2] Dr. Hu is an assistant professor in the Department of Mechanical Engineering and Applied Mechanics, the University of Michigan. Phone (313) 936-2921.

[3] Dr. Zhang is an assistant research scientist with the University of Michigan. Phone (313) 647-7419.

[4] Participating organizations of IRW are AISI, Alcan, Allen-Bradley, ANSYS, Auto Body Consortium, Battelle, Chrysler, DuPont, EWI, Ford, GM, ITI, Johnson Controls, Medar, Lamb, PICO, Robotron, RoMan, Sensotec/InTech R&D, Tower Automotive, and the University of Michigan.

1. Introduction

Resistance spot welding (RSW) is the predominant means of joining sheet metals in manufacturing automobile bodies. There are typically about 2,000 to 5,000 spot welds in every passenger car. However, because of a lack of confidence in the quality and consistency of the welds, a typical automobile has more welds than required to maintain its design requirements. Destructive testing is still used to determine if good welds are present. In addition, a reduction in vehicle weight and thus improved fuel efficiency, will depend on the introduction of advanced materials, such as aluminum, in automotive body structures, which will demand the development of different approaches to resistance welding.

1.1 Resistance Welding Process

Resistance welding is a process of joining two or more metal parts together in a localized area by application of electric current, based on the principle of joule heating. The heat is provided by the resistance furnished by the metal parts to the flow of current through the electrode tips, and a pressure is also provided by these same electrodes through air or hydraulic cylinders. Resistance welding involves complex interactions of electrical, thermal, mechanical, metallurgical, and other processes. The core of resistance welding is the welding machine. The schematics of a typical resistance welding machine are shown in Figure 1.
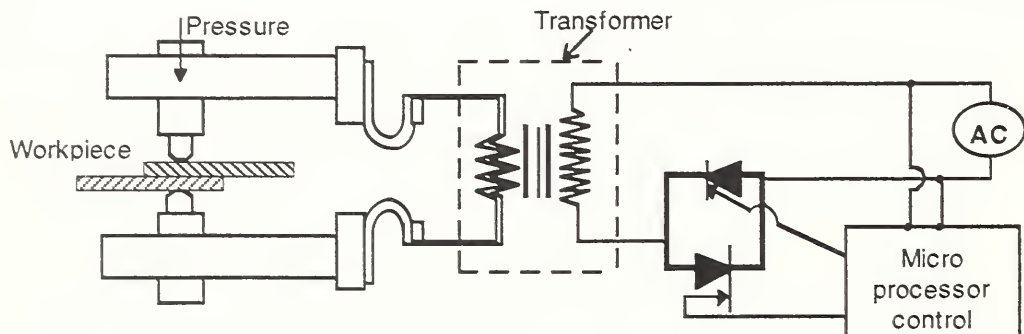
Figure 1. A resistance welding machine.

Many variables affect the performance of a welding machine. The influential ones include: (1) mechanical characteristics (e.g., welder/gun rigidity, electrode design/geometry, welding force, and production conditions); (2) electrical characteristics (e.g., current, voltage, power factor), and (3) mechanical /electrical characteristics (e.g., secondary design, type, and condition of welder). Other important factors that influence weld quality are materials, surface conditions, time, and human factors.

The state-of-the-art in resistance welding is, and has been for many decades, constant voltage or constant current control. Constant voltage or constant current control has proven successful for high weld integrity, and has been applied to a wide range of materials. Unfortunately, constant voltage is an open-loop control technology, and does not compensate for system or workpiece variations. An improvement has been the development of constant current control systems. Several variants of this technology have been introduced [Ferguson, 1986; Androvich, 1992]. This approach does compensate for some manufacturing variations [Gould et al., 1993]. However, it is still an open-loop approach, and cannot guarantee the quality of every weld.

Closed-loop controls for resistance welding have been studied since the 1950s. Feedback signals have generally been selected to reflect some inherent weld characteristics. These have included tip voltage, dynamic resistance, displacement, force, temperature, acoustic emission, etc. Many of these systems have proven successful on bare steels in controlled environments [Dickinson et al., 1980; Stiebel et al., 1986; Nagel and Lee, 1984; etc.]. However, the advent of coated steels, as well as the degree of variability inherent in current automotive manufacturing, have minimized the effectiveness of such units. In automotive applications today, more than 99% of resistance welding systems still use constant current control.

It is now recognized that real closed-loop control is not possible without an appropriate model. In previous efforts on closed-loop control, simple single-factor quality measures were assumed, providing very simple control models. It has been realized that resistance welding is a multi-variant process and requires appropriately high-order models [Neid, 1984; Tsai et al., 1991]. To date, however, these models are not of sufficient sophistication to handle the range of variables inherent in resistance welding, so feedback control has not been possible.

Recently, learning-based control algorithms, or neural networks, have been applied to resistance welding [Messler et al., 1994]. These approaches are experience based, relying on observed proven responses to predict future process responses. They do offer, however, the advantage of simpler models (given their empirical nature) for controlling the resistance welding process. Still, these approaches are very preliminary and experimental in nature.

Considerable work has been done on the welding of coated steels and more recently on the welding of aluminum. Much of this work has concentrated on either studying the wear mechanism of electrodes or trying to extend electrode life; less work has been carried out on the fundamentals of the welding process and no detailed work has been carried out in the US in understanding the significance of equipment and tooling design.

In summary, most of the past and state-of-the-art research efforts focused on the performance of a single weld, and tended to ignore machine characteristics, tooling, and production conditions. Also, when monitoring and controlling resistance welding, usually only one or two variables are monitored as feedback for control, in spite of the fact that resistance welding processes are multivariate in nature and the interactions among the variables are complex. These factors contributed to the limited success of these efforts.

To develop effective, intelligent welding monitoring and control systems, a systematic investigation of the impact of process variables (both controllable and uncontrollable) on weld quality is needed. Then, a multivariate approach to process monitoring and control will be developed based on the systematic investigation and understanding of the RSW process variables. Machine characteristics, both mechanical and electrical, and production conditions need to be systematically analyzed and incorporated in the design and control of RSW processes. The developed multivariate monitoring and control systems should be tested for resistance welding of a wide range of materials, including conventional steels, high strength steels, coated steels, and aluminum alloys.

In September 1995, the Intelligent Resistance Welding Consortium (IRW) was formed to tackle these difficult tasks. NIST-ATP funding jump-started the research and development activities. Participants in the consortium are AISI, Alcan, Allen-Bradley, ANSYS, Auto Body Consortium, Battelle, Chrysler, DuPont, Ford, GM, Johnson Controls, Medar, Lamb, PICO, Robotron, RoMan, Sensotec/InTech R&D, and Tower Automotive, with the University of Michigan, the Edison Welding Institute, and the Michigan Industrial Technology Institute participating in the research and integration activities.

This paper will provide an overview of the program and highlight the achievements from the past one and a half years.

2. IRW Program Overview

2.1 Goals and Objectives

The **overall goal** of the IRW program is to develop advanced technology and the knowledge base required to ensure quality and consistency of resistance spot welds.

Specific technical goals of the IRW program are:

- To improve the scientific understanding of resistance welding processes through modeling and analysis

- To develop sensors that are compact, robust, and cost effective for shop floor process monitoring

- To develop intelligent resistance welding monitoring and control systems

- To integrate IRW technologies for the expanded enterprise

2.2 Individual Projects

In order to achieve these goals, six technical projects and one integration project were identified at the beginning of the program. These projects address individual technical goals within the overall program goal. They are each led by a project leader from industry and an associate project leader from the universities or research institutions, with participation from multiple companies. These projects are described briefly below, and their relationships shown in Figure 2.
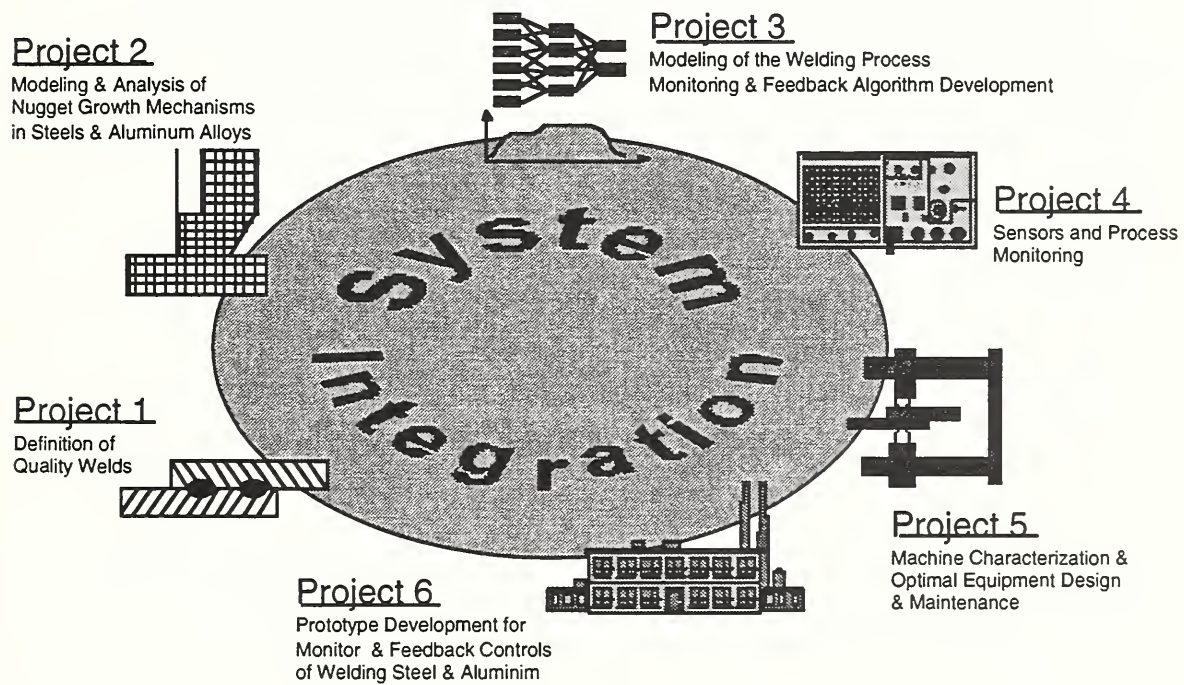
Figure 2.   IRW technical projects.

Project 1:   Quality Definition of Resistance Spot Welds

The goal of this project is to improve the scientific understanding and practical applications of quality definition and quality attributes, and to define the quality requirements of multiple welded auto body components. This project is vital to the success of the Intelligent Resistance

Welding (IRW) program. The first half of the project focuses on single welds. The relationships between weld attributes and performance characters will be developed, which will serve the IRW program in developing in-process monitoring and control systems.

Project 2: Modeling and Analysis of Nugget Growth Mechanisms

This project will develop finite element models and other physical models to simulate the growth of weld nuggets, and study the sensitivity of weld growth to process parameters. Computational models include electrode model, electrode/sheet model (both electrical/thermal and thermal/mechanical). Material modeling includes contact models, melting, and solidification.

Project 3: Modeling and Control Algorithm Development

This project aims to deliver algorithms for monitoring and control of resistance spot welding (RSW) processes in both steel and aluminum. These algorithms will embody knowledge of correlations between on-line *(in-situ)* measurements and the resultant physical weld attributes. It is thought that artificial neural networks (ANNs) may provide the best vehicle for establishing these correlations.

Project 4: Sensors for Process Monitoring and Control of Resistance Welding Processes

The goal of this project is to develop sensors for monitoring and control of resistance welding of steels and aluminum. This involves:

1) Developing and validating sensors required for monitoring resistance welding

2) Testing and evaluating the sensors and algorithms developed for welding of steels

3) Testing and evaluating the sensors and algorithms developed for welding of aluminum

Project 5: Machine Characterization and Optimal Equipment Design and Maintenance

The objectives of this project are to:

1) Develop a good understanding of welding machine characteristics, such as relationships between resistance spot welding equipment variables and weld quality matrices.

2) Develop criteria for optimal equipment design and maintenance for welding steels and aluminum alloys.

3) Correlate welding process and equipment characterization parameters with monitored signals for real-time feedback control systems.

Project 6: Development of Monitoring and Feedback Controls Systems for Steel and Aluminum

The goal of Project 6 is to demonstrate the commercial viability of the monitoring and control concepts resulting from Projects 1 – 5. At a minimum, commercial viability implies that the control and monitoring concepts are cost-effective to implement, cost-effective to maintain, and can be operated by the existing plant personnel with a moderate amount of additional training. What can be done in the laboratory or in theory on a computer cannot always be successfully implemented on the factory floor. The major objective for Project 6 is to discern which of the methods, sensors, and algorithms developed in Projects 3, 4, and 5 have the greatest possibility of commercially success. This will require extensive input from the designers, builders, maintainers, users, and buyers of welding equipment.

Project 7: System Integration and Multi-Disciplinary Planning

The goal of this project is to improve the effectiveness of the research teams. Our approach is to apply current best practice to foster the interaction among the teams, and leverage each other's work, and insure deployable results.

3. Highlights of Achievements

During the first 18 months, significant progress has been made in all the IRW projects. However, due to the space limitations and the proprietary nature of some of the results, this paper will highlight the following specific achievements:

1) Quality definition of resistance spot welds

2) Finite element modeling and simulation of resistance welding process

3) Sensors and data acquisition system for process monitoring

4) Establishment of a fault tree for analyzing resistance welding machines

5) Integration and commercialization activities

3.1 Quality definition of resistance spot welds

Currently, most manufacturers use button diameter after a peel test as the single indicator of weld quality. A state-of-the-art survey of the standards/requirements of RSW quality has been conducted and the results indicate a wide discrepancy among the standards used by automotive

manufacturers, the aerospace industry, and the military. Figure 3 shows a comparison of some standards and specifications.
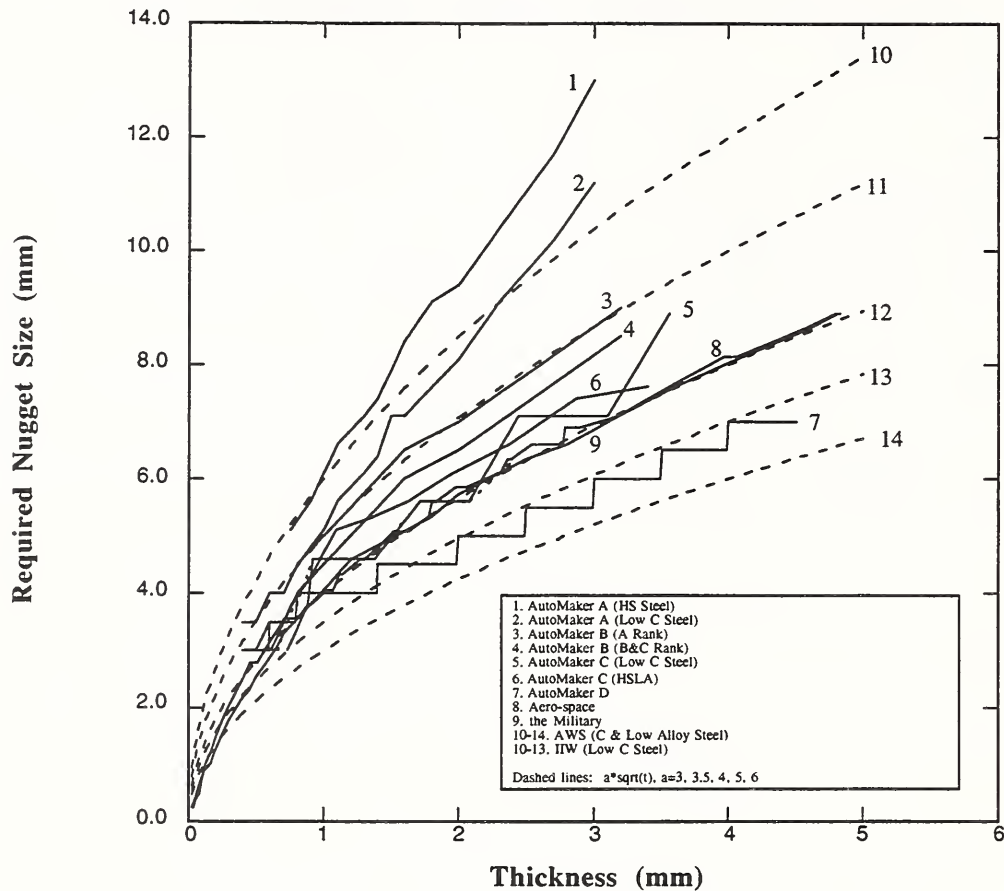
Figure 3.   Comparison of standards/specifications on weld nugget size requirements.

One of the reasons for the discrepancy shown in Figure 3 is the uncertainty of what constitutes good weld quality. Additionally, the lack of standardized tests makes it harder to reach a consensus on weld quality among practitioners. In this project, a series of tests have been conducted on coupon size sensitivity for the tensile shear test (the most commonly conducted test in weld strength measurement). A methodology has been developed to find the optimal coupon size. Materials tested include commonly used steels and aluminum alloys of various gauges. Dynamic strength of spot welds is an important index of weld quality. But because of the complexity it involves, dynamic testing of spot welds is not commonly performed. To overcome the shortcomings of current dynamic tests, a simple, yet accurate testing device has been developed for dynamic strength measurement of welded joints.

3.2 Modeling and simulation of resistance welding process

A fundamental understanding of the resistance welding process is essential to the development of monitoring and control algorithms. Modeling provides a good approach toward this understanding. Coupled electric-thermal and thermal-mechanical finite element simulations have been carried out to capture the dynamic characteristics of nugget growth. Specifically, the model has been used to study the impact of electrode geometry and current wave form on nugget growth. These models are applied to bare steels, galvanized steels, and aluminum. A good understanding of the resistance welding process has been obtained from these analyses. Figure 4 shows a simulation of nugget growth, as well as a comparison between experiments and modeling for bare steel welding.

(a)

(b)
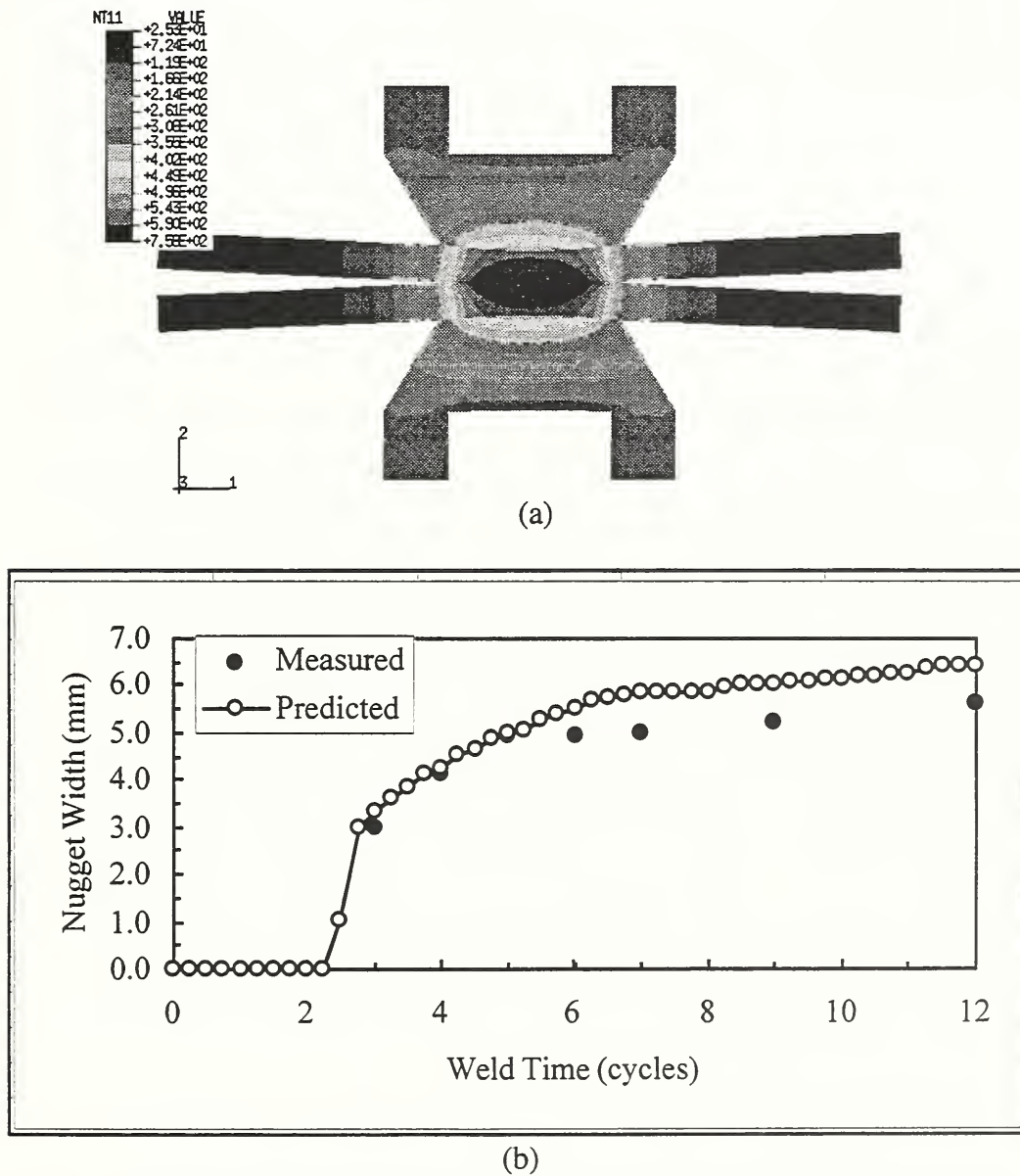
Figure 4. FEM simulation of nugget growth and comparison with experiments.

In addition to nugget growth modeling, numerical, analytical, and experimental studies have also been conducted to understand some important aspects of resistance spot welding. A contact resistance model has been developed that includes electrical/thermal conduction models for a single contact and multiple contacts.

Efforts have also been made to predict and prevent expulsion. Expulsion is usually considered undesirable, although it is prevalent in the production environment. An example of expulsion is shown in Figure 5a. A general expulsion detection model, together with procedures for its implementation, was developed by the IRW program. A comparison of the model with experiments is shown in Figure 5b, with good agreement.

(a)   (b)
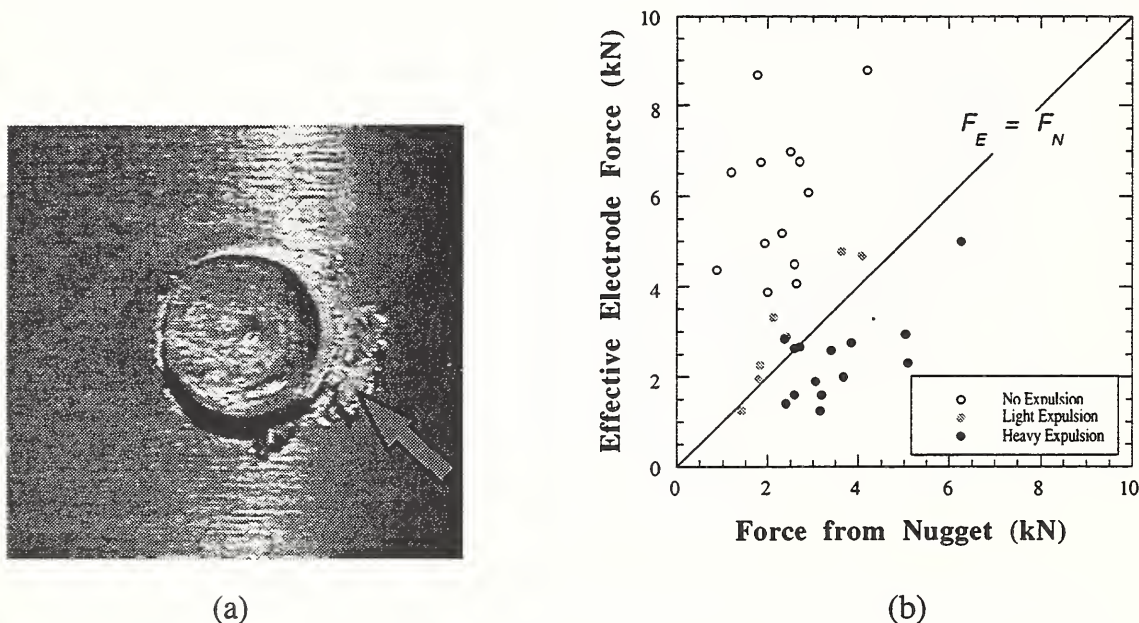
Figure 5.   (a) A weld button with expulsion.   (b) Expulsion model verification.

3.3 Sensors and data acquisition system for process monitoring

The ultimate objective of IRW is to develop intelligent process monitoring and control systems for resistance welding. These systems will use sensors to obtain process conditions, and apply advanced signal processing and control techniques for process monitoring and control. Various existing sensors have been evaluated and new potential sensing technology explored. For example, the possibility of using fiber-optical sensors for displacement monitoring of nugget growth was evaluated with great promise.

The data acquisition system developed by the IRW team consists of various sensors, a CPU, AD cards, a signal conditioning unit, and a barcode scanner. The interface of the data acquisition system is shown in Figure 6.
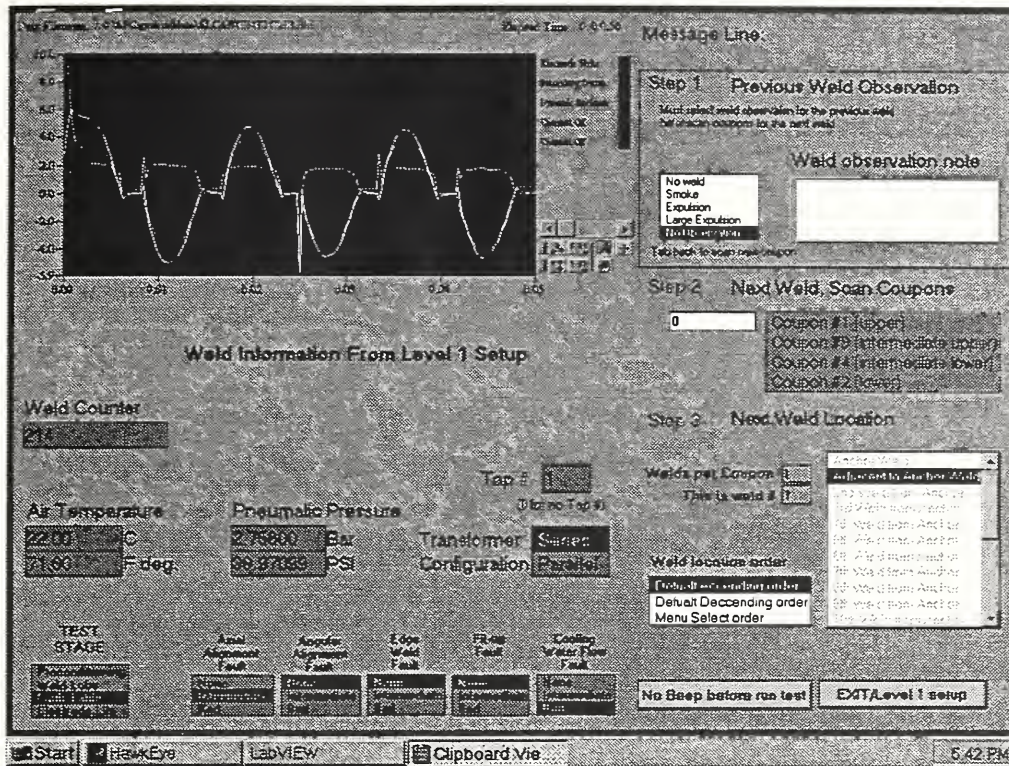
Figure 6. IRW data acquisition system interface.

3.3 Monitoring and control algorithm development

This program aims at delivering algorithms for monitoring and control of resistance spot welding (RSW) processes in both steel and aluminum. These algorithms will embody the knowledge of correlations between on-line (*in-situ*) measurements and the resultant physical weld attributes. It is thought that artificial neural networks (ANNs) may provide the best vehicle for establishing these correlations. On the other hand, to enhance the understanding of the RSW process, we are also working on a model-based approach. It could offer more insights to help our ANN-based approach. These algorithms will first be implemented as non-destructive weld monitoring, and eventually in a form to manipulate weld current and possibly other input signals in real-time. Besides artificial neural networks, possibilities of using other methods, such as fuzzy logic, will also be explored.
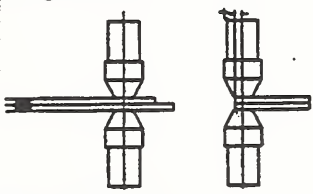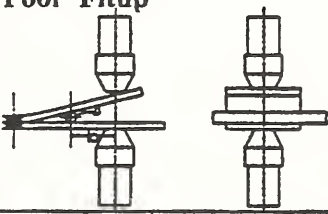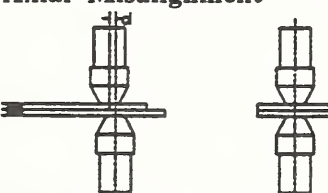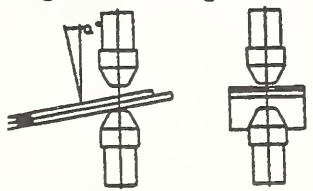
A set of five data acquisition systems have been developed for five experiment sites (Table 1), and all of them have been installed and are operational. Design of experiments (DOE) that mimic the various fault conditions (Table 2) from production environments have been carried out at the University of Michigan site. The purpose of the DOE weld experiment is to develop a data collection methodology so that the data used for the development of monitoring and control algorithms is representative of an actual welding environment. In order to investigate variance in the spot welding process created by varying the parameters, the data collection

335

procedure will be performed under several conditions. This process is to be repeated on different machines, for different power sources, with various electrode geometry, and for various types of steel and aluminum. This data will be used to develop the monitoring and control algorithms.

Table 1. IRW Experimental Site Matrix.

Material	Machine Type	Power Source	Test Site
Steels	C-Gun	Single Phase AC, MFDC	Medar
	Pedestal	Single Phase AC	UM
	Press	Single Phase AC	Candid Logic
Aluminum Alloys	Scissors Gun	MFDC	Alcan
	Pedestal	Single Phs. AC, Single Phs. DC	Chrysler CTC
	Pinch Gun	Single Phase AC	Lamb

Table 2. Faults of the DOE Experiments.

Fault	Definition
Edge Weld	Electrodes are aligned, the position of the test weld is near or on the side edge of the coupon. Position test weld so that there is a distance "t" from the center of the surface diameter of the bottom electrode to the side
Poor Fitup	Electrodes are aligned, a plastic-coated wire with diameter D is placed in between the two coupons being welded together. The block is placed approximately half-way between the shunt weld and the test weld.
Electrode Cooling Water Flow Rate (No figure)	Electrode cooling water flow rate is defined by the sum of the readings on two flow meters attached to each electrode.
Axial Misalignment	Electrodes are shifted horizontally off center. Instead of electrodes being aligned vertically, one electrode is shifted horizontally to the left or right. The electrode is still vertical, there is no tilt.
Angular Misalignment	Electrodes are aligned, the coupon is tilted up to the right or left, so that it is at an angle with the electrodes.

336

3.4 Establishment of a fault tree for analyzing resistance welding machines

A working model of a resistance welding fault tree has been developed, which has now been placed on the Web. The model can be accessed by anyone with access to the IRW homepage. It can be found in the Resources page in the Download area under IRW Project Documents. (Go to IRW Project 5). Currently the model is somewhat skeletal and still subject to change, but an impression of the appearance and functioning can be gained by the user. The structure of the fault tree is shown in Figure 7.
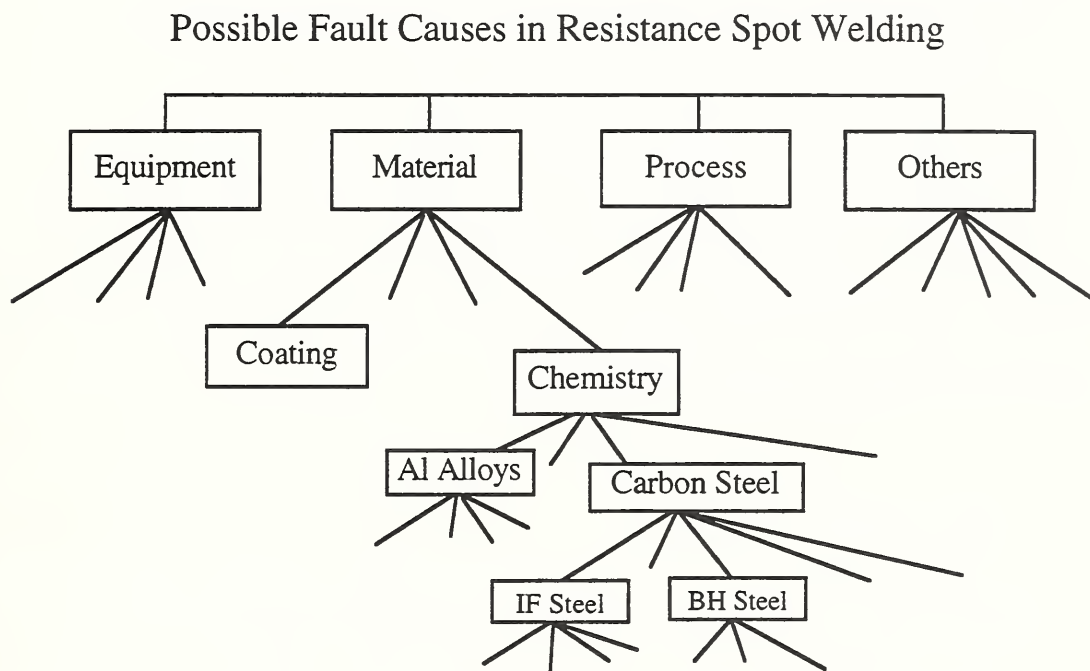
Figure 7.   Structure of the fault tree.

3.5 Integration and commercialization activities

Managing such a large program with so many participants is a difficult task. At the very beginning, the program put together an integration team that is responsible for ensuring the information flow among IRW projects. The end goal is to ensure that the results of the research projects within the IRW program are integrated with each other and with the manufacturing environment to form an integrated agile system that is readily modifiable, expandable, and transferable to meet user needs. The integration effort is divided into four primary tasks. The first task, communications, focuses on methods to improve communications between and within project teams. The second, infrastructure, is focused on

methods to support research project teams through common systems. The third, the plant floor workbench, ensures the deployability of research results onto the shop floor. The final task, the design workbench, ensures the deployable results of the research into the engineering environment.

A number of potential commercializable products were identified through brainstorming by the team members. Several of them are under active development, such as the impact tester for welded joints, and the expulsion prediction system.

4.   Summary

The Advanced Technology Program brought together manufacturers and research institutions to work together on some generic problems in resistance spot welding of steels and aluminum. It would be extremely difficult, if not impossible, for individual companies to deal with a problem of such scale. Significant progress has been made during the past year and half in the Intelligent Resistance Welding program, through a collective effort of industrial members and research personnel of the research institutions.

5.   References

1.   Androvich, D.A. (1992). "Resistance welding using constant current heat control", *Sheet Metal Welding Conference V*, AWS Detroit Section, Detroit, MI.

2.   Dickinson, D. W., Franklin, J. E., and Stanya, A. (1980) "Characterization of Spot Welding Behavior of Dynamic Electrical Parameter Monitoring", *Welding Journal*, June 1980, pp. 170-s to 176-s.

3.   Ferguson, H.S. (1986). "Programmed constant-total weld pulse current control", *Sheet Metal Welding Conference II*, AWS Detroit Section, Detroit, MI.

4.   Gould, J., Kimchi, M., and Mitchelle, T. (1993). "Effect of fit up conditions when using constant current control systems for resistance spot welding", *SAE Technical Paper* No. 930451, SAE.

5.   Messler, R.W., Jou, M., and Li, C. J. (1994). "A fuzzy logic control system for resistance spot welding based on a neural network model", *Sheet Metal Welding Conference VI*, AWS Detroit Section, Detroit, MI.

6.   Nagel, G. L., and Lee, A. (1984). "A new approach to spot weld feedback control", *Sheet Metal Welding Conference I*, AWS Detroit Section, Detroit, MI.

7.   Nied, H. A. (1984). "The finite element modeling of the resistance spot welding process", *Welding Journal,* 63(4), pp. 123-s to 132s.

8.   Stiebel, A., Ulmer, C., Kodrack, D. and Holmes, B. B. (1986). "Monitoring and control of spot weld operations", *SAE Technical Paper* No. 860579, SAE.

9.   Tsai, C. L., Dai, W. L., Dickinson, D. W. and Parritan, J. C. (1991). "Analysis and development of a real-time control methodology in resistance spot welding", *Welding Journal*, December 1991, pp. 339-s to 351-s.

Precursory Study for Manufacturing Bi-Metallic, Light Weight, Automotive Components

James A. Duggan
Chief Engineer - Advanced Design
Dana Corporation - Spicer Driveshaft Division

Richard A. Marando
Lead Project Engineer - Hydroform Development
Dana Corporation - Parish Light Vehicle Structure Division

Michael F. Miller
Lead Project Engineer - Global Testing
Dana Corporation - Parish Light Vehicle Structure Division

Gerald L. Myers
Senior Design Engineer - Advanced Design
Dana Corporation - Spicer Driveshaft Division

Boris A. Yablochnikov, Ph.D.
Project Engineer - Lead Scientist
Dana Corporation - Spicer Driveshaft Division

Introduction:

Dana Corporation, headquartered in Toledo, Ohio, is one of the world's largest automotive parts suppliers employing 48,000 people at 508 facilities located in 29 countries with global sales nearing $8 Billion annually.

With the first practical universal joint from the Spicer Driveshaft Division in 1904 and heat treated vehicle frames from the Parish Light Vehicle Structrures Division in 1905, Dana Corporation has set industry standards for power transmission and vehicle frame technology. Spicer Driveshaft Division, the oldest and largest independent driveshaft manufacturer with its headquarters and research facility located in Holland, Ohio, has affiliate operations in 17 countries worldwide. The Parish Light Vehicle Structures Division is North America's largest structural component supplier producing full frames, engine cradles and suspension crossmembers for the light truck, sport utility and passenger car markets. Parish, with its headquarters and largest manufacturing facility in Reading Pennsylvania, has affiliates in 6 countries of North and South America.

Continuing research of joining technologies is taking place at the Spicer Driveshaft Division headquarters and research facility in Holland, Ohio. Dana's hydroform development facility is located in the Parish Division Reading, Pennsylvania, Technical Center which houses a

5,500 Ton hydroforming press and a 6.375" Multiple Radius CNC bending machine supported by a fully accredited structural test lab with MTS Servo Hydraulics and RPC 3 test equipment.

Abstract:

Dana Corporation and the National Institute of Standards and Technology (NIST) have entered into a cooperative agreement, through the Advanced Technology Program (ATP). Under the Motor Vehicle Manufacturing Technology (MVMT) focus group, the goal of this project is to develop and commercialize a manufacturing process for producing light weight bi-metallic automotive components through Hydroforming and/or Magnetic Pulse Welding. Resulting technology from this project will provide the automotive manufacturing community the ability to exceed today's strategies for meeting current and future weight reduction, cost, accuracy and environmental requirements.

Dana proposes to make automotive components for drive train and frame structures utilizing a tubular hydroforming process to form tubular work pieces to a final shape, and a unique welding process to assemble the formed work pieces. The products of the hydroforming and welding processes are lighter weight components with increased stiffness, without welding filler metal, thermal effects, or fumes and can consist of similar or dissimilar materials. The welding process is capable of metallurgically joining aluminum to aluminum and aluminum to steel. This selectivity of similar or dissimilar substrate materials allows a designer to choose the material best suited for the application.

The two processes can be utilized independently or in conjunction. This paper addresses each process independently. Section 1, Hydroforming, defines the tubular hydroforming process in all of its variations, and offers a logic based approach for designing the most cost effective hydroform part that meets vehicle system requirements. Section II, Magnetic Pulse Welding, describes the method of metallurgically joining (i.e., welding) similar and dissimilar materials without the negative effects of sustained arc welding.

SECTION I - Hydroforming

When discussing the tubular hydroforming process, a distinction is usually made between low-pressure hydroforming and high-pressure hydroforming. The distinction refers to the magnitude of the pressure applied internally to force a tubular blank into conformance with the die cavity during the forming process. The proponents of low-pressure hydroforming will tell you their typical application requires no more than 5,000 psi. Conversely, those who espouse to high-pressure hydroforming will tell you that most automotive applications require 15,000 to 30,000 psi internal forming pressure. Beyond internal forming pressure, the lines of differentiation between the two processes become blurred and conflicted. Make no mistake, there are significant differences between the two processes and selecting the wrong one for your product can have expensive consequences.

Whether you are talking to an equipment manufacturer or a component supplier, both have made significant financial and emotional investments in which ever process they are working with. Therefore, it is difficult for them to be completely impartial on the subject of which process is most suited to your application. Product teams who wish to apply the hydroforming technology are forced to wade though a sea of conflicting information in an effort to determine the best process for their application. Since each process has it's own set of design limitations the High / Low decision must be made very early in the product development cycle when the core product characteristics are typically not fully defined. As the product evolves, unanticipated interference areas, critical quality relationships, attachment requirements, crash tests and analysis results will all drive changes to the structure. Any one of these issues has the potential to make your first decision the wrong decision and ultimately drive cost into your program and your product.

Superimposed Tension Forming

From a process standpoint:

The primary difference between low and high-pressure hydroforming is the way the material is moved into corners of the die. With the low-pressure process, material slides into the corners of the die. With the high-pressure process, material expands into the corners of the die.

In application, the low pressure process begins with a tubular blank substantially equal in perimeter to the finished part cross sections and involves lightly pressurizing the blank before closing it in the die cavity. This initial pre-closing pressure gives the blank a greater resistance to collapse during the re-shaping process allowing the blank to take the shape of the die cavity as the die is closed upon it.

The high-pressure process employs a tubular blank slightly smaller than the finished part perimeter allowing the tube to be closed in the die cavity without the assistance of pre pressurization, and subsequently, uses a substantially higher pressure to expand the tube into conformance with the die cavity.

From a product standpoint:

The primary difference between low and high pressure hydroforming is:

SUPERIMPOSED TENSION.

The low-pressure process uses internal pressure to force a tubular blank to bend or fold at desired locations allowing the blank to take the shape of the die cavity. Therefore, parts produced using this process are subject to residual stresses in the bend zones commonly referred to as spring back. To illustrate this, Figure 1 shows the cross section of a typical bend identifying the compression zones (yielded and un-yielded), the neutral axis, and the tension zones (yielded and

un-yielded). In the two un-yielded zones, the individual grains have not been compressed or stretched enough to exceed the yield limit of the material. Therefore, the material in the un-yielded zones is stressed below the elastic limit and wants to "spring back" to its original state creating residual stresses which work to unbend the part upon release of the bending forces.
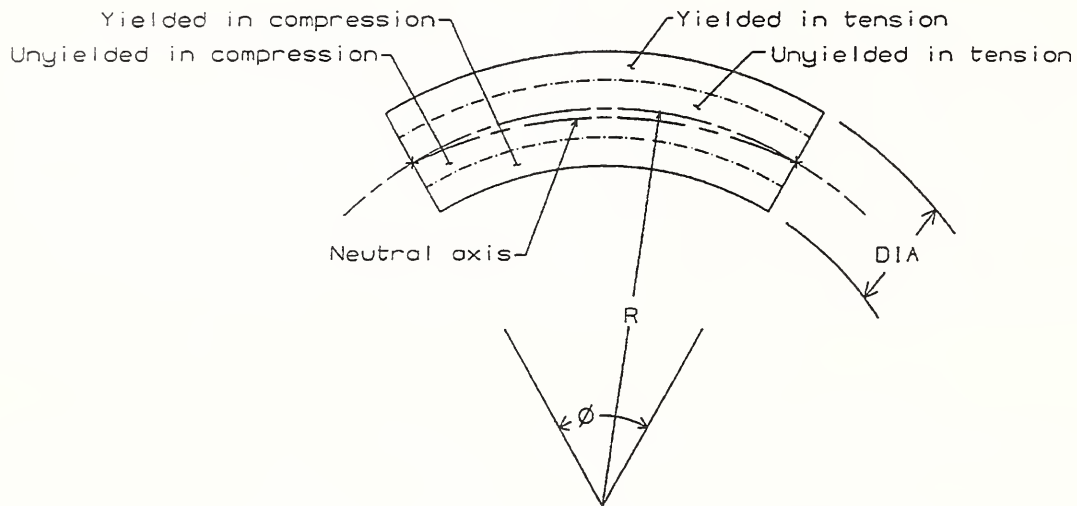
Figure 1 - Illustration of stress zones of tubular workpiece.

Conversely, the high-pressure process stretches material into the corners of the die effectively performing a process known as "Superimposed Tension Bending". The entire cross section of the bend is yielded in tension thereby eliminating un-yielded zones and competing residual stresses which are the basis for spring back. Theoretically, Superimposed Tension eliminates spring back by forcing the neutral axis off the part. Practical demonstrations of this process indicate spring back is not completely eliminated as shown in Figure 2. However, the remaining spring back is negligible allowing the high-pressure hydroforming process to be categorized as a Net Shape process.
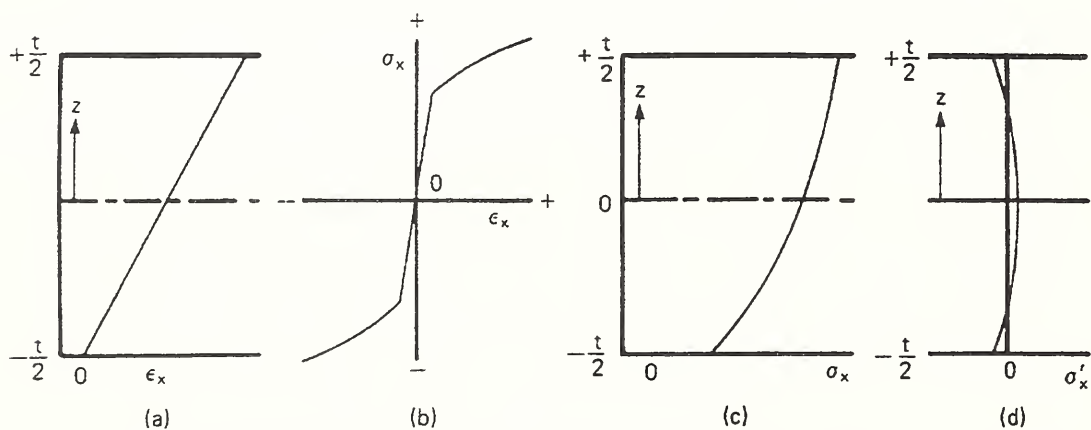
Figure 2 - With sufficient tension, the neutral axis moves out of the sheet so the strain is in tensile across the entire section, (a). With the stress-strain curve shown in (b), the stress distribution in (c) results. After removal of the moment, elastic unloading leaves very minor residual stresses, as shown in (d).

While there are other differences between low-pressure and high-pressure hydroforming, none are as significant in their effect on either the product or the process as is superimposed tension. Therefore, from this point forward, low-pressure hydroforming will be simply referred to as hydroforming and high-pressure hydroforming as **SUPERIMPOSED TENSION FORMING** or **S.T.F.** for short.

The internal pressure required to create Superimposed Tension for a given part is a function of the material thickness, the yield strength and the Inside Radius of the tightest bend on the part. Therefore, it is not possible to define a pressure breakpoint between Hydroforming and S.T.F. for all parts. However, it is reasonable to assume that **S.T. F.** requires a substantially higher pressure than does hydroforming.

The Cost Of Pressure

Internal forming pressure has a direct relationship to Piece Cost !!

Piece cost is a summation of the material, labor and burden costs associated with the production of a given part. Burden cost is a summation of a variety of process cost factors, typically classified as either fixed or variable. Fixed burden factors exist for the support of every product produced in a given facility and are independent of the process methodology. An example of fixed burden would be the heating costs for the facility. Variable burden reflects those costs associated directly with the production of a given part. An example of variable burden would be the annual maintenance cost for the equipment used in the production of a given part. The maintenance costs for the **S.T.F.** process are higher than the hydroforming process due to the degree of sophistication and relative unavailability of the pressure generation equipment used in the process. For example, basic pressure intensification and pressure delivery equipment used in the hydroforming process can be purchased from hundreds of companies around the world. Those same components within **S.T.F.** process are produced by only a handfull of specialty companies and are therefore more expensive to repair and replace.

Another portion of variable burden is capital depreciation. As stated previously, the S.T.F. process requires a higher internal forming pressure than does the hydroforming process. Both of the subject forming techniques utilize a hydraulic press to react the die separating forces generated by the internal forming pressure, as shown in Figure 3. Logically, the **S.T.F.** process requires higher press tonnage to react the higher separating forces generated by the higher internal forming pressures. Higher press tonnage requirements equate to higher capital cost for the **S.T.F.** Process.
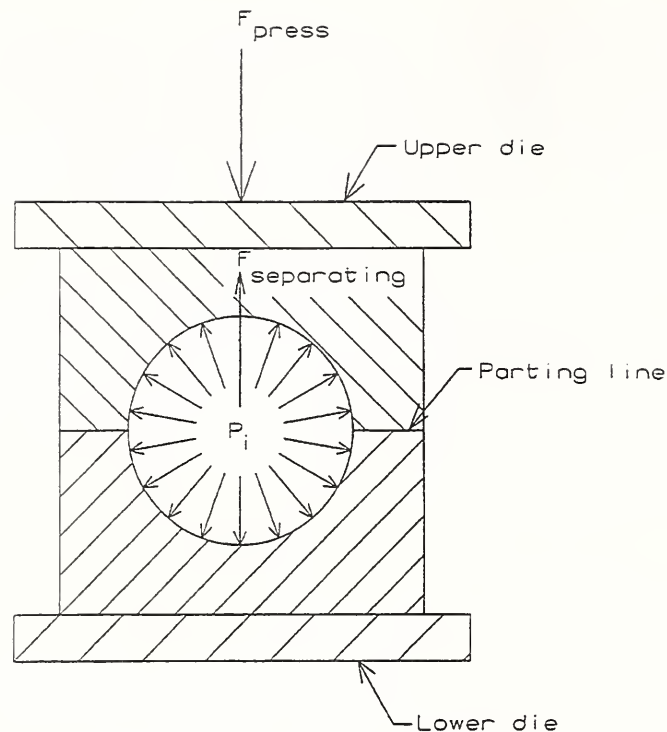
Figure 3 - Illustration of forming pressure and force.

Since the **S.T.F.** process uses higher internal forming pressures, it requires a more robust tool to ensure pressurization without deflection. It is logical to assume the larger, more robust tooling will be more costly than that required for the hydroforming process.

In practice, the high production volume of the typical automotive application provides ample opportunity to recover the increased capital and tooling costs driven by the **S.T.F.** process thus minimizing the vehicle level cost impact. The costs induced by the **S.T.F.** process are not excessive when the down stream benefits of net shape processing are considered. However, when compared directly to the hydroforming process, **S.T.F.** is more expensive.

In comparison to traditional stamp and weld techniques, it is generally accepted that the more stamped and welded components replaced by one hydroformed or **S.T.F.** member, the more cost effective the process. Essentially, the cost of applying either the hydroforming or the S.T.F. process is offset by the elimination of value added content on the traditionally processed part (i.e.: welding wire, assembly labor, material overlap, inventory, handling costs, etc.). Since the S.T.F. process is more costly to employ than the hydroform process, the threshold for economic feasibility is higher for the **S.T.F.** process than the hydroforming process. The tooling cost for either process is almost always less expensive than tooling for a traditional stamp and weld assembly when the assembly tooling is included.

Design Flexibility

When applied to a part larger than a bread box, the hydroforming and **S.T.F.** processes are fairly similar in terms of design flexibility. Both processes require a pre-bent tubular blank and are therefore subject to the same bending do's and don'ts. The importance of the bending process should not be under-estimated. There are many practical examples where a hydroform program was dropped because the tubular blank could not be bent.

The hydroforming process, as stated earlier, relies on the tubular blank being nearly equal in perimeter to the finished part. Therefore, when designing for the hydroforming process, the finished part should be designed with 0% expansion. Any expansion that does occur during the process will result in wall thinning. The inside bend radii (usually expressed as a ratio to wall thickness - Ri/t) can theoretically be as tight as 1t. However, a tight radius part is difficult to manage in a production environment, and therefore not a recommended approach. Generally speaking, a 4t radius is considered a robust target for design.

The **S.T.F.** process can, to some extent, expand the tubular blank without wall thinning inside the end-feed zones. The ability to end-feed material into the expansion zone reduces as you traverse the length of the part. The effects of end feeding are virtually eliminated after the first major bend or form change. The amount of expansion possible within the end-feed zone is a function of the distance from the end of the tube, the type of material, the geometry of the finished part and a host of process variables. Therefore, it is impossible to state a fixed expansion percentage limitation within the end feed zone for all parts. For discussion purposes, a typical automotive structural component produced from mild carbon steel will have an expansion limitation closer to 15% than 50%, unless the expansion zone is limited to the very end of the tube. Any expansion occurring outside the end feed zone will result in wall thinning. Since the **S.T.F.** process relies upon stretching metal into the corners, it is desirable to allow for approximately 2% expansion in areas outside the end feed zone. Theoretically, it is possible to produce 1t bends with the **S.T.F.** process. However, it is impractical due to the extremely high internal forming pressure required to stretch metal into the tight radius. This in turn drives press tonnage and capital cost. We have found designing with 4t or 5t inside radii yields a readily formable part which does not require excessive internal forming pressure.

Hole perfing is one other design related issue considered when working with the hydroforming or **S.T.F.** process. Both processes have the same limitations in terms of where and how many holes can be perfed in the die. The key difference is the amount of deformation around the perfed hole. Since the hydroforming process uses a relatively low internal forming pressure, its ability to react the shear forces generated by the perfing operation are limited. Even with a tapered punch face, there remains a greater degree of deformation around the hole. The **S.T.F.** process is able to provide a much flatter surface around a perfed hole. However, some nominal amount of deformation does occur in most applications. In either case, large lightening holes and holes requiring a perfectly flat interior surface can not be achieved within the forming operation and must be post processed. Due to the relatively high cost of relocating an imbedded

perf unit. we recommend post processing (by laser, fine plasma or high speed punch) any holes likely to change location frequently during the life of the product.

The Quality Of Pressure

In today's highly competitive global automotive market, quality is an expectation not a sales feature. Both the hydroforming process and the **S.T. F.** process will yield a part with greater dimensional stability than a traditional stamp and weld assembly by simply eliminating the thermal distortion induced during the seam welding process. The **S.T.F.** process takes it a step further and effectively eliminates spring back in the bend zones yielding a near net shape part.

We recently conducted a study comparing the **S.T.F.** process to a traditional stamp and weld process. A production side rail (inner and outer half shells seam welded together) was reproduced using the **S.T.F.** process. The **S.T.F.** part demonstrated an average 8-fold improvement in Cp values over the stamped and welded component when the production tolerances were applied. The Cp average for all 18 points checked in the study was 8 with the low at 5 and the high at 20. Distribution charts for the **S.T.F.** formed part and the stamped and welded assembly, checked at the same location over 31 pieces are shown in Figures 4a and 4b respectively.
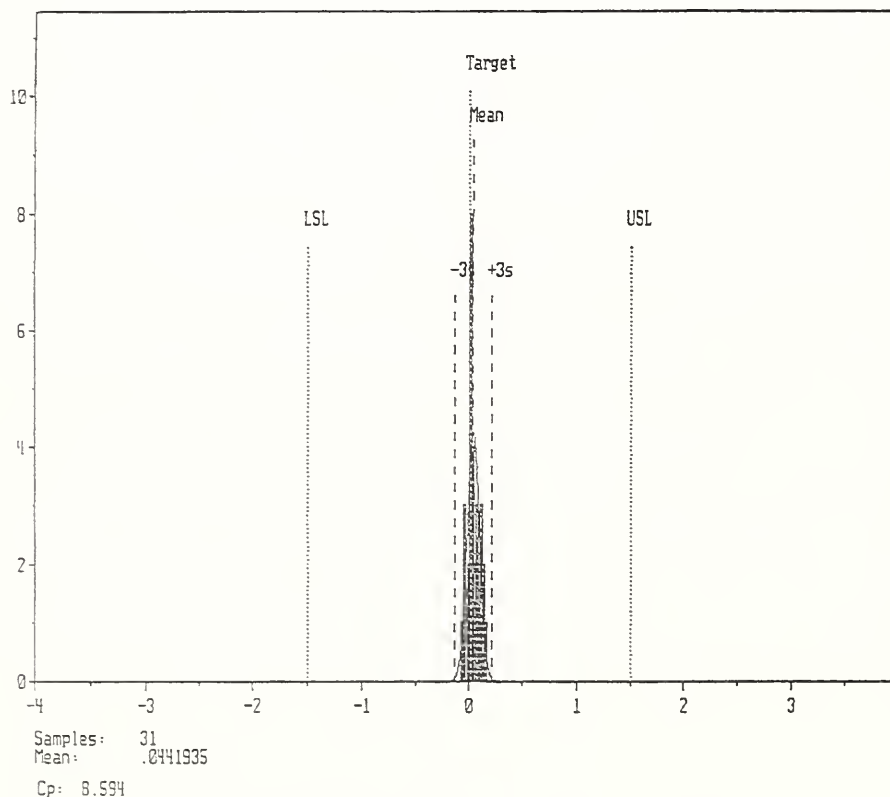
Figure 4a - Measured Dimensional accuracy of **S.T.F.** process.
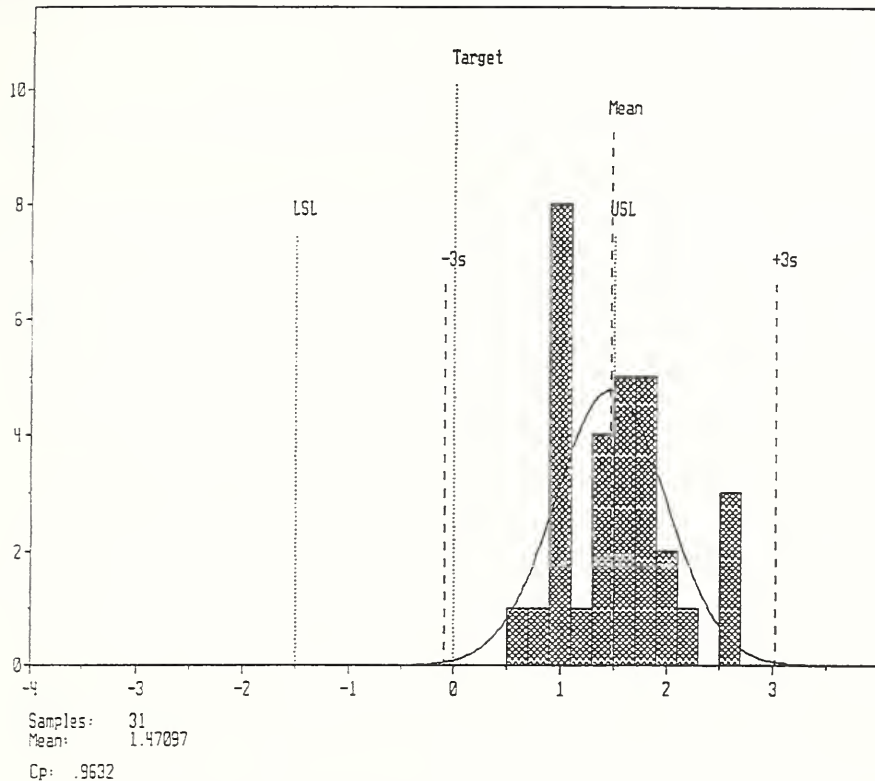
Figure 4b - Measured dimensional accuracy of Stamp and Weld process.

Durability / Structural Integrity

It is widely known that welded structures, when dynamically loaded, are much more likely to fail at, on, or around a weld than they are anywhere else on the structure. By eliminating the seam weld on a given part, it is logical to assume the opportunities for weld related failure are correspondingly eliminated. Therefore, the structural integrity of the part is improved.

A key component in the durability of a structural assembly is part fit-up. In a typical automotive structure, numerous brackets are welded to the main members to facilitate attachment to a number of vehicle systems. One common source of durability failure is poor bracket fit-up. Assuming the tooling used to produced the mating parts is properly designed and well maintained, poor bracket fit-up can usually be attributed to one of 3 factors; spring back within the bracket. spring back within the main member and thermal distortion resulting from the seam welding process within the main member. If the main member is produced with the hydroforming process, the thermal distortion factor will be eliminated which effectively eliminates 1/3 of the variables contributing to poor bracket fit up. If the main member is produced with the **S.T.F.** process, both the thermal distortion factor and the main member spring back factor will be eliminated which effectively eliminates 2/3 of the variables contributing to poor bracket fit up . By eliminating variables in the bracket fit-up equation, it is logical to

assume that good bracket fit-up is more likely, and therefore, the durability of the structural assembly will be improved.

Meeting The Vehicle System Needs At The Lowest Possible Cost

Unless you are certain at the onset of your program that some aspect of the hydroforming process is unacceptable, begin your design efforts with the hydroforming process in mind, but develop the production cost model around the **S.T.F.** process. If you can use the hydroforming process to effectively meet vehicle system needs, using the **S.T.F.** process would be an unwarranted expense. If, during the design evolution, it becomes evident the **S.T.F.** process is needed to meet system needs, your financial plan will support the approach.

Prototype tooling built to withstand the forming pressures of the **S.T.F.** process will allow you to evaluate the quality of both processes and provide objective data from which to select the production process. This approach is best facilitated through a source that has both hydroforming and **S.T.F.** process capabilities, but can be done through a cooperative arrangement with independent sources.

SECTION II - Magnetic Pulse Welding

Solid State welding is not a new concept for metallurgically joining similar and dissimilar metals. A variety of driving mechanisms can generate the required pressure to drive materials together on an elemental basis. The US Mint uses a clad metal alloy of silver, copper, and zinc to form the coins in circulation today. The sheets of alloy are pressure formed using rolls to squeeze the metal together and form the clad metallurgy. In attaching tubular elements, the oil piping industry uses pressure generated by an explosive blast to join pipeline sections underground. Magnetic forming has been widely used to shape and attach large diameter tubular components, such as driveshafts, by using a torque transmitting profile and cold forming the tubing around the end-fitting. For these large diameter sections, the attachment is strictly mechanical. Smaller diameter mechanical strength tubes, 25 mm or less, have been metallurgically bonded for a variety of applications including welding nuclear fuel rods. Larger tube diameters up to 60 mm have been welded, however the tubing is 1100 series pure aluminum and too weak for structural applications.

The objectives of this portion of the MVMT project are to: 1) develop a process, and the supporting equipment, to metallurgically bond steel components to large diameter (e.g., 100 mm and greater) mechanical grade aluminum tubing (e.g., 6061-T6 extruded tubing); 2) commercialize the process with a fixed diameter coil for the manufacture of high volume automotive driveshafts and 3) develop an agile process to accommodate closed-loop automotive structures.

What appears to be a scaling factor to current Magnetic Pulse Welding (MPW) technology has become an improved triggering mechanism and machine layout to achieve an extremely low inductance device with minimal parasitic losses. This system optimization allows energy levels on the order of 50 kilo Joules to be repeatably released and provide a metallurgical bond between steel and 6061-T6 aluminum as well as all aluminum structures.

The basic triggering mechanism utilizes a discharge annulus with a space defined by an inner and outer surface. The space is evacuated to form a vacuum or partial vacuum which creates a disconnect to prevent premature discharge of the capacitor banks during the charging cycle. Once the capacitor banks reach a preset voltage level on the order of several thousand volts, a spark initiator creates a plasma which allows current flow, which can be as high as 1,000,000 amperes. The current flows through a multi-segment, electrically continuous helical coil inductor which creates an intense magnetic field perpendicular to the current flow. The magnetic field sets up strong reactive eddy currents on the surface of the aluminum tube inside the inductor. The eddy currents create a magnetic field of their own which happens to be of the same magnetic pole as the inductor magnetic field. Since magnetic fields of the same pole repel and the inductor has a has sufficient strength to resist the forces generated by the pulse, the driveshaft tube accelerates inward toward the stationary steel or aluminum component. The impact is sufficient to create a metallurigical bond between the materials as evidenced by a scanning electron microscope image of the steel/intermetallic region/aluminum microstructure. Two Energy Dispersive X-Ray Spectrometer (EDS) techniques were also employed to determine if welding/diffusion had occurred along the steel/aluminum interface. The first EDS method used x-ray dot mapping to form an elemental image, pixel by pixel, of the interface region. The evidence of a multi-elemental region of aluminum and iron of at 7 microns or more was detected. The second technique employed in our investigation was a linescan analysis which plots the relative intensity/counts of each element along the scanned surface. Intermetallic regions of 10-20 microns were detected at the weld interface.

Sensing of Gas Metal Arc Welds

T.A. Siewert, R.B. Madigan, and T.P. Quinn
National Institute of Standards and Technology,
Boulder, Colorado

Improvements in electronic equipment for data capture and processing are opening new opportunities for more intelligent welding equipment. Specifically, welding knowledge is being embedded into control systems, so the welding equipment is able to make intelligent decisions about the weld quality and, ultimately, to keep the parameters in the range where only high-quality welds are produced. This paper discusses some of the knowledge about gas metal arc welding that has been reduced to a form suitable for electronic control.

There are two general areas in which electronic control systems can improve weld quality:
- in semiautomatic welding, control systems can improve weld quality by detecting and correcting problems faster than a human, and
- in fully automatic welding, control systems can emulate the eyes and ears of a skilled welder, going beyond the present reliance on the inherent stability of gas metal arc welding

However, a control system cannot function without process information. Sensing systems provide this information by mimicking the welder's senses and judgement. The wider the range of sensory data, the better the ability to maintain the weld quality. Continual improvements in sensing strategies and sensing systems are following from the sequential improvements in electronic equipment.

Among the various sensing strategies, through-the-arc (using the arc current and voltage signals) or imaging (cameras with image processing) strategies are most common in commercial systems. Both of these sensor strategies are well developed for seam tracking(1,2). This report shows the ability of through-the-arc sensing to provide additional data on shielding-gas quality and contact-tube wear(2-3).

Electrical signals serve as a very reliable sensor technique, with further advantages of being inexpensive and nonintrusive. When combined with a sensor of arc light intensity, the current also provides information about the arc length and the individual droplet transfer events(4,5). Other sensing strategies (airborne ultrasound, plateborne ultrasound, and imaging in the infrared range) have been shown to provide information about shielding gas quality, weld penetration, and flaw formation(6-8). All these sensing strategies are the subject of continuing studies, and additional discoveries are expected.

A key component of sensor development is confirming that the sensor data are being interpreted correctly. Often the electrical signals that are gathered during welding are compared to later examination of the weld bead surface, microstructure, or macrostructure (such as by NDE techniques). For correlation to events in the arc, the electrode can be imaged during welding by using a low-power laser to illuminate the arc from the back, then the arc light can be filtered with a narrow-band filter centered on the laser wavelength. The result is a shadowgraph of the electrode tip and the arc region. The image of the electrode tip and arc region can be recorded with a high-speed video for subsequent study and image processing at low-speed. We have found that by imaging at 1000 Hz, we are able to resolve all but the fastest transfer events (such as the transition between streaming spray and rotating spray transfer). Simultaneous recording of the electrical sensor data and the high-speed video data permits cross correlation of the data, assuring that the electrical sensor data is unambiguously related to actual welding events.

Figure 1 illustrates a method of detecting contact tube wear that uses the statistical variation in the voltage record over the life of the tube. The tube wear was confirmed with metallographic sectioning of the contact tube. The data were taken with a pulsed power source, so the voltage record for a new tube shows substantial variation, with a period near 0.01 s (the period of the pulsing). However, the envelope of data for the new tube (at the bottom of the figure) is consistent for periods near 1 s and does not contain sharp changes in voltage. In contrast, the voltage records for a worn tube (near the top of the figure) show substantial fluctuation over a period near 1 s. This is an example of a welding problem that a welder learns to recognize during
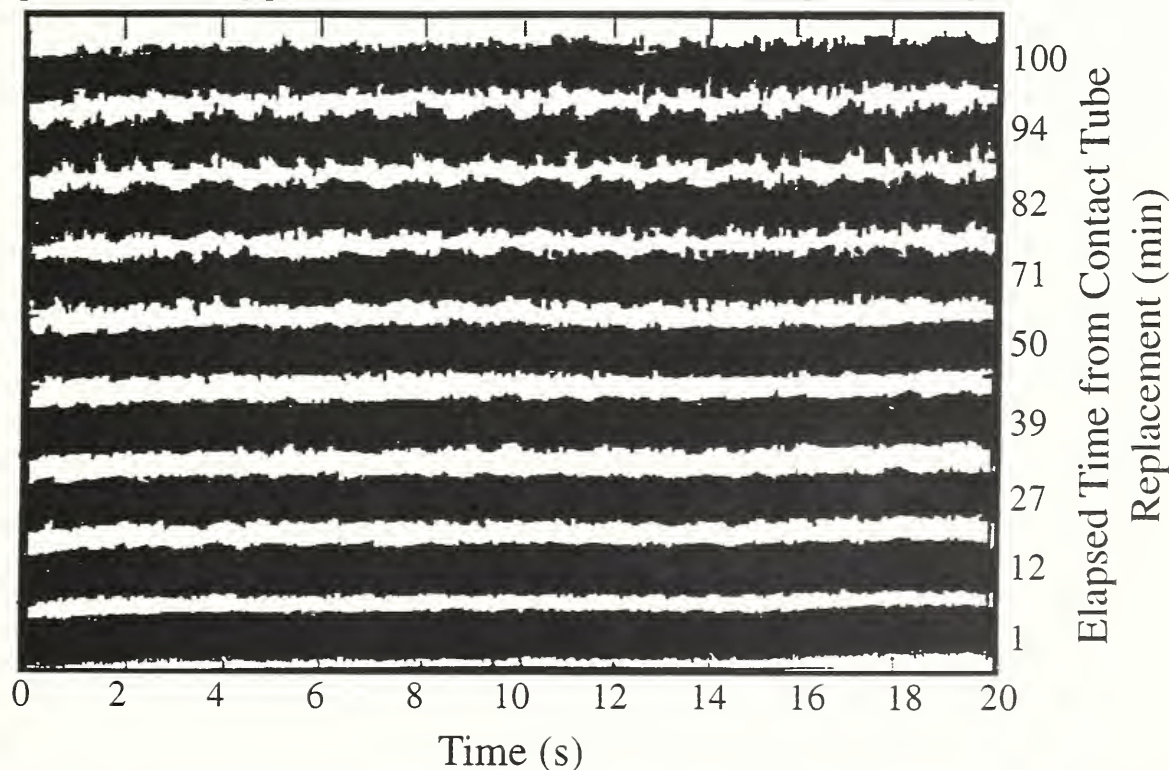
Fig. 1 - Sequential 20-s samples of the voltage record during the life of a contact tube. The envelope of the voltage variation becomes ragged as the tube wears.

apprentice training, but whose characteristics are difficult to define or measure without electronic data processing. The wear occurs in a nearly uniform manner over much of the contact tube life (often near one hour or more), but the signal characteristic of wear is cyclical over a period of several seconds, as either the cast and helix of the electrode cause the electrode to rub against various surfaces of the contact tube, or as stick/slip occurs at the electrode/tube boundary. The wear appears as a gradually increasing fluctuation in the voltage signal (with a constant-current power source; or in the current signal with a constant-voltage power source). Detection of the pattern shown in Figure 1 was reduced to a computer program and implemented on a personal computer. The program has been shown to monitor wear of contact tubes for a variety of welding parameters and contact tube alloys (3).

Degradation of shielding-gas quality is another welding problem that the apprentice welder learns to recognize, but whose characteristics are difficult to quantify. When the uniform flow of shielding gas around the arc is disrupted by cross drafts or spatter in the gas nozzle, weld appearance or properties degrade; in extreme cases, porosity develops. The effect can be sporadic when caused by air flow through an open door or persistent when caused by spatter in the gas nozzle. We reduced the characteristic pattern in the electrical signals to a computer algorithm, enabling successful automatic detection of the problem.

Melting of the contact tube in gas metal arc welding is another common problem. We have found that the period between interruption of the electrode feed in gas metal arc welding and melting of the contact tube is on the order of 0.1 s (9). Figure 2 shows that today's control technology can detect the wire (electrode) feed interruption and shut down the welding power before melting occurs, typically with reliability and speed better than those of a human. Other events, such as droplet transfer, often occur in intervals of less than 0.01 s, which require a response time of less than 0.001 s for active
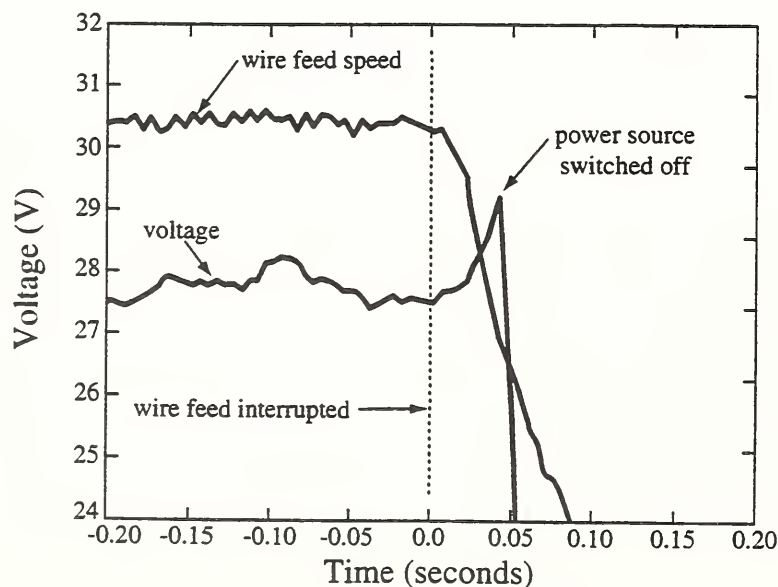
Figure 2. Voltage signal variation after the electrode feed is interrupted. Unless the power is switched off (here at 0.05 s), the contact tube will melt after about 0.1 s.

control. This is a challenge for today's low-cost control systems, but will become easier as computer technology continues to advance.

The arc length controls the power or heat input to the weld, while arc light intensity is determined by the volume of the plasma and its temperature. Since temperature remains fairly constant over much of the welding range in GMAW, the most important parameters in determining the arc light intensity are the diameter and length of the arc. Thus arc length can be determined from the diameter and the light intensity of the arc.

The arc diameter is determined primarily by the current, which can be monitored by a nonintrusive Hall-effect transducer located anywhere on the power cable. The arc light intensity can be monitored by a simple photodiode located in line of sight of the arc. These two sensors can be combined in a control system to maintain the desired arc length. A parametric study of different optical filters and light sensor positions has shown that a blue-band filter was suitable for controlling the arc length to within a one-standard-deviation range of 0.7 mm in a set of 30,000 measurements. Figure 3 summarizes the data from these experiments, covering arc lengths from 5 to 15 mm.
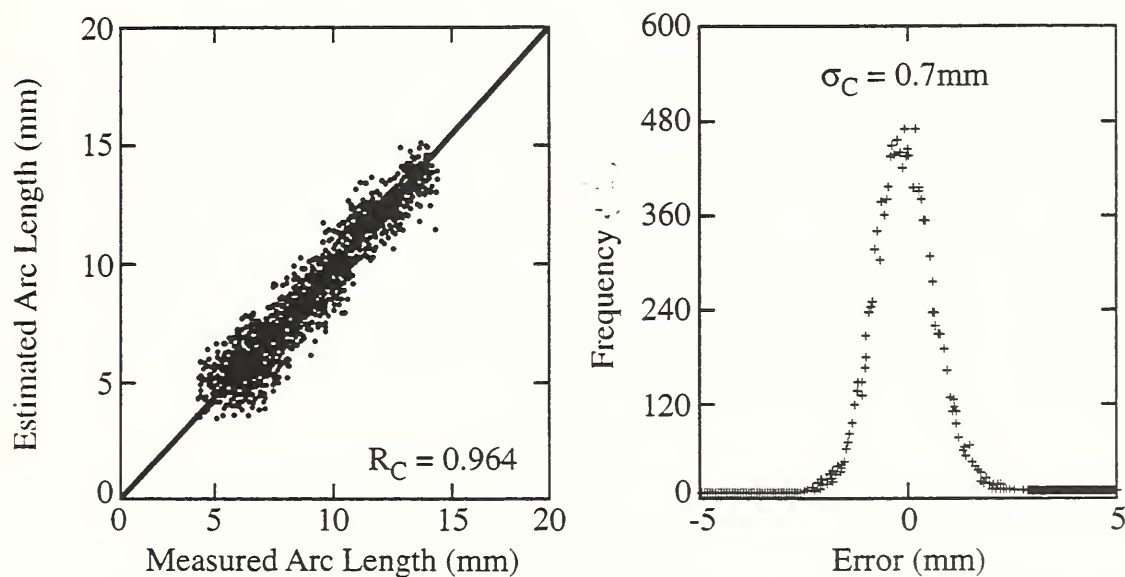
Figure 3. Estimated versus measured arc length and error histogram resulting from using a BBP (blue-band pass) optical filter.

We built a control system around this arc-length measurement technique. As a test of this controller, a GMAW weld was produced up a 10 mm high steel ramp. Figure 4 shows that without the control system, the arc length dropped from 6.5 to 2.5 mm. When the control was activated, the arc length remained steady near 5 mm, changing by less than 1 mm.
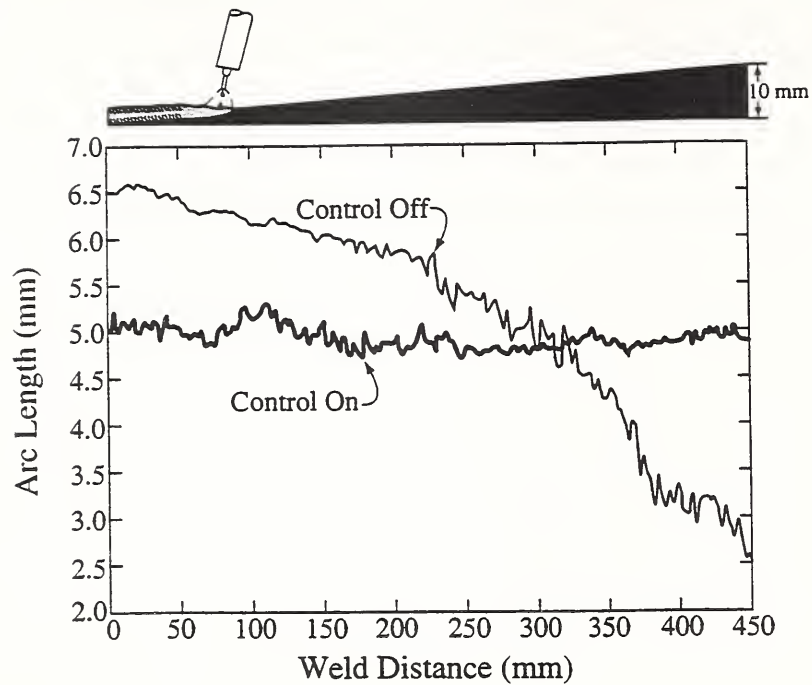
Figure 4.   Measurement of arc length while welding up a 10 mm high ramp, with and without arc length control activated.

In addition, the arc-light sensor has a rapid response, fast enough to detect individual droplet transfer events in a gas metal arc weld. In this case, detachment of individual droplets modulate the arc length, often in the frequency range near 200 Hz. High-speed video images confirm that the sawtooth shape seen in the photodiode output correlates to repetition in the incremental advance of the electrode tip. The electrode tip advances until the molten droplets detach. Figure 5 shows this correlation between the video images and the photodiode output, calibrated in arc length. This information has been used to develop a real-time monitor of the droplet transfer rate, where fluctuations in the transfer rate indicate instabilities in the arc.

In summary, through-the-arc sensing has been shown to detect common welding problems such as loss of or contamination of the shielding gas or blockage of the wire feed. Through-the-arc sensing can also provide process information on arc length and droplet transfer rate. In these two cases, the appropriate response to deviations is known and a control system has been developed and demonstrated.

There are also opportunities for further improvements in mimicking a human's ability to identify and correct welding problems. Here, we need to understand how a trained welder responds to difference conditions and then reduce those rules to an algorithm suitable for an automatic system. Human response times are on the order of 1 s, especially after a long, uneventful period of welding. This duration is close to the solidification time of the weld pool, and therefore, it is tolerable in manual welding (10). If human response times were slower, manual welding never would have been possible. For much faster events, only automated systems can react in time.
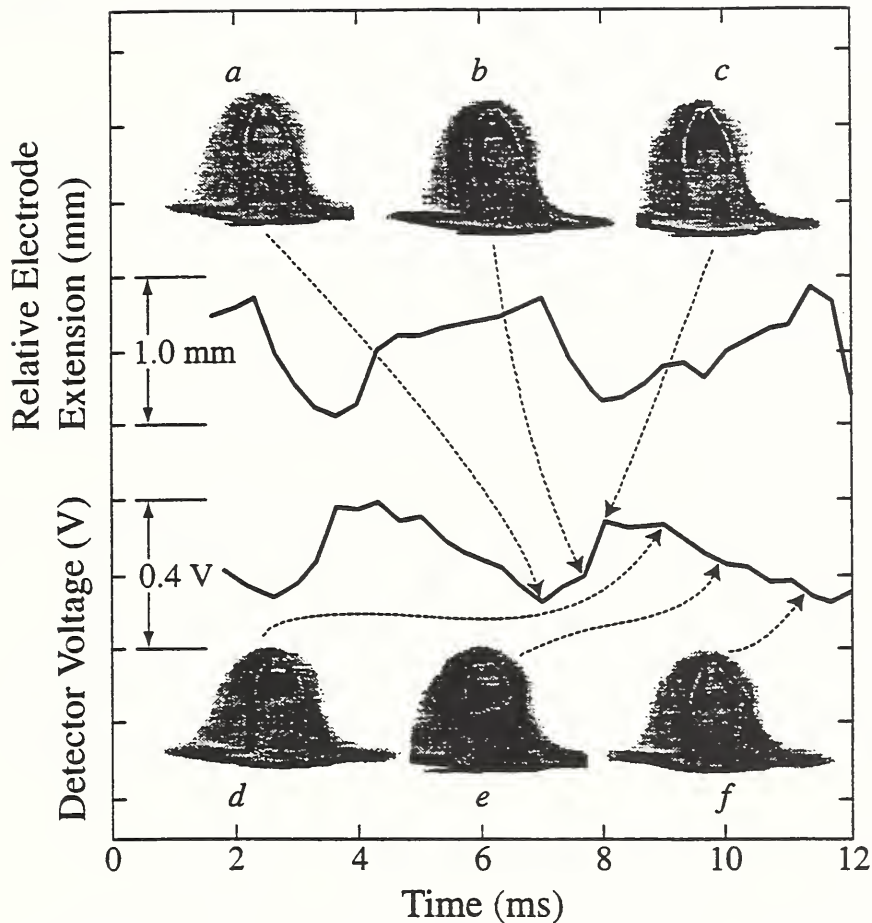
Figure 5. Processed video images showing how droplet frequency can be measured from the arc light intensity.

References

1.   Agapakis, J.E., Katz, J.M., Koifman, M., Epstein, G.N., Friedman, J.M., Eyring, D.O., and Rutishauser, H.J., 1986, "Joint Tracking and Adaptive Robotic Welding Using Vision Sensing of Weld Joint Geometry," *Welding Journal*, 65 (11), pp. 33-41.

2. Richardson, R.W., 1986, "Robotic Weld Joint Tracking Systems – Theory and Implementation Methods," *Welding Journal*, 65 (11), pp. 43-51.

3. Quinn, T.P., Madigan, R.B., Mornis, M.A., and Siewert, T.A., 1995, "Contact Tube Wear Detection in Gas Metal Arc Welding," *Welding Journal*, 74 (4), pp. 115s-121s.

4. Control of Gas-Metal-Arc Welding Using Arc-Light Control, R.B. Madigan, T.P. Quinn, and T.A. Siewert, National Institute of Standards and Technology Interagency Report 5037, 1995.

5. Siewert, T.A., Madigan, R.B., Quinn, T.P., and Mornis, M.A., 1992, "Through-the-Arc Sensing for Monitoring Arc Welding," *Proceedings of 3rd International Conference on Trends in Welding Research*, eds. David, S.A., and Vitek, J.M., ASM International, pp. 1037-1040.

6. Matteson, M.A., Morris, R.A., and Raines, D., 1992, "An Optimal Artificial Neural Network for GMAW Arc Acoustic Classification," *Proceedings of 3rd International Conference on Trends in Welding Research*, eds. David, S.A., and Vitek, J.M., ASM International, pp. 1031-1036.

7. Song, J.-B., and Hardt, D.E., 1993, "Closed-Loop Control of Weld Depth Using A Thermally Based Depth Estimator," *Welding Journal*, 72 (10), pp. 471s-478s.

8. Carlson, N.M., Johnson, J.A., Larsen, E.D., Van Clark, A., Schaps, S.R., and Fortunko, C.M., "Ultrasonic Sensing of GMAW: Laser/EMAT Defect Detection System," *Proceedings of 3rd International Conference on Trends in Welding Research*, eds. David, S.A., and Vitek, J.M., ASM International, pp.859-863.

9. Siewert, T.A., R.B. Madigan, and T.P. Quinn, presentation at 1993 AWS Annual Meeting (1993).

10 Cook, G.E., K. Andersen, and R.J. Barrett, "Recent Trends in Welding Science and Technology," ASM International, Materials Park, Ohio, 891 (1990).

Dynamic Magnetic Compaction (DMC) Technology
For High Density/Net Shape P/M Parts

J. P. Barber, D. P. Bauer, B. Chelluri
IAP Research, Inc., 2763 Culver Avenue, Dayton OH 45429-3723

W. F. Jandeska
General Motors Powertrain, 895 Joslyn Avenue, Pontiac MI 48340-2920

T. Grady and D. Score
Delphi Automotive Systems, 7601 East 88th Place, Indianapolis IN 45256

T. Cadle and J. Mandel
Zenith Sintered Products Inc., N112W18700 Mequon Road, Germantown WI 53022-8209

Introduction

The automotive industry is constantly striving to reduce cost, weight and emissions while improving fuel economy and NVH (noise, vibration, harshness). Improved efficiency combined with reduced vehicle size and weight require gears with high power density. Powertrain gears are the single most expensive component in automotive transmissions. The total value of transmission gears in the United States alone is greater than $1,000,000,000 per year. High performance gears with superior mechanical properties at a lower cost are hence constantly sought. The performance needs include enhanced tooth bend fatigue strength, higher surface contact fatigue and the tight dimensional control such as concentricity, tooth involute and minimum lead error and run out.

Today, most gearing is produced by machined cast, forged or wrought gear blanks and is the only process of choice for providing the required power density, the precision geometry and high performance. Conventional P/M can not achieve the required dimensional control or the material density to provide the required mechanical properties. These gears are significantly more expensive than the ones made by conventional Powder Metallurgy (P/M) method. Conventional P/M processes yield density in the range of 7.2-7.3 g/cc. Forged P/M processing is not a viable option because of its inability to provide sufficient dimensional accuracy.

The importance of achieving high density in P/M is well established. Figure 1 shows the typical dependence of strength and ductility on density for a steel powder. Improved power density in gears via P/M requires net shape fabrication of a conventional, axial pressed and sintered nickel steel (FN 04 xx) gears at a density of at least 7.6 gms/cc.

Dynamic Magnetic Compaction (DMC), a new P/M compaction technology being developed under an ATP program, has demonstrated that the high density ranges of 7.5-7.7
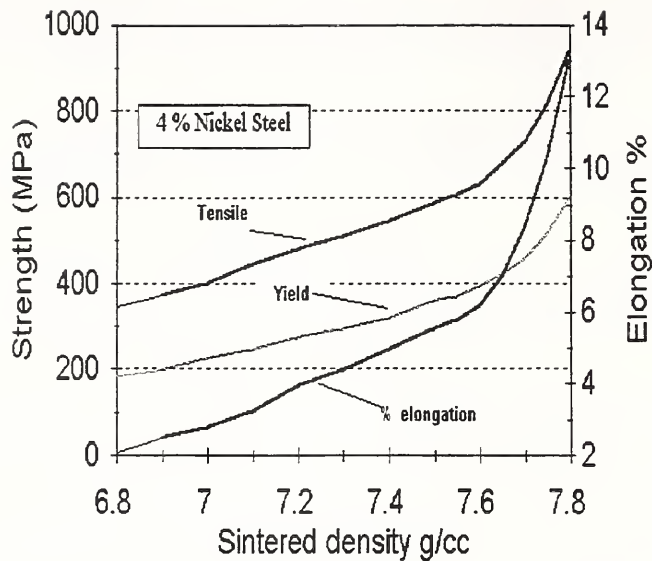
Figure 1. Strength of conventional, axial pressed and sintered nickel steel powder as a function of density.

g/cc in steel powders is attainable. The net shape capabilities of the DMC process are being investigated and detailed testing is being carried out. We plan to insert DMC powder compaction technology into the established, low cost conventional or standard (STD) P/M process. The present STD P/M and the envisioned DMC P/M processes are compared in Figure 2. Each process starts with the selection and mixing of powders. Either prealloyed or elemental powders may be used and additives may be included to facilitate part extraction, sintering etc. The powders are then loaded into die and compacted. The density obtained during compaction is critical to final product with almost all properties improving with higher density. The pressed "green form" is then sintered
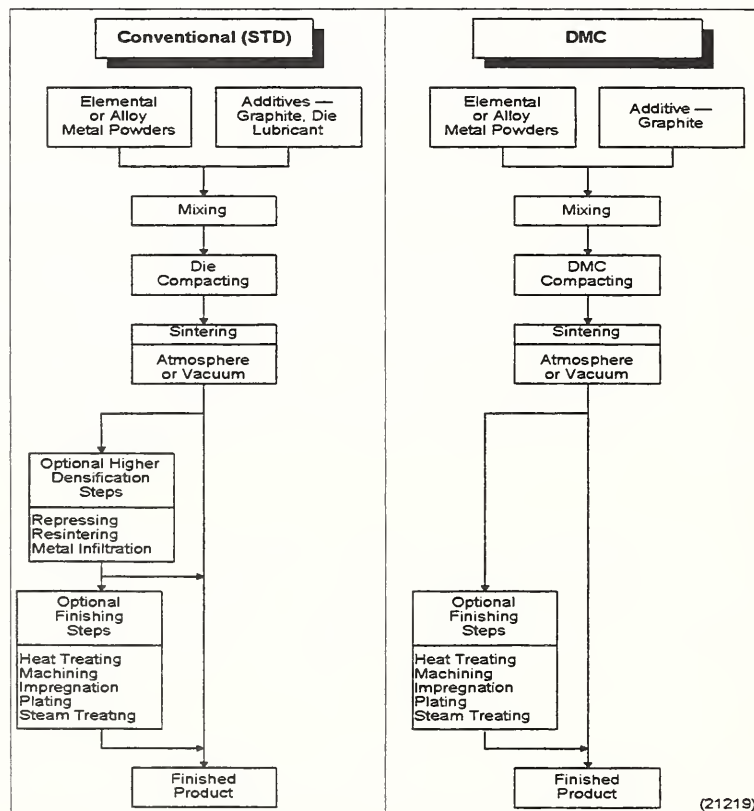
Figure 2. Conventional and DMC P/M Processes.

360

to promote inter particle bonding which establishes the mechanical properties. The STD P/M process includes optional steps for increasing density; DMC achieves high density in a single step. Optional finishing steps may be used for some parts in both processes.

What is Dynamic Magnetic Compaction?

DMC is an IAP patented technology for powder compaction which uses pulsed magnetic pressures to densify powders as conceptually illustrated in Figure 3. The compaction occurs in less than 1 milli second and ongoing results show that the compressibility of steel as well as other powders is enhanced under these conditions. We demonstrated that 4400 steel alloy powders can be compacted to densities of 7.6 g/cc in a single pulse.

Magnetic pressures can also be applied in configurations, such as radially in ward or outward, which cannot be duplicated in standard pressing, thus promoting densification in many useful cylindrically symmetric shapes not now possible. Radial compaction is ideally suited for compaction of high length to diameter (L/D) ratio specimens. DMC technology does not require lubricants in the powders, thus subsequent burn off problems do not exist. Cold isostatic pressing also radially compacts, but the "dynamic" nature of the process enhances the density results significantly.
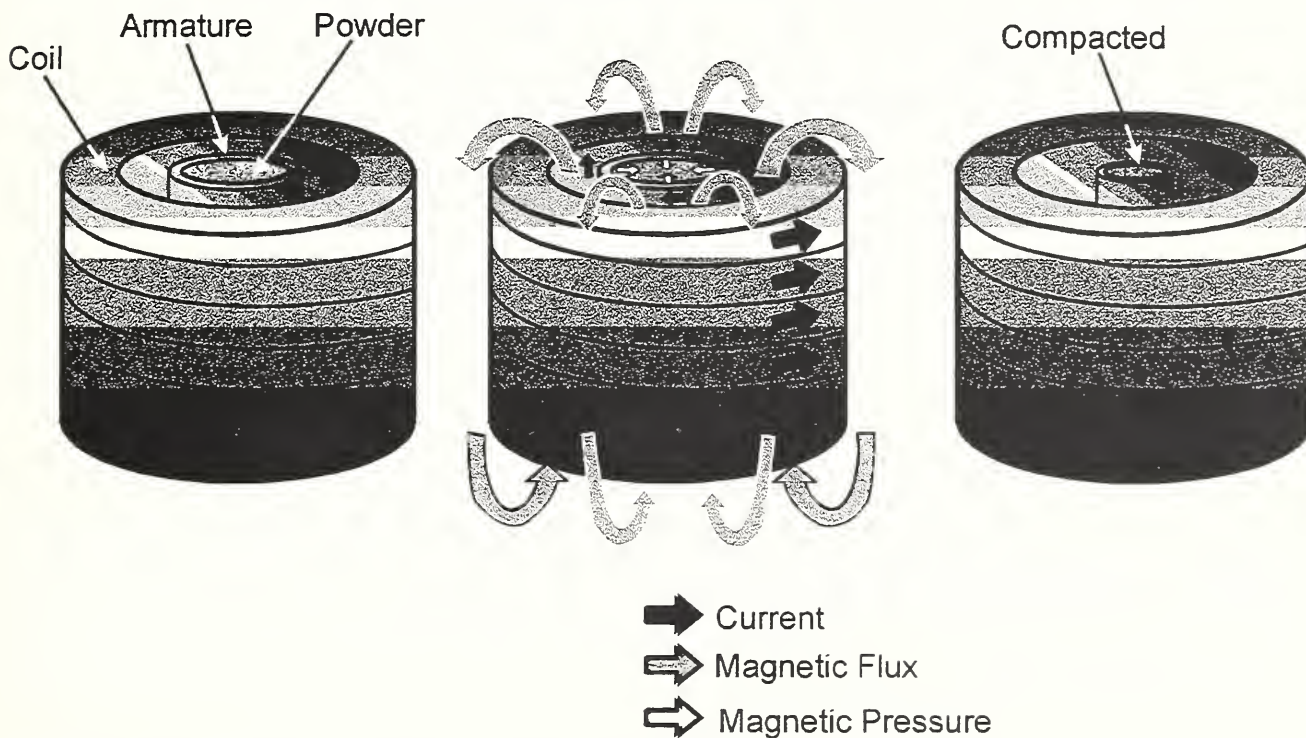
Figure 3. Dynamic Magnetic Compaction Process

Results

In the current ATP project , we are developing two DMC products: a new process for forming powder metal into very high density net shape compacts/parts; and a new machine to

implement densification process. We are on schedule to demonstrate the technical and economic viability of the DMC equipment and process to make gears and magnetic components. The following sections summarize the results in the specific areas.

DMC Process Results

Figure 4 shows the density enhancement of DMC process as compared to conventional die press. Mechanical property data on test pieces and gear samples is in progress. The net shape fabrication results are being gathered at the present time.

22447
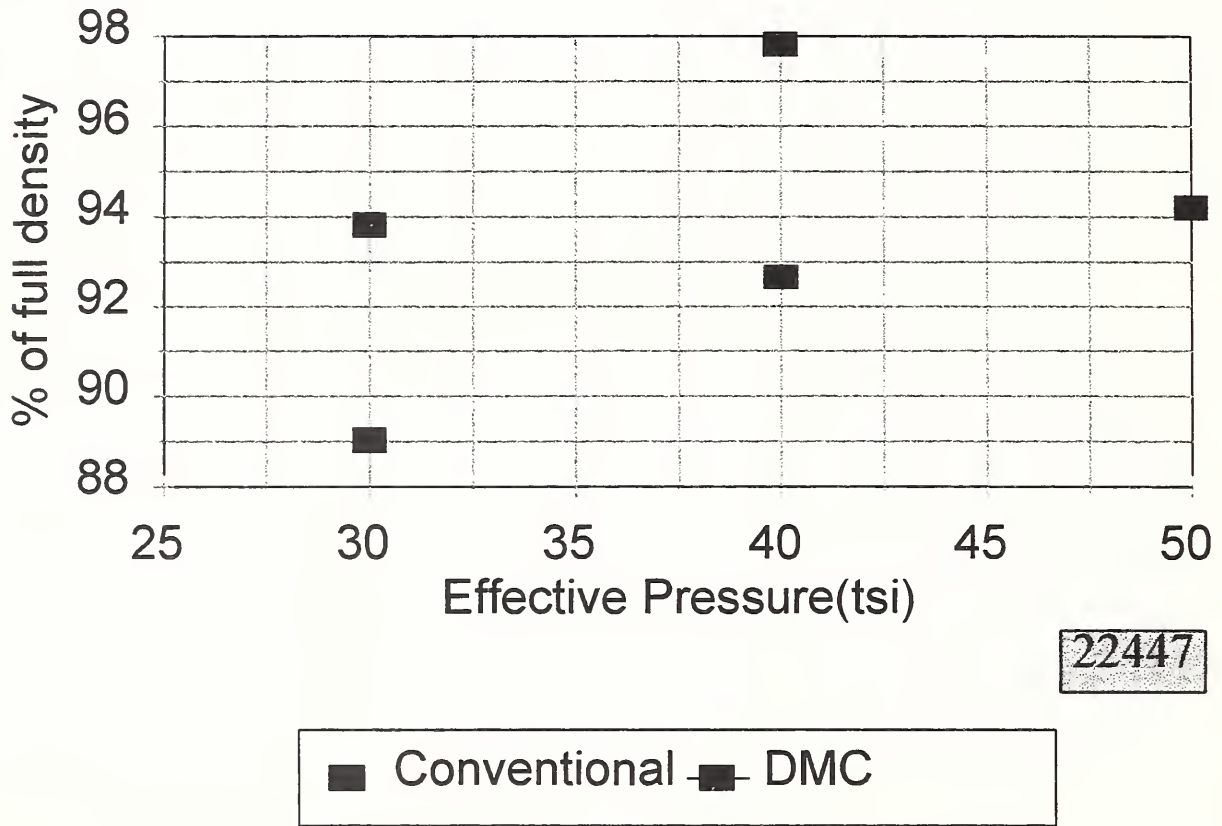
Figure 4. **Compressibility of Steel Powder.**

DMC Equipment Results

At the conclusion of Phase I, IAP will build a prototype DMC machine for installation at Zenith Sintered Products, Inc facility. The progress on prototype equipment is according to plan and on schedule. We have a concept design for production cell. We have also developed power supply designs for commercial DMC systems to handle the demands of the high rates of auto industry. We have developed for the first time designs for high field coils (40 Tesla) that can handle high volume demands. The production rates of 4/minute have been demonstrated with DMC equipment and production rates of 15/minute can be achieved with the new power

supply designs.

Cost Studies

Over all process economics for DMC technology were developed with the key elements such as pulse power equipment, magnetic tooling, armatures and other relevant operating costs. The comparison of the cost with the competing processes are being made. The preliminary production cost estimates show that DMC is highly cost competitive for manufacturing high power density transmission gears if the dimensional accuracy in gear teeth can eliminate current machining operations.

Commercialization

Our long term goal is to have broad and growing application of the DMC technology to an ever wider market where part performance and net shape processing are important. Obviously, with the ever present auto industry emphasis on lighter weight and lower cost, the automotive market potential goes beyond gears and ferrous based materials. Furthermore, like the rest of the P/M industry, (where automotive represents about 70% of the market), DMC will find non automotive applications, especially where machined castings and forgings are used. The path to this market potential requires one step at a time.

The current ATP project is the first phase of an envisioned three phase commercialization process. Our plan for meeting first phase goals is for IAP to build a prototype DMC machine and use the system to fabricate gears for testing at GM. Joint Venture partners will work closely during this phase to iron out any problems with equipment and adjust the processing steps to simulate production "like" conditions of manufacturing. This machine will be used to identify improvements in the construction and operation of future production machines. Gear testing at GM will allow the process to gain credibility with product release engineers at GM. At the conclusion of Phase I we will have complete data to evaluate the potential of the DMC process to manufacture gears.

Following a positive assessment of DMC viability by all joint venture partners, the second phase of commercialization will enter the established auto industry procurement/production cycle. Success in Phase II will provide high industry wide visibility. The third phase of DMC commercialization will be one of growth which will begin as IAP and Zenith sales engineers repeat the Phase II cycle but armed with production proven results from the first product.

Conclusions

DMC is a promising new P/M compaction technology which appears to have the potential to produce high density materials with excellent properties. We plan to bring this innovative P/M technology to the prototype stage at the end of ATP project. This technology has the potential to become the next generation powder metal (P/M) process offering the automotive industry (and others) a new class of products in between current P/M technology

and that of forged product, a nitch that provides the opportunity for weight savings with significant cost reductions. We are in the process of doing this by inserting Dynamic Magnetic Compaction (DMC) into the standard, low-cost, Single Press Single Sinter (SPSS) P/M process. This will permit us to produce net shape products with less than ½ the porosity of today's P/M products at a cost well below today's high density forged and wrought products. The successful introduction of this technology will benefit our Joint Venture Partners, the United States automobile industry, and the nation.

Session VIII. Optics and Ultrasonic Techniques

Direct Diode Laser Systems for Machining/Materials Processing
A. Schoenfelder and R.J. Lang, SDL; D. Gherasim, Utilase

Non-contact Film Thickness Gage
J.S. White, F. P. LaPlant, and J.W. Dixon, Autospect

Fast Volumeteric X-Ray Scanner for Three-Dimensional Characterization of Critical Objects
F. Hopkins, SMS; A. Pfoh, and R. Isaacs, General Electric; and J. Hetzner, GM

Direct Diode Laser Systems for Machining/Materials Processing

A. Schoenfelder, R.J. Lang, and D. Gherasim*
SDL Inc., 80 Rose Orchard Way, San Jose, CA 95134
*Utilase Systems, Inc., 20201 Hoover, Detroit, MI 48205

Introduction

Because of high efficiency, reliability, fiber-optic beam delivery, and compact size semiconductor diode lasers have been the preferred source for many optical systems applications. In recent years, continued improvements in high-power, high-brightness diode laser arrays have enabled new applications for diode laser sources in areas such as soldering, plastic welding, epoxy curing, heat treatment of thin metal parts, and sintering for rapid prototyping. Improvement in diode laser sources is occuring at the device design and materials level to increase reliable power output levels, and in advanced optical methods for beam combining the power outputs of more than one diode laser array, while maintaining acceptable brightness.

Figure 1 presents a log-log plot of power versus Etendue (size-divergence product) that provides a high-level view of the applications space for laser sources and material processing applications. The diagonal lines indicate the lines of constant brightness. The shaded regions in the upper right section of the plot correspond to the indicated application requirements for aluminium/steel cutting, aluminium/steel welding, and heat treatment that are currently served by industrial CO_2 and lamp-pumped Nd:YAG lasers. The individual points indicate the positions of many commercially available diodes. The grey points denote performance of the direct output beam, whereas the black points correspond to the performance of the associated device when the output beam is fiber-coupled. Elongated grey-shaded regions denote arrayable modules with potential for brightness conservation. As can be seen, diode laser systems have the potential to directly compete in applications that are currently served by gas and solid-state laser systems as well as by conventional material processing technologies. Moreover, diode laser systems can even enable new processes in so-called hybrid systems which consist of conventional material process technologies combined with a diode laser system.

High Power CW Diode Lasers for Solid-State Laser Pumping

Although the principal subject of this paper is direct-diode applications, the importance of diode pump arrays to industrial solid-state lasers is of such potential significance that subject must be addressed. For example, designers of solid-state laser systems have been actively involved developing laser diode pumped high-average power, solid-state industrial laser systems capable of producing multi-kilowatt power outputs that still retain the high beam quality unique to diode-pumped laser systems. Precision laser machine tools based on diode laser pumping have been designed and are now under development. The extent of use of such systems will be limited only by the cost and reliability of the laser pump arrays.

Advanced diode pump arrays with increased lifetime and lower cost are being developed by SDL under support of the DARPA PLM program [1]. The advanced package architecture is designed to support an average of 120 W/cm^2 operation under cw or pulsed operating conditions up to a maximum average power of 240 Watts for an array comprised of sixteen 20 W bars. Tests on prototype arrays show lifetimes that extrapolate to greater than 10,000 hours.

The advanced package design has the potential to offer low-cost (≤$25/W) provided the production volumes are sufficiently high. Such volumes can be projected from a substantial replacement of lamp-pumped industrial solid-state lasers by diode-pumped solid-state lasers. However, the rate of replacement of lamp-pumped lasers by diode-pumped lasers is controlled by the price sensitivity of the application. Therefore, markets with a tolerance for higher diode prices are engaged first; and as the annual diode pump array volume continually builds, prices come down and applications markets with greater price sensitivity will be entered. Replacement of low power (up to ~10 W) lamp-pump solid-state lasers by diode-pumped solid-state lasers is well under way, and the laser industry is on a trajectory to initiate introduction of higher power (≥ 50 W) diode-pumped solid-state lasers.

Direct Diode Stacks for Heat Treating Applications

Currently, many surface heat treatment applications use multi-kW Nd:YAG laser with a spot size of a few mm and a scan speed of ~200 cm/min. The main shortcoming of the present Nd:YAG laser system is the depth of focus and in some cases the size of the laser system itself. Surface heat treatment with presently available Nd:YAG laser systems is labor intensive because the stand-off distance must be constantly regulated to keep a constant intensity on the surface of the part. Having a large depth of focus (~1 cm) is important because it would accomodate details of the process associated with the vagueries in surface morphology, part orientation, and path complexity, with a minimal adjustment of the stand-off distance. Also, being able to vary the format of the spot, as well as having large spot sizes (~1 cm^2) is very desireable. Finally, the heat treatment unit should have the capability to do die repair, which requires about 3kW focused into ~2 mm spot.

Heat treatment using a direct diode laser system would offer a more compact system relative to the multi-kW solid-state lasers currently employed, which is an advantage. To the extent that the ~1 cm stand-off distance can be realized with direct-diode systems, further advantage is possible. High power diode laser arrays for heat treatment applications have been realized with power outputs of 400 W in a 2x2 cm aperture and a 290 mrad by 20 mrad divergence of the emitted beam.

Increased Power Output and Reliability of Semiconductor Diode Lasers

Although diode laser arrays have demonstrated cw power output capabilities in excess of 100 W for a 1 cm wide bar, the reliable cw power output level of commercially available 1 cm wide bars is typically 20 W for a median lifetime of >10,000 hours. The limitations on reliable power output level are primarily due to thermal management, but at higher power levels (~1 MW/cm^2) catastrophic facet damage is the limiting factor. Note that the >10,000 hour median lifetime is arguably the accepted industry standard for commercial 15-20 W cw diode laser arrays, though there as of yet are no formal industry standards for high-power diode laser arrays.

The unit cell of the cw bar is the broad-area stripe laser. At present, commercially available 100 μm wide broad-area stripe lasers provide 1 W cw power outputs with a median lifetime of 400,000 hours at 25° C operating temperature. Recent research and development at SDL efforts have focused on improved device design and material quality have given rise to 100 μm broad area lasers operating at 2.7 W and 60 ° C that under initial lifetesting (see data in Fig. 2) extrapolates to a median lifetime of 50,000 hours [2]. Incorporation of these design and material improvements

into 1 cm bar could lead to power output levels of 40 W or possibly more; but in order meet the 10,000 hour requirement, the cooling of the bar must be improved to efficiently remove the increased thermal power dissipated in the diode array at these high power levels.

Conversion efficiencies of 1 cm diode arrays are typically abut 40%, but can range from as low as 30% to as high as 50%. This same sort of efficiency range should be expected from high power (>20 W) bars as well. Therefore, if one doubles the power from 20 W to 40 W while maintaining the same conversion efficiency, the thermal power dissipated in the bar increases by a factor of two. In order to maintain a constant active temperature rise going from a 20 W to a 40 W bar, the 40 W bar must be mounted in a package which has a thermal resistance that is a factor of two lower than current commercial packages. Maintaining the bar active layer temperature within acceptable limites is critically important for meeting the >10,000 hour lifetime requirement.

High-power high-brightness direct-diode sources

One of the most recently developed high-power, high-brightness direct-diode sources included in Figure 1 is the lensed 2-dimensional array of angled-grating distributed feedback (α-DFB) laser diodes [3], which could one day possibly evolve to produce brightness levels required for industrial metalworking applications. The α-DFB laser is an advanced-design semiconductor laser comprised of a broad-area stripe with an intracavity grating that provides both spectral and spatial mode discrimination. The grating is oriented at an angle (refered to as α) with respect to the facets so that it provides both longitudinal (spectral) mode control and lateral or spatial mode control. Single α-DFB emitters have demonstrated 1 W cw power output with a lateral beam divergence of 0.29° (5 mrad) as compared to ~12" for a conventional broad-area stripe laser. Linear monolithic arrays of α-DFB lasers have demonstrated power outputs as high as 11 W cw with diffraction-limited emission from each emitter. Output powers up to 20 W cw have been achieved with a 3x diffraction-limited operation of each emitter across the array [4].

Most recently a lensed two-dimensional array of 10 α-DFB bars has been fabricated. A picture of this diode laser array (operating at ~25 W cw and focused to ~400 μm spot) cutting a razor blade (0.2 mm stainless steel) is displayed in Fig. 3. This lensed array produced a collimated beam of 10.1 x 9 mrad from an aperture size of 1 x 1.7 cm. At a maximum operating power of 60 W cw, the α-DFB array yielded a brightness of ~0.4 MW/cm^2-sr. It is anticipated that the α-DFB stack could be improved to yield a brightness of ~7 MW/cm^2-sr. It should be emphasized that the α-DFB laser stack is still in development and only through continued future development of advanced prototypes deployed in beta site testing will it be possible to assess accurately the commercial potential for the α-DFB laser stack in the materials processing market.

[1] N. W. Carlson, L. Marabella, Reports on Topical Meeting of the Laser Society of Japan: Advanced Solid State Lasers and Their Applications, Nov. 19, 1996.
[2] S. O'Brien et al., accepted for presentation at LEOS 97, San Francisco, Nov. 10-13, 1997.
[3] A. Schoenfelder, S.D. DeMars, S. O'Brien, R.L. Lang, CLEO 97, Baltimore, May, 19-23, 1997: 20 W, high-brightness angled-grating DFB laser array, paper CMA1.
[4] A. Schoenfelder et al., accepted for presentation at LEOS 97, San Francisco, Nov. 10-13, 1997.
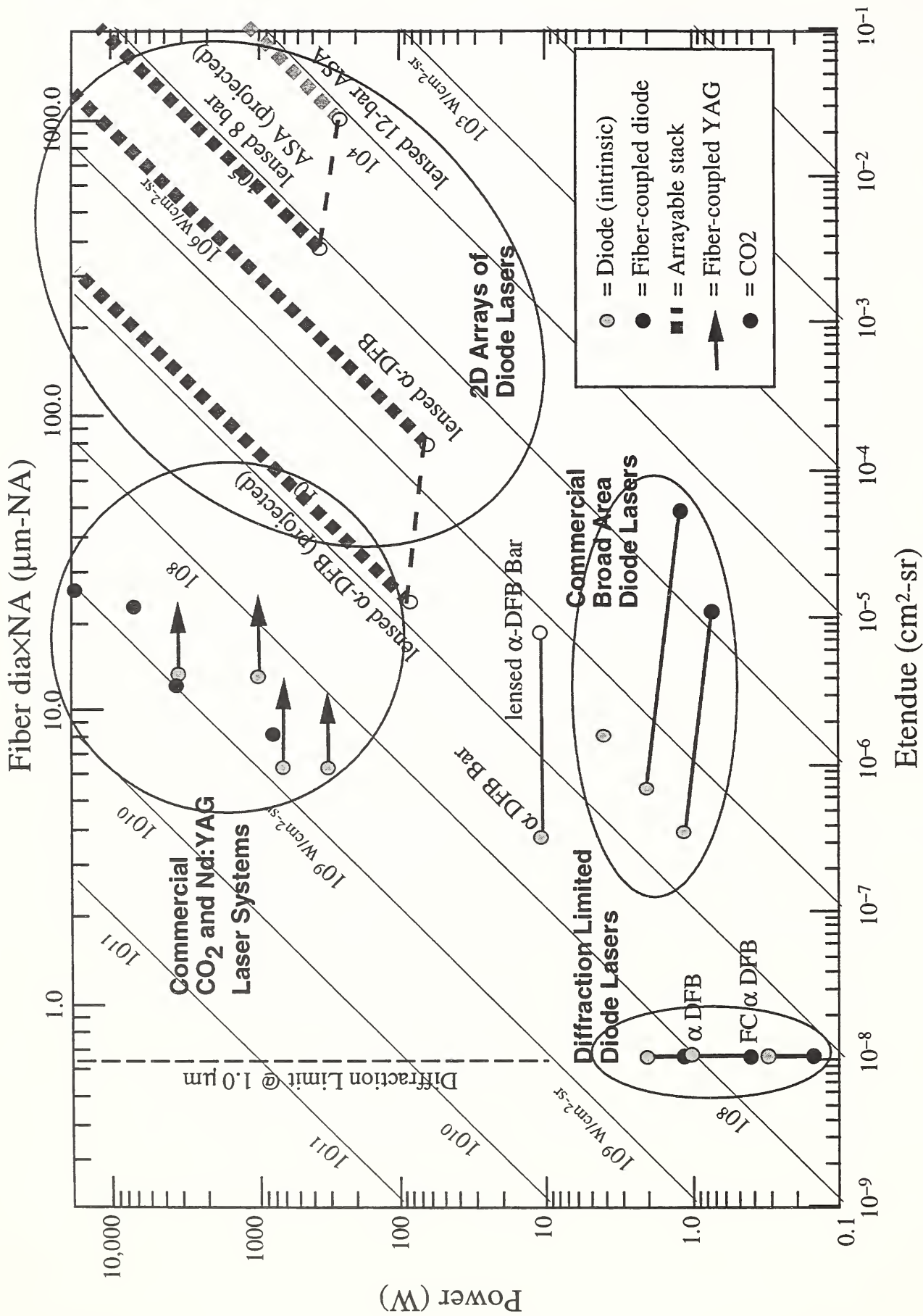
Fig. 1. Brightness, power, and etendue of laser diodes and systems.
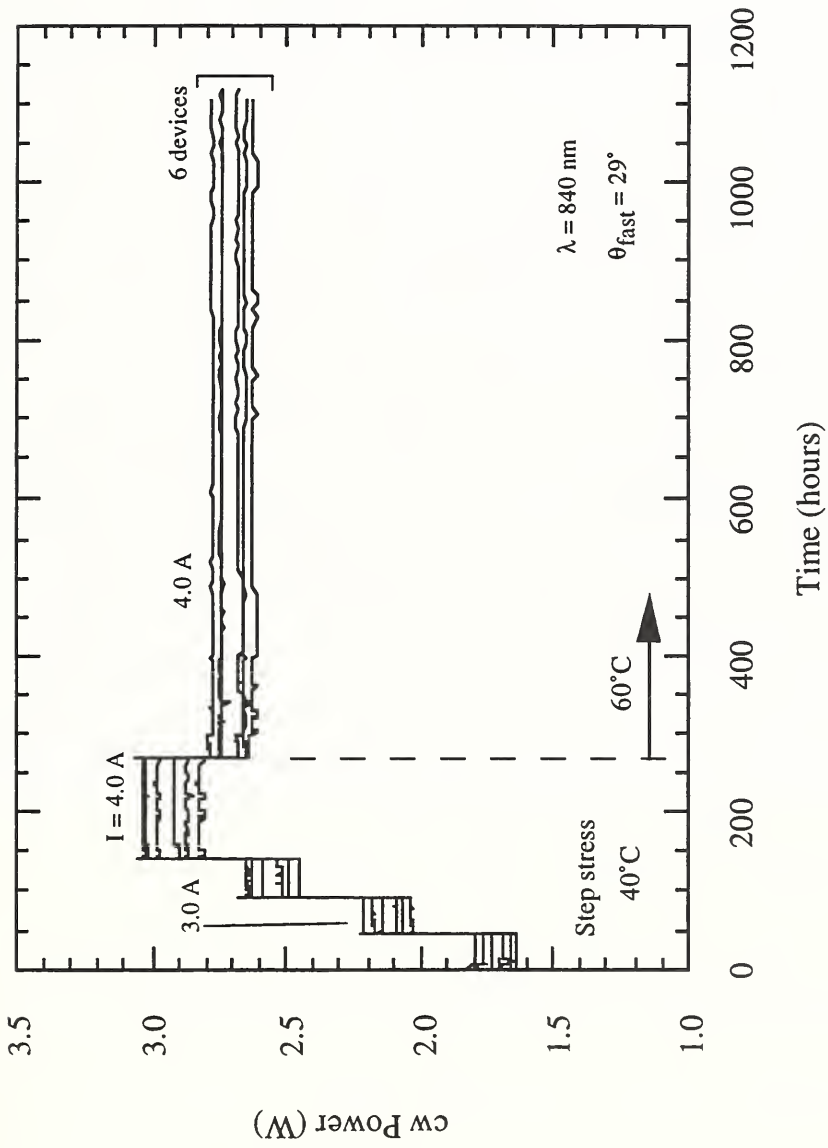
Fig. 2. Preliminary lifetest data of improved high-power 100 μm broad-area lasers.
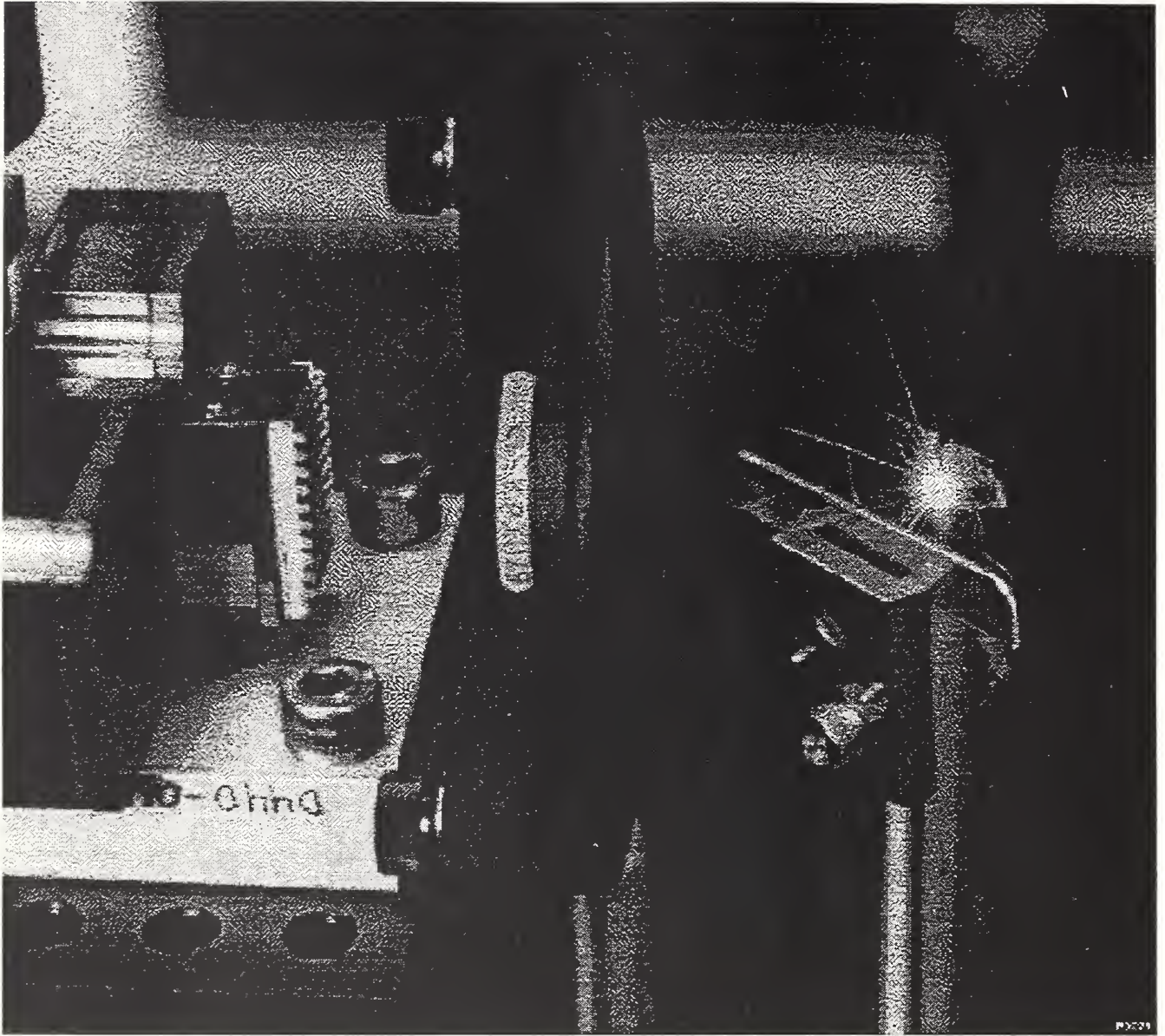
Fig. 3. Photograph showing an a-DFB laser array cutting a 0.2 mm thick razor blade.

Non-contact Film Thickness Gage

J.S. White, F.P. LaPlant and J.W. Dixon
Autospect, Inc.
4750 Venture Drive
Ann Arbor MI 48108

Introduction

One of the most complex and costly processes in the assembly of an automobile is the painting of the vehicle body. In addition to providing the desired appearance (e.g., color, gloss), the applied coatings are also required for body durability. Both appearance and durability of the films are directly affected by the thickness of the applied materials. Unfortunately there is insufficient control in the application process and film builds can vary widely on the same vehicle body as well as from vehicle to vehicle. The biggest problem with applying the coatings too thick is the simple material waste, probably exceeding $500 million per year for the U.S. Big 3 automotive manufacturers alone. Also, a coating that is too thick will often exhibit adverse appearance effects (e.g., orange peel) and lead to other film build problems (e.g., sags, pops). A coating applied too thin will result in insufficient durability (e.g., corrosion protection, chip resistance). Clearly there is a need for a <u>non-contact</u> gage to monitor <u>on-line</u> the thickness of the applied coatings (post process gages do exist). The goal of this NIST Advanced Technology Program (ATP) project is to develop such a gage for use in the automotive manufacturing process. The gage will need to measure multiple spots on every vehicle at various location in the process and eventually will be integrated into a process control system. We refer to such a gage as the Wet Film Measurement (WFM) system. An example automotive painting process, showing where WFM gages would be useful, is illustrated below.
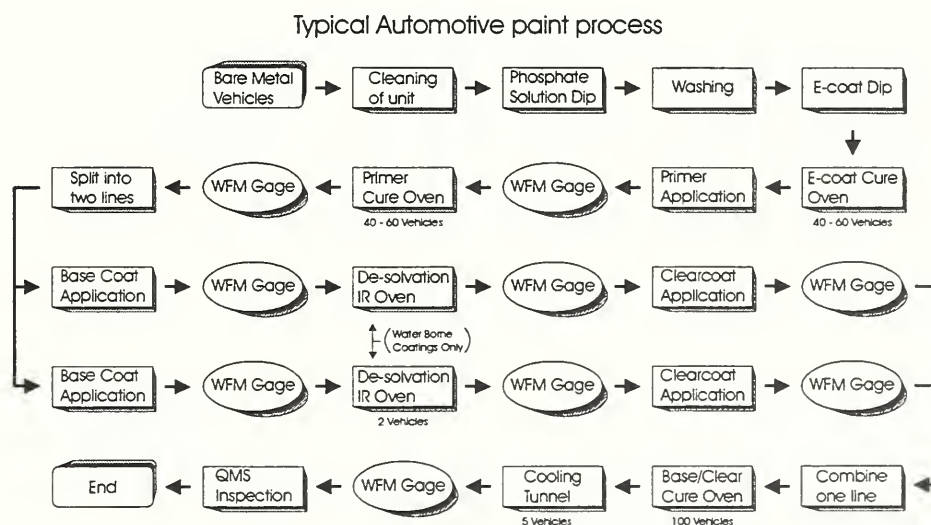
Figure 1.

Experimental

Various important WFM gage characteristics are listed below:

Required
- Non-contact
- Non-destructive
- Real-time
- Unaffected by factory environment
- Safe

Desired
- Insensitive to body position and tilt
- Works on all automotive substrates
- Reports all coatings rather than total thickness
- Little calibration required
- Low maintenance

We have investigated a number of techniques which exhibited promise in making non-contact film thickness measurements, keeping the above characteristics in mind. Of the techniques considered, we have chosen to pursue the development of Laser Ultrasonics (LU) for this application. The basic experimental set-up for Laser Ultrasonics consists of four major sections; two lasers, an interferometer, and optics to focus and collect the reflected light. The basic LU experimental setup is outlined below.

Basic Laser Ultrasonics WFM Setup

Generation Laser
Nd:YAG OPO, 1570 nm
10 ns pulse, <5 mJ/pulse

Fiber Optics

Interferometer
Confocal Fabry-Perot

Detection Laser
Doubled Nd:YAG, 532nm
>2 MHz linewidth, 200 mW CW

Focusing Head

Painted Sample Surface
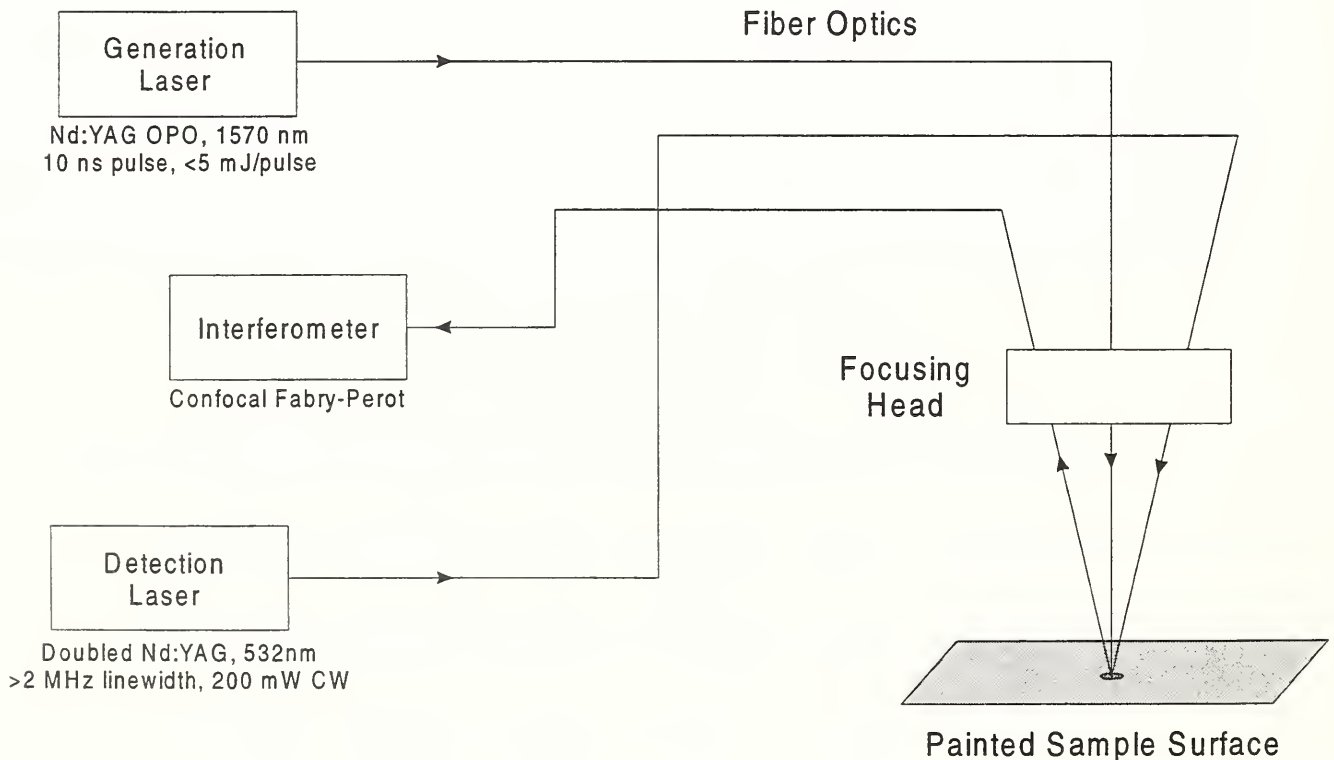
Figure 2.

The generation laser delivers a very short pulse (<10 nS) which results in extremely rapid localized heating of the sample from the absorption of the light energy. This heating produces a density gradient and then elastic waves (ultrasound) in a sample. For wet samples, some of the solvent may be ablated by this energy pulse (the energy densities used are far below that which can ablate or damage solids). The mass of the solvent ejected from the sample generates an opposing force into the sample, thus greatly increasing the ultrasound generation efficiency. The detection laser is used in concert with the interferometer to monitor the sample surface motions at ultrasonic frequencies by collecting the reflected detection laser light and sending this signal light to the interferometer. The ultrasonic motions of the surface induce a Doppler shift in the frequency of the detection laser, and this shift can be isolated and recorded by the interferometer. Example data from a wet paint film sample is shown in Figure 3.
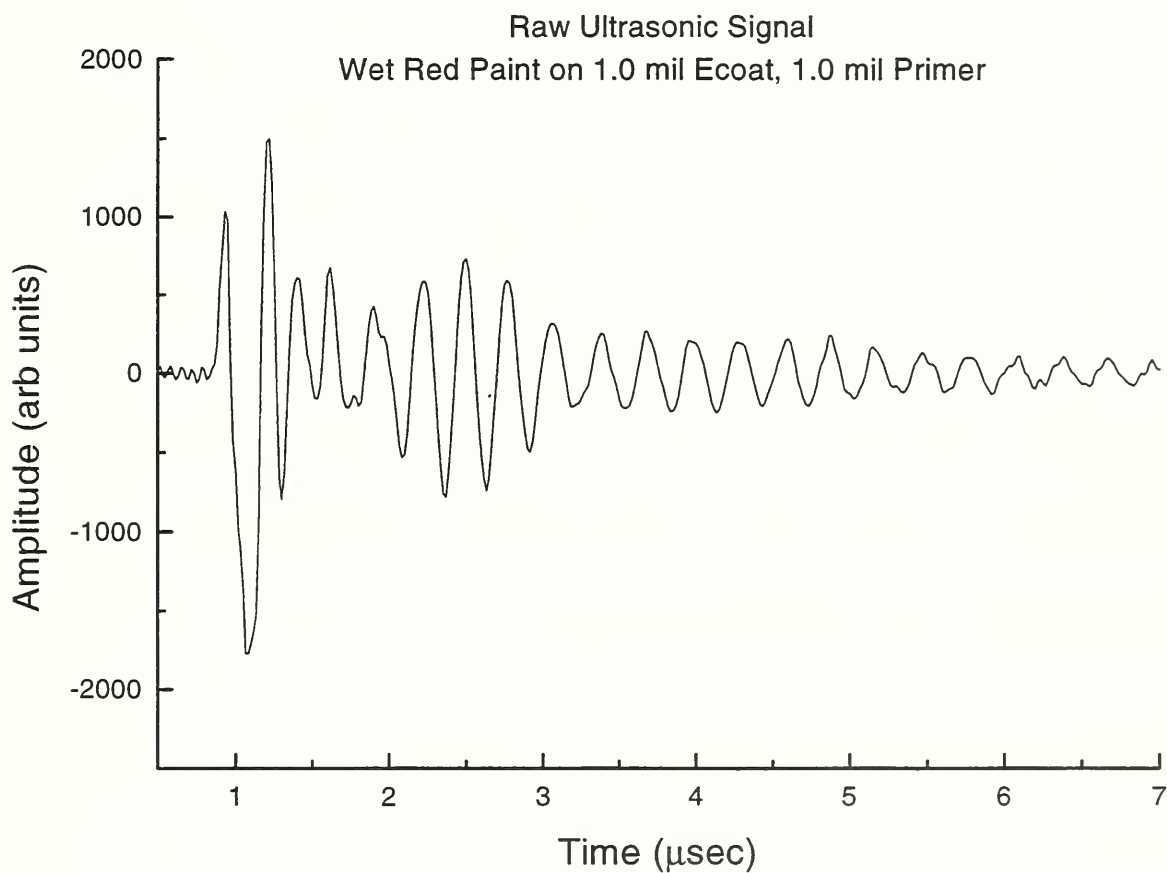
Figure 3.

The speed of sound of the material and thickness of the film will determine a characteristic resonance frequency for the film. Since these resonant frequencies can be easily extracted from the raw data, e.g., with a Fast Fourier Transform (FFT), the thickness can be calculated if the material's speed of sound is known. This is the method our gage uses: monitor-

ing the ultrasonic signal from a wet film, extracting the resonant frequencies and then calculating the film thickness from these frequencies. An example calibration curve for a white solvent borne base coat measurement is shown below. The data shown is for single shot measurements; no averaging is required.

Wet White Solvent Base Coat
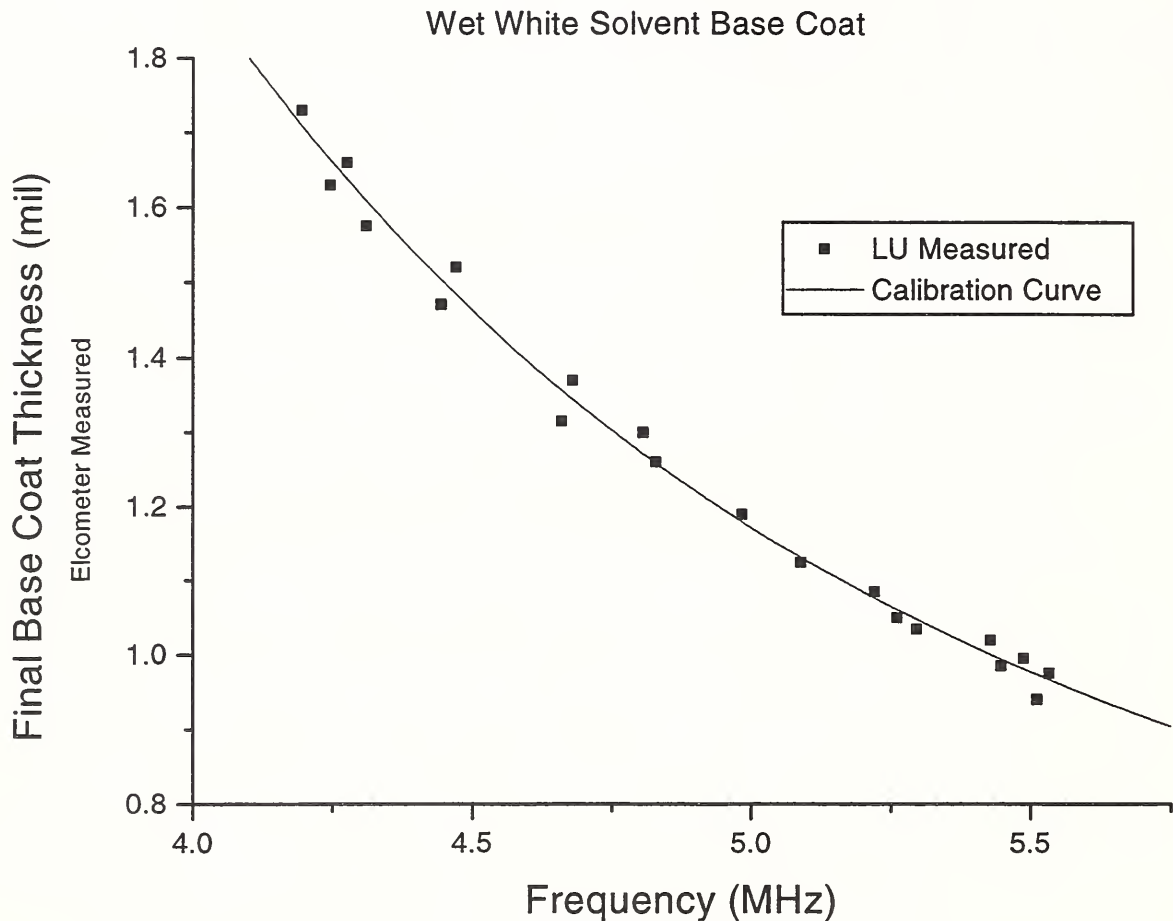
Figure 4.

The application of any gage for automotive film build measurements is especially difficult because a large number of spots on any individual body will need to be measured at various locations in the process (e.g., primer, base coat, clearcoat). Also, most assembly plants have two base/clearcoat paint lines which increases the number of applicable locations. This implies that a large number of sensors, or robots to position individual sensors, will be required. Laser Ultrasonics is well adapted to this requirement as it is possible to multiplex the equipment. Selecting specific fiber optics groups to carry; 1) the light energy to the correct sample position and 2) the reflected signal light from that position back to the interferometer allows one set of equipment to provide a large number of sensors (400) covering all potential locations within the entire assembly plant. No robot or other positioning equipment would be required. Figure 5 below illustrates this concept showing just a few sensors (9) per location.

On-Line Multiplexing of Laser Ultrasonics
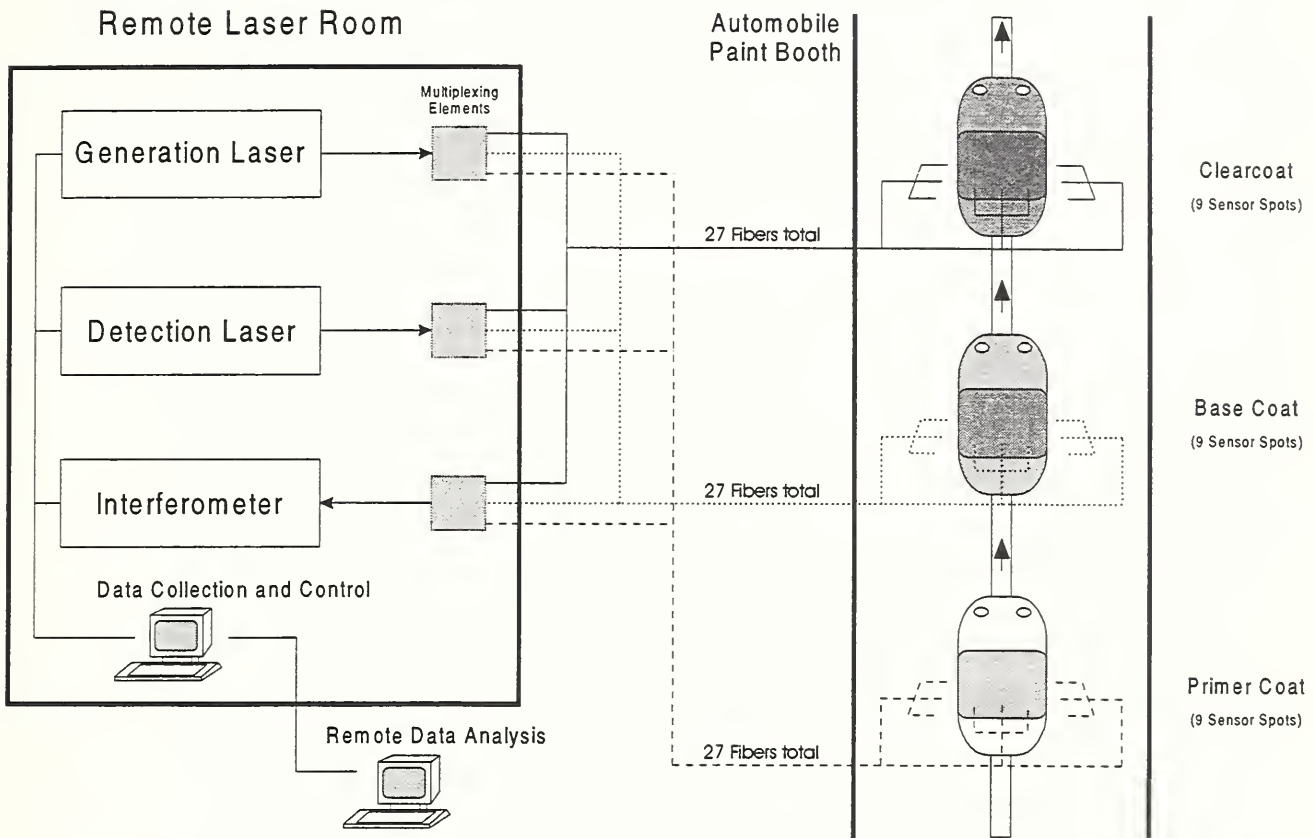
Figure 5.

Specific advantages of this technique, when applied to the on-line measurement of a wet automotive paint films, are highlighted below:

- **Extremely fast measurement.** The required data is collected in ~5 μs. This implies that we are not sensitive the motion or vibrations of the automobile body because the body is motionless on this time scale.

- **Large standoff sensor.** We are currently using a 6 inch standoff sensor. Larger standoffs (6-7 feet) have been demonstrated.

- **Large depth-of-field and insensitivity to sample angle and tilt.** Our current optical design has a depth-of-field of ±2 inches and tilt range of ±5°. Improvements of these values are not unlikely.

- **Extremely high precision.** Because the signal of interest is a frequency, the amplitude of the signal is not critical or even important as long as sufficient reflected light is collected. This point also helps explain the large depth-of-field and tilt insensitivity. Thus, we are not sensitive to the large variety of drifts and calibration problems that affect measurements that need to measure the amplitude of a signal.

- **Can measure all paint materials, both wet and dry.** Laser Ultrasonics is able to make measurements on all automotive paint materials (e.g., metal and/or mica bearing paints). We have tested a few dozen different types of automotive paint materials and have always been able to take measurements without problems.

- **Can make measurements on all metal substrates.** Any metal substrate (e.g., steel, aluminum) of varying thickness can be inspected. It may be possible to report the metal thickness, if desired.

- **Only minimal, extremely rugged components required in the spray booth.** Due to the large depth-of-field and insensitivity to sample tilts with this sensor type, dynamic positioning equipment (i.e., a robot) is not required. Also, since the laser light is carried on optical fibers, no electrical power is required in the spray booth. The only sensor materials in the spray booth will be glass lenses, glass and plastic optical fibers, and metal for the sensor housing and support arch for the sensors. However, if desired, the sensor can be mounted on the end of a robot arm.

- **Minimal line space required.** An additional advantage of not needing robots is that a measurement location should only require 3' - 4' of line space.

- **Extremely large number of sensors possible with multiplexing.** An additional subsystem will allow multiplexing of the signal lasers and reflected light to and from a large number of sensors. There is no theoretical limit to the number of sensors possible. Practical limits are set by the maximum repetition rate for the pulsed generation laser. 400 sensor positions are expected to be easily achieved, but again is not limited to this value.

- **Central secure location for all expensive and sensitive components.** Due to the use of optical fibers, it is not necessary to locate the main system components near the paint area. Any convenient central location between the paint lines will be adequate. The upper limit of fiber transmission has not yet been experimentally determined, but a distance of 150 M from the central location to the paint line sensor is currently achievable. We are developing methods to increase this distance quite appreciably.

- **Matches currently collected data.** The two most common current forms of compiling film build data are either Elcometer Electromagnetic Induction or the PELT gage. Both of these methods acquire numerous small spot samples of film build over the car body. Laser Ultrasonics would collect similar data (spot size ~3 mm diameter) and thus this new data should correlate well with previously collected data.

- **Operation in harsh environments.** Due to the minimal and rugged components used in the sensor head, Laser Ultrasonics should be able to collect data in virtually any process location (e.g., near or even inside a curing or dehydration oven, cooling tunnel, etc.).

- **Safety.** As a result of locating the lasers in a central location, the system should be quite safe. Access to areas where potentially dangerous laser light levels could exist can be easily and completely controlled. With this configuration, there is the added safety factor of no electronics in the spray booth and the avoidance of much of the necessary explosion proofing of the equipment.

The perceived limitations of the technique are:

- **Can currently only measure total film build.** As compared to the available PELT gage, the non-contact measurement capability of Laser Ultrasonics sacrifices sensitivity and thus does not currently allow the measurement of the individual film thickness layers. However, we do not believe that this inability will be a significant detriment for this application due to the following three advantages of the technique: 1) the ability to measure both wet and dry film builds, 2) multiplexing, allowing an extremely large number of sensors and 3) the extremely high precision of the measurement.

  With the large number of sensors, it is easy to measure the sample (i.e., auto body) before and immediately after the application of a new film. Thus, the thickness of that layer can be reported in real-time and the buildup of all layers can then be followed in the same manner. In many ways, we believe this may be a better measurement method <u>for process control</u> in that the determination of an individual film thickness does not require the entire film build process to be finished before the measurements can be made. Also, intermixing of layers is not a concern as we measure each layer's thickness individually.

  A common problem with this type of measurement procedure is that the error in the reporting of an individual thickness is increased because two separate readings are required. However, the extremely high precision of the Laser Ultrasonic measurements means that this additional error does not significantly contribute to the overall imprecision in the measurement. Care in indexing repeatable sampled positions will have to be ensured.

- **High initial cost.** The initial cost of this type of system will be relatively high. The technique was not rejected because the measurement application, within an automotive assembly plant paint booth, is a high cost procedure under any circumstance. However, the cost per sensor ratio for Laser Ultrasonics will easily be far less than the other techniques considered due to the large number of sensors available with multiplexing. That is, one Laser Ultrasonics system should be able to make measurements at numerous locations for both paint lines in a plant, and in fact should address any paint film thickness measurement need within the entire plant. Another point in favor of Laser Ultrasonics is, again, no positioning (i.e., robots) or other equipment are required. Robots are expensive and their real cost may not adequately factored into the total system cost (e.g., maintenance).

- **Laser eye safety concerns.** A recent change in the generation laser used has tremendously reduced the eye safety concerns. However, the system still uses lasers (Class IIIb) and thus precautions will be absolutely necessary. An advantage in this application is that the measurements will typically be taken in locations where human presence is not allowed.

Conclusion

The application of Laser Ultrasonics to the on-line non-contact film thickness measurement of wet automotive paints is proposed and the perceived advantages and disadvantages are discussed. The precise non-contact measurement of both wet and dry film thicknesses has been demonstrated. The installation of a beta measurement system in an automotive assembly plant with occur in the very near future. All of the advantages listed above would apply to the application of this technique to other non-contact film thickness measurements.

Advanced Technology Program No. 70 NANB5H1148

"Fast Volumetric X-ray Scanner for Three Dimensional Characterization of Critical Objects"

Scientific Measurement Systems, Inc.
General Electric Company
General Motors Corporation
EG&G Reticon

Substantial progress has been made toward the original objectives for this program toward the goal of implementing an advanced x-ray computed tomography (CT) system capable of rapid, three-dimensional (3D) imaging of automotive and other industrial components. Work on a number of critical advancements in hardware and software has yielded encouraging results which are leading to a specific system configurations for a volumetric CT (VCT) prototype system. The early empirical work has utilized a breadboard system at General Electric Aircraft Engines in Cincinnati, Ohio.

The three primary areas for hardware development have been the production of a fast, high quality area detector, a two dimensional collimator, and parallel processing which will support the computational throughput required to provide reconstruction of a complete three dimensional matrix of a component in a matter of minutes. The detector research has centered on evaluating through computer simulation and, subsequently, through empirical measurements, the performance of various scintillators in conjunction with new, proprietary photo-conversion and signal readout electronics. Based on favorable results, provided partly by a breadboard using of x-ray energies up to 450 keV, optimum components have been selected for incorporation in a prototype. Measurements of a number of factors, including detector efficiency, spatial resolution, and minimization of and correction for x-ray scatter contribution, have been made, in addition to the generation of preliminary VCT data sets. Figure 1 presents one such 3D data set for a jet engine turbine blade. Supplementary measurements are currently in process using the breadboard.

Benchmark evaluations of possible computing platforms have been made involving a number of Intel, Motorola, and DEC processors. An intermediate reconstruction platform will be used for the early work with prototype in mid 1998, with the intention of making a final assessment of the final hardware for the prototype in the latter part of 1998 to allow for imminent advancements in computing to be most fully utilized. The computing speeds encountered to date imply that the computing needs will be met by a reasonably priced parallel configuration of at least two of the processors tested.

The primary software development has been in the area of exact 3D reconstruction algorithms. Several different algorithmic approaches have been

formulated and tested for accuracy (particularly reduction of distortion of off mid-plane regions) and possible adaptation for parallel processing. One of the approaches has been chosen for incorporation in the prototype during 1998. Optimization of speed of reconstruction and data flow using the computer(s) of choice, the selection of which will occur in early 1998, will be a major effort during that period. Testing of the final prototype is scheduled for the latter part of 1998.

The use and formulation of 3D data of the type depicted in Figure 1 has been explored using several commercially available software analysis packages, as well as specialized software provided by team members. Encouraging results have been recorded, in terms of automation of three dimensional comparisons with CAD and engineering models and compatibility for input directly into a variety of such programs for surface and solid rendering of the geometries as measured with VCT.

Figure 1a. Volume rendering of the airfoil region of a hollow, jet engine turbine blade, derived from a VCT data set.
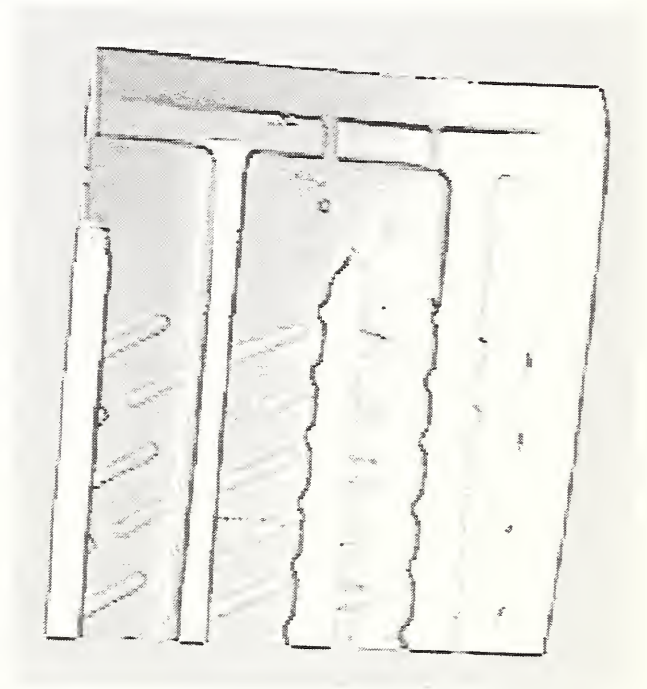
Figure 1b. A masked region of the data in Fig. 1a, showing the interior of the airfoil

Session IX. Information Technology

Managing Abnormal Situations in the Process Industries I:
Automation, People, Culture
E.L. Cochran, Honeywell Technology Center

Managing Abnormal Situations II: Collaborative Decision
Support for Operations Personnel
P.T. Bullemer, T. Cochran, S. Harp, and C. Miller,
Honeywell Technology Center

Future of Supervisory Systems in Process Industries: Lessons
for Discrete Manufacturing
I. Nimmo and E.L. Cochran, Honeywell

Integration Support for Large JV ATP Projects: Why? How? So
What?
V. Parunak, R. VanderBok, and M. Brown, Industrial
Technology Institute

Managing Abnormal Situations in the Process Industries I: Automation, People, Culture

Edward L. Cochran
Honeywell Technology Center

Consider the following incidents, all of them missed opportunities to demonstrate excellence in the management of complex systems:

- The left engine on a British Midlands 737 fails catastrophically. The flight crew mistakenly shuts down the *right* engine, and the aircraft ultimately crashes short of the runway while attempting an emergency landing.

- An air traffic controller clears a 737 to land on a runway at LAX, forgetting that she had cleared a commuter aircraft to hold for take off on the same runway. The 737 collides with the smaller aircraft with horrific results.

- The U.S.S. Vincennes, a guided missile cruiser engaged in combat operations, detects an aircraft leaving a commercial airport in an uninvolved country. The crew concludes the target is a threat and ultimately shoots down a commercial airliner, killing 290.

- The navigation of the lead ship in a squadron of destroyers on a night sortie breaks down, and the bridge crew, misperceiving lights on shore and believing they have passed a dangerous shoal, turn the ship directly into the hazard. It runs aground, as do most of the ships following behind.

- A process operator begins to start up equipment on the Piper Alpha production platform following a shutdown for maintenance. Unfortunately, the maintenance crew has not yet finished reconnecting the pipes, and escaping vapors explode, leveling the platform and killing 167 workers and rescue personnel.

- The operator of a chemical plant disbelieves a level reading that has led to the activation of safety interlocks and prevented the initiation of a batch process. He overrides the interlock and the reactor explodes.

- An operator sent to open a valve on one of four identical process units opens the valve on the wrong unit, leading to a catastrophic fire.

These incidents all have one thing in common: The people involved weren't really at fault, and the causes of the incidents should not be listed as human error. Instead, these incidents are just a few examples from a long and tragic list of cases in which the complexity of the systems involved combined with the communications requirements of the specific situations led to performance demands that individual humans are fundamentally not capable of sustaining.

In all of these cases the result was an accident, and therefore newsworthy. However, it is likely that, although not newsworthy, nonoptimal management of complex systems is ubiquitous, and has enormous financial impact in terms of waste and inefficiency. We are only beginning to understand the magnitude of the problem, and the true nature of its cause.

System complexity is a function of the number of components, the number of changes that can occur, the interactions among those changes, and the speed with which those interactions propagate. In practice, the number of components and relations is less important than the possibilities for change in the behavior of those components: People can often learn to deal with static complexity given sufficient training, but once the possibility of change is introduced, difficulty increases exponentially.

When systems become too complex for one person to manage, the management task is divided among members of a team of people. This works for a while, but the advantages of having more people to think about the problem is offset by the need for communication among those people, and by the introduction of new kinds of problems resulting from various kinds of communication lapses. People not only must be capable of understanding the situation and developing a successful response, but they must rely on each other in developing their understanding and determining the best course of action. There is a delicate balance to strike here, lest the communication demands consume all of the attention of the people who were added to the system to help manage it.

As systems become more complex, their human supervisors need to be more capable.

As the numbers of people involved increases, the amount of communication required increases.

Eventually, too much is required of the individuals charged with supervising the system, and breakdowns occur. To blame these breakdowns on human error is akin to criticizing a novice juggler for her inability to juggle 15 balls at once. While standing on someone's shoulders. On a unicycle.

Since the capabilities of individual humans are limited and change too slowly for us to rely on evolution as a solution, there are only two options:

- Reduce the apparent complexity of the system, or

- Reduce the need for, or increase the efficiency of, communication.

The remainder of this paper describes a number of ways of achieving these goals. I begin with an overview of some of the problems experienced by individuals and teams in dealing with complex systems, which will inform our concluding discussion on promising solution approaches.

The Role of Individual Humans in Complex Systems

Many current views of human action in complex systems use a model similar to that in Figure 1. The model in Figure 1 is a variation of a decision-making model adopted by the Chemical Manufacturers Association.
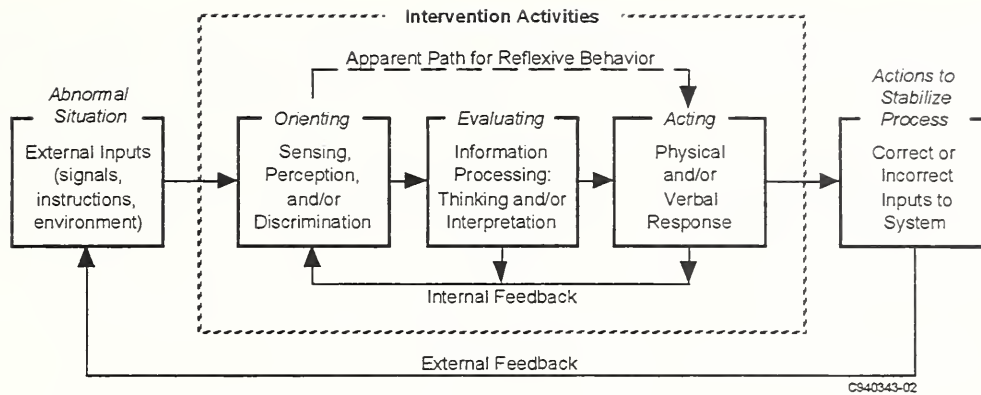
Figure 1: **Standard model of human interaction with complex systems.**

In this model, there is an outside world which generates events, some of which are perceived by humans, processed internally, and ultimately generative of responses—actions which initiate new cycles of activity.

This model is often used to explain and categorize the kinds of errors that can arise during different stages of the Orient-Evaluate-Act cycle. For example, operators may fail to notice key information that is present on their displays, leading to inadequate orientation and inaccurate evaluation. Such errors can be minimized by increasing the conspicuity of the information or reducing the workload of the operator. There has been increasing emphasis on the rigorous application of human factors principles to the user interfaces to complex systems to identify and minimize such problems.

However, those who apply this model in that traditional way often make two assumptions which are only approximately true. Worse, these assumptions break down precisely in the same way and for the same reasons that management of complex systems often breaks down. The result is that our traditional approaches to developing better ways to manage complex systems have a large blind spot.

The first assumption is that, in describing the activities of the operator, the outside world is the standard reference. Thus, we tend to think that if turning on a pump lowers the level in a tank, and the operator needs to lower the level in a tank, the operator should turn on the pump. If that act is not carried out, a mistake is declared. Most of the time this way of thinking is perfectly adequate.

However, the world that *really* matters is the one in the head of the observer. If the operator does not know that turning on a pump lowers the level of a tank, or does not know that the particular tank with the problem has a pump that can be turned on, then it is pointless to expect the operator to turn on the pump. The "mistake" is not the operator's—the operator is after all acting perfectly consistently with respect to the world as they know it. Instead, the problem is with the way the operator learned about the system. That process is not represented in the model at all.

Abnormal situations frequently arise because the physical world is not congruent with the world in the head of the operator [Indeed, all of the examples at the beginning of this paper such examples].

The second assumption that users of the Orient-Evaluate-Act model often make is related to the first. We assume that there is a single set of "facts" to be had, which should, ideally, serve as a reference for evaluating the performance of operators and the system itself. This assumption serves us well most of the time, especially when the facts in question concern physical things that can be observed and discussed and agreed upon. The assumption that facts are facts is clearly not true in the realm of political opinion, religion, and other cultural matters, and in fact there is a long tradition in Western philosophy to distinguish between "facts" and "beliefs" and to develop a method (the "scientific method") of extending the realm of fact.

The problem is that individuals rarely distinguish between the two, and therefore it is wrong for us to assume that individuals will be orienting, evaluating, and acting solely with respect to "the facts." Consider the operator of a process unit faced with an unexpected and severe upset. The plant manager has stated in writing that the policy in such instances is to shut the unit down to prevent damage and to permit an orderly recovery. All operators have read and signed this policy. The operator in this instance attempts to ride the disturbance out, and, after some drastic control moves, ultimately succeeds. The "facts" would indicate that the operator in this case made a mistake by persisting, in direct opposition to standing policy, with the attempt to recover from the upset without shutting down. However, what if the last four times this event has occurred, operators who have kept the plant running have been rewarded, and those that have followed established policy have been ostracized?

When conflicts such as this exist, people do the best they can, forming beliefs about what the facts are based on their own observations, knowledge, and experience[1]. When the need to act in accordance with the "facts" arises, the results can be unpredictable.

A different example of a problem with the use of "facts" arises in complex systems due to the relative rarity of the kinds of problems that people have been tasked to prevent. In essence, the issue is that humans are not very good at understanding and explaining events with which they have little experience. College-educated adults may believe that objects that are dropped from a moving car fall straight down, or that objects swung around on a string and released will spiral away. Humans are also fundamentally flawed statisticians, and fall prey to all manner of inaccurate assessments of low probability events, which causes problems when those assessments are acted upon as "fact."

The problem for us as designers of complex systems is that we tend to design them as if the users will understand the facts and determine the true state of the world and act accordingly. As we have seen, the facts are quite malleable and the perceived state of the world may be inaccurate. Abnormal situations frequently arise for this reason. [Indeed, all of the examples at the beginning of this paper such examples].

Richard Feynman makes a similar point in a discussion of the Challenger accident:

> "Let us make recommendations to ensure that NASA officials deal in a world of reality in understanding technological weaknesses and imperfections well enough to be actively trying to eliminate them.... And they must be realistic in making contracts, in estimating costs, and the difficulty of the projects. Only realistic flight schedules should be proposed, schedules that have a reasonable chance of being met. If in this way the government would not support them, then so be it. NASA owes it to the citizens from whom it asks support to be frank, honest, and informative, so that these citizens can make the wisest decisions for the use of their limited resources.
>
> "For a successful technology, reality must take precedence over public relations, for nature cannot be fooled."

---

[1] The contributions of culture to experience can significantly affect the results, as we will see later in this paper.

The simple orient-evaluate-act model needs to be viewed as a more complex system of interactions between perceivable events and the knowledge and beliefs of an individual, leading to the construction of an understanding of the situation upon which responses are based. This expanded model is depicted in Figure 2.
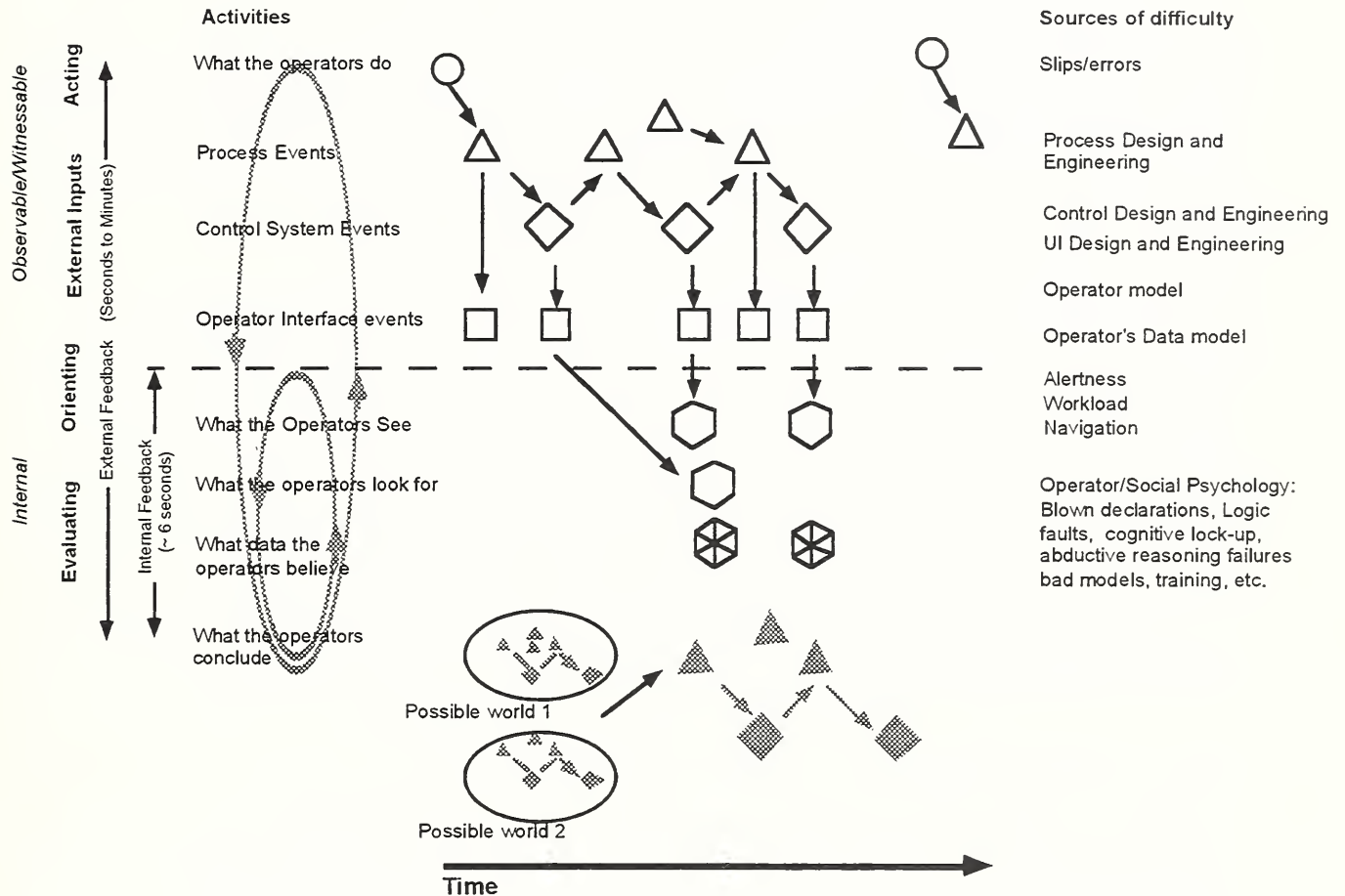
Figure 2: Elaborated model of human interaction with complex systems.

Operations Teams and Complex Systems

Complex systems are usually managed by *teams* of individuals. As a result, all of the issues discussed above are present for each individual member of the operations team. In addition, a number of new issues emerge as a consequence of the need to solve problems collaboratively.

Shared understanding of how the world works.

As important as it is for the lone system operator to have thorough and accurate knowledge of their system, it is even more so for members of operations teams, for two reasons. First, teams are responsible for larger, more complex systems, and no single person may be capable of understanding the whole system. Instead, all of the members of the operations team depend on each other's understanding of parts of the system, and gaps in that understanding are both more likely, and more likely to have greater consequences.

The second reason that thorough and accurate knowledge of the system is critical for members of operations teams is related to the first: The team member with inaccurate knowledge is likely to lead the rest of the team down the wrong path *even when they presumably know better.*

There are many reasons for this. The more knowledgeable team member(s) may incorrectly yield to the authority, experience, or position of the other, or be overly influenced by irrelevant social factors such as friendship (or lack thereof) or be shaken by stress, excitement, fear, or other emotional reaction to the situation.

In the British Midlands incident described at the beginning of this paper, at least one of the pilots may have acted on the "knowledge" that fresh air to the cockpit is provided by the right engine. The smell of smoke may have led to the assumption that the right engine was the one that had failed—an assumption that was not challenged by the other pilot nor shaken by the evidence of the instruments.

Shared understanding of the desired state of the world

It is critical that the members of the operations team develop a consistent understanding of their short- and long-term goals—the formal and informal rules and guidelines for the operation of the system they are responsible for. We have already seen that the "facts" about the desired response to an incident may be not be the same for operators and plant managers, and indeed this is an example in which the lack of shared vision within an operations team can lead to nonoptimal results.

The same sort of problem (albeit on a much shorter time scale) happens hundreds of times a day in the U.S. when team members miscommunicate—when one or more members of the team fail, for whatever reason, to receive the message that another member intends to send. From that point forward, the team members are operating with respect to different views of the world. As a result, the wrong valves get opened, airplanes land on the wrong runway, ships run aground, and plants run short of raw materials.

Shared understanding of the current state of the world

It is critical that the members of the operations team not only develop an accurate model of the state of the system they are responsible for, and its desired state, but the state that the system is actually in. When a members of an operations team believe different things about the state of a plant, trouble often results. [Note that post incident review teams typically label the beliefs of the team member(s) who is determined to have been incorrect as "beliefs" or "assumptions" (often described as "unfounded"). In fact, the team member(s) who are determined to have been *correct* often were working with beliefs or assumptions as well—they just happened to be accurate. All of the team members—correct as well as incorrect—act consistently with the world inside their own heads.]

When an operator believes that maintenance is complete on equipment at a plant at the same time that a maintenance worker believes that the equipment is still out of service for maintenance, an incident is likely. Whether the operator or the maintenance worker is "correct"—that is, has a set of assumptions more consistent with the "true" state of the plant—is irrelevant, as this changes only the nature of the incident, not the likelihood that an incident will occur. Some of the most horrific incidents in the history of technology (including the loss of Piper Alpha oil rig with most of those on board and the $1.6 billion fatal explosion and fire at the Phillips plant in Texas City in 1989) can be traced to this precise problem.

The problem worsens when the state of the plant is unfamiliar (as it is during upsets), when the understanding has to be developed rapidly (as it is during upsets), when the team is inexperienced in working together on the particular problem involved problem (as it is during upsets), and when accurate understanding is important to the selection of an appropriate response and ultimately to the outcome of the event (also as it is during upsets). Reviewers of the Vincennes incident describing

the chaotic environment on the bridge and in the ships combat information center —the ship had been maneuvering violently, and firing weapons at surface targets—concluded that it played a significant role in the incident.

Shared understanding of the plan to reduce the discrepancy between actual and desired states of the world.

Even when the operations team has a shared vision, an understanding of the how the process works, and an accurate diagnosis, they may fail to develop a shared understanding of the recovery plan. This is relatively rare, and usually happens in combination with one of the previous problems described above. Nevertheless, since most processes are complex enough that several options are available to recover from any undesirable state, it is important that the operations team agree on and operate consistently with respect to a single recovery plan.

Opportunities for error arising from the need for communication

Given the need for every person on the operations team to act consistently and accurately with respect to the state of their process, it is not surprising that errors occur. In fact, the relatively low rate of incidents stands as testament to the effort that has been devoted to creating just this kind of shared vision across process teams.

Many good analyses of the causes of individual human errors exist, and need not be summarized here. In an ideal culture of team operations, most errors of this type would be quickly detected and corrected, given the interlocking responsibilities and oversights of the members of the operations team.[2]

Since most operations teams do not yet operate with respect to completely consistent perspectives, they are subject to several kinds of errors which are unique to team settings and which are extremely difficult to overcome. Awareness of these team-induced errors has been growing, and attempts are being made to address the problem (e.g., Cockpit Resource Management training initiatives in aviation), but the study of this area is still in its infancy.

Several examples of these errors are described below.

Inadequately qualified assertions

Assertions are statements about our observations. While we may believe these statements are descriptions of the way things "really" are, in fact all we know is how we ourselves observe them. An assertion is "true" if we could provide a witness who would concur with our observation, and it is "false" if a witness would dispute it. The act of making an assertion commits us to truthfulness, but assertions may be qualified and thus relieve us of the need for absolute precision (this paper is a fine example).

However, accurate communication requires that we be able to assess the need for such qualifications in our sharing of observations, facts, urgency, and so on, with other members of our team. This requires either significant knowledge of the other people on the team (as is the case in close-knit teams of long standing), or absolute adherence to procedure (is as the case in naval carrier operations or air traffic control.) Most process plants—indeed most civilian operations of all kinds) operate as if the team has the requisite self-knowledge to accurately communicate in the

---

[2] An exception occurs when the team structure is designed to depend on the complete trust of members of the team in an experienced team leader, who must make accurate decisions in isolation. The need for this team structure is rare—formation acrobatics is one example—and when the leader of the team does make a mistake, the cost can be high: There have been instances in which entire formations of planes followed the leader into the ground.

absence of rigorous procedures. This is risky, as demonstrated by the large number of audio recordings of incident communications in which amplifications and clarifications of observations are required in order to get an appropriate response.

Inaccurate declaration of knowledge

Declarations are decrees—impositions of aspects of our own world, such as our beliefs, on others. They can be loosely described as determining a view of the world that those subject to the decree will abide by, "accurate" or not. Because declarations represent the views of at least one person, but not others (else imposition would not be necessary), they are usefully described as valid or invalid, according to the power of the person making the declaration, as well as accurate or inaccurate.

The power we grant to some people to make valid declarations is called authority. Thus valid declarations (whether accurate or not) are made by those with authority, and invalid declarations (whether accurate or not) are made by those without authority. It is easy to understand how valid but inaccurate declarations can be a source of problems, as when a flotilla of destroyers follows the admiral's ship into the shoals.

Inaccurate declarations are in fact a major cause of the progression of abnormal situations to accidents, because they are relatively frequent, and because they create a world—a set of shared beliefs—in the operations team, whether it is accurate or not. For example, in the presence of conflicting temperature readings and no other witness, a declaration is required to determine which reading is "correct"—and that declaration will be used to coordinate subsequent action. The declaration may be valid—the shift supervisor may have the authority to make it—but it may not be accurate. This is a critical point in any situation, because from then on the operations personnel will act in accordance with the world they have created, until it becomes so obviously inconsistent with further events that the declarations are retracted. Unfortunately, humans are very good at weaving consistency out of noisy or incomplete observations— "obvious inconsistency" must be very obvious indeed.

Ineffective culture of authority

A more insidious kind of problem arises when the exercise of authority is ill-defined, as when incident command functions are not specified, or those with authority do not exercise it appropriately. Authority is important, because it enables declarations to be valid. Declarations are one of only two ways of defining shared knowledge, and by far the most efficient in terms of time.[3]

Hierarchies often exist primarily to permit declarations to be made in the presence of unresolved questions. In the absence of authority, decision making can become a social process, and anyone on the team becomes capable of making a de facto declaration on the basis of field knowledge, social influence, bravado, self-certainty, or other factors, whether appropriate or not. There is no time for an effective inquiry process, and a consensus "state of the world" is thus reached in an ad hoc way.

Many incidents attributed to so-called "garden path thinking" or "cognitive lock-up" can be viewed in these terms. A human declares that a radar blip represents an attacking fighter, or that safety interlocks have inappropriately engaged. The declaration is not questioned, even though authority structures are supposedly in place to ensure that it is. Subsequent actions, consistent with the inaccurate declaration, lead to incidents.

---

[3] The other way to define shared knowledge is through a process of inquiry—questioning the sources of differences in assertions until they are explained through more comprehensive sharing of context.

Throughout the Piper Alpha oil platform fire, a neighboring production platform pumped product to Piper Alpha (literally adding fuel to the fire) because the supporting platform manager assumed that he had no authority to do otherwise in the absence of an order from Piper Alpha to stop production. [Piper Alpha's control room and communication capability was destroyed in the initial explosion].

Confusion between requests, promises, offers, and assertions

We should not expect operations teams to be students of discourse and rhetoric, but in fact in the absence of clear procedures or team with enormous experience working together, the opportunity for confusion is rampant. Consider the opportunities for problems inherent in the confusion between requests, promises, offers, and assertions:

- *Promises* are the linguistic acts that allow us to coordinate action with others (both implicitly and explicitly). Four fundamental elements are involved in a promise: A speaker, a listener, a condition of satisfaction, and a completion time. The act of making a promise commits the promiser in three domains: sincerity, competence, and reliability.

- *Requests* are solicitations of promises from someone else. They can be thought of as a promise of the form "I [perhaps conditionally] promise X if you promise Y." X can stand for tangible and intangible conditions of satisfaction, e.g., "I will be pleased," "I will be better able to make a diagnosis," "I will get you safely out of this." Y carries the form of the promise in return, e.g., "[Listener] will fetch a reading from the gauge;" "[Listener] will close valve 17 and open valve 18." A request, when answered with a declaration of acceptance, creates a promise. If a request is declined no promise has been made. [A request which cannot be declined is not a request, it is a declaration.] The act of making a request implies that we can competently assess our knowledge of the system, and it commits us to acting consistently with an implicit assertion we thereby make, that the *other* person is sincere, competent, and reliable.

- *Offers* are a request for a promise of the form "If I promise Y, will you promise X?" where X and Y are defined as above. Thus a simple "May I help you?" if it is sincere, carries a social meaning such as "I will help you [the promise] if you would be pleased for me to do so [promise as condition of satisfaction]." The commitments, structure, and implications of making offers are very similar to those involved with making requests; only the agent of conditionality is reversed.

This is obviously a brief and highly simplified treatment of an extremely rigorous and extensive domain of study. Successfully transferring information from one person to another requires the successful application of myriad knowledge about ourselves and our colleagues, otherwise confusion results: We make requests of people who are not able to carry them out, raising doubts about our own competence and the validity of further requests. We unknowingly accept a request that we are unable to fulfill, calling into question our competence and/or sincerity. Ultimately, we may not know what is being asked of us, or why, or what the consequences will be for failing to meet the request.

If that were not challenging enough, the linguistic domain of *pragmatics* enables every one of the above distinctions to be rendered false; for example, an assertion ("It is cold in here!") can be a request ("Please close the window!"). An offer ("May I review your monthly report") can be a declaration ("Give me your monthly report")—especially if the speaker has authority over the listener. And so on. Pragmatics requires a further layer of social knowledge to be negotiated, one that focuses on the implicit or implied motivations, values, and characteristics of the people with whom we interact. Pragmaticians study linguistic acts from both sides—the intent of the speaker and the interpretation of the listener—and the social mediations can be complex indeed. The fact

that humans communicate so well is testimony to how finely-tuned and sophisticated are our abilities to act within a social environment.

Summary: Teams and culture in complex systems

We have argued that all of these linguistic acts imply social commitment; they can only function from the background of shared social practices and a shared ability to recognize and act in accordance with those practices. An operations team in a complex process, then, must understand how the process works, and their goals for the process, and the state of the process, and the actions to required to close any gap between the state of the process and their desired state. But they need to know more.

They need to understand themselves, their team members, and their culture.

As a rule, those charged with designing, implementing, operating, and maintaining complex processed pay little if any attention to these issues, if they are aware of them at all. The result is that we handle the requirements imposed by the need to collaborate with others in process operations in the same informal way that we handle the need to collaborate with others in planning a PTA fund-raiser, remodeling a kitchen, raising a family, and every other endeavor typical of life.

As a rule, people are intuitively very good at this, as the continued existence of PTAs and remodeling contractors, to say nothing of civilization itself, amply demonstrates. Nevertheless, mistakes are made; misunderstandings occur; backtracking is necessary; failure is a possibility.

There are two primary reasons why this situation is not tolerable in complex systems:

- The complexity of the systems and the speed with which events propagate make our intuitive methods of gaining alignment very inefficient and error-prone.

- The consequences of error are too high.

Lessons for managing automation in complex processes

Given the forgoing discussion, we are left with two primary issues that need to be addressed if we are to progress further to successfully manage systems of higher complexity, with larger operations teams.

First, we need to find better ways to manage the apparent complexity—to assure that team members have consistent views of all the things that we have argued they must.

Second, we must find better ways to support real time collaboration and the exchange of information as accurately, quickly, and richly as the operations teams of complex systems will require.

Consistency

There are a number of ways to ensure consistency in the understanding of systems.

Training

Rigorous, high quality training is mandatory, especially including training for the rare events that test the operations team the hardest. As technology improves, the extension of simulation-based training from domains such as aviation is becoming more economical. But good training need not be expensive: "What if?" training, role playing, and scenario development are all very effective in building shared views and expectations.

Reporting

Incident tracking and reporting is critical to the deliberate and disciplined sharing of insights and experience gained from unusual events with everyone who might benefit from them.

Communications

A formal communications policy must be in place that ensures that consistent messages are sent from the top to the bottom and from one end of the plant to the other.

Authority

An incident command structure must describe the authority of those involved—and those involved must live up to that authority.

Procedures

Most important of all, consistency requires that the operations culture make a disciplined practice out of creating, evolving, and following *procedures*. Procedures incorporate knowledge compiled when there is time to think so as to benefit those in situations in which there isn't. Procedures clarify ahead of time the goals, expectations, and information requirements of unusual situations. Communications procedures ensure that the appropriate information is available to the appropriate people at the appropriate time. [Good radio procedures alone can reduce uncertainty in message content by a factor of two, and perhaps even more in an industry that often refers to equipment entirely in easily confused alphanumerics, such as J-159A].

The U.S. Navy, over the past thirty years, reduced the rate of mishaps by a factor of six, from 19.3/100,000 hours of flying time in the early 1960s to under 3.0 in the early 1990s through a comprehensive effort to develop best practices and instill absolute procedural discipline in using those practices.

Collaboration Support

Once we have achieved some measure of consistency within the operations team—and only then— we can begin to develop ways for the members of the team to collaborate quickly and accurately. [The introduction of collaboration support technology before consistency is achieved is not helpful and may in fact be harmful in that it might permit the more rapid exchange of inaccurate information.]

The goal will be to ensure that each member of the operations team has access to the right information, at the right time, for the right reasons. The next paper describes an ambitious approach in this area.

Conclusion

We have seen that the use of operations teams to the management of complex systems introduces a number of extremely challenging issues that must be solved if progress is to be made. Most of these issues result from the fact that our normal approaches to human interaction, and our skills in communicating with others, are simply not adequate to keep up with the rapid changes that occur in processes. Even if we could keep up with these systems as well as we do in our other daily activities, the consequences of the occasional mistakes are too great for us to tolerate.

To make further progress, effort needs to be devoted to the development of deliberate, principled approaches to the establishment of rigorous and consistent operations cultures. Only then will further advances in automation technology be likely to succeed.

Managing Abnormal Situations II: Collaborative Decision Support for Operations Personnel

Peter T. Bullemer, Ted Cochran, Steve Harp and Chris Miller
Systems and Software Technologies Department
Honeywell Technology Center

Introduction

The largest economic disaster in U.S. History (not due to natural causes) was a $1.6 billion fire and explosion at a petrochemical plant in 1989. This accident represents an extreme case in a gamut of minor to major process disruptions, collectively referred to as abnormal situations in this paper. Most abnormal situations do not result in explosions and fires but are costly nevertheless, resulting in poor product quality, schedule delays, equipment damage, and other significant costs. The inability of the automated control system and plant operations personnel to control abnormal situations has an economic impact of at least $20B annually in the petrochemical industry alone (Nimmo, 1995).

The Abnormal Situation Management (ASM) Joint Research and Development Consortium, led by Honeywell, was formed in 1992 to develop the technologies needed to allow plant operations personnel to control and prevent abnormal situations. In addition to Honeywell, the ASM Consortium includes the U.S. operations of the seven largest petrochemical companies (Amoco, BP, Chevron, Exxon, Mobil, Texaco and Shell), two innovative real-time industrial process controls applications software vendors (Applied Training Resources and Gensym), and a specialty chemical company (Nova Chemicals). In addition, the ASM Consortium has three university affiliates in Dr. Jim Davis of Ohio State University Department of Chemical Engineering, Dr. Venkat Venkatasubramanian of Purdue University School of Chemical Engineering and Dr. Kim Vicente of University of Toronto Department of Industrial Engineering.

In 1994, the U.S. National Institute of Standards and Technology (NIST) Advanced Technology Program (ATP) awarded a 3.5-year, $16.6 million cooperative agreement to the ASM Consortium. The goal of the NIST-funded Collaborative Decision Support for Industrial program is to demonstrate the technical feasibility of collaborative decision support technologies for improving the performance of operations personnel.

This paper presents the system-level solution, AEGIS (Abnormal Event Guidance and Information System), currently under development in this NIST ATP program. We begin the discussion with a prototypical characterization of the nature of collaborative decision-making in the process operations environment of the petrochemical industry. Next, we present the collaborative decision support vision as characterized by six primary functional roles of the AEGIS prototype. In addition, we discuss several technological innovations addressing key challenges in collaborative human-machine interaction, system architecture, and system customization tools. Finally, we conclude with summary of program progress in implementation of the technical innovations and future research plans.

Collaborative Decision-making in Process Operations Environment

Most oil refineries and petrochemical plants use distributed control systems (DCS) to simultaneously control thousands of process variables such as temperature and pressure. The major human role in this control is to supervise these highly automated systems. This supervisory activity requires: monitoring plant status; adjusting control parameters; executing pre-planned operations activities; and detecting, diagnosing, compensating and correcting for abnormal situations. Increased demands for higher efficiency and productivity in these industries are resulting in tremendous increases in the sophistication of process control systems through the development of advanced sensor and control technologies. However, these sensor and control technologies have not eliminated abnormal situations and will not in the future. Consequently, operations personnel continue to intervene to correct deviant process conditions.

The persistent paradox in the domain of supervisory control is that as automation technology increases in complexity and sophistication, operations professionals are faced with increasingly complex decisions in managing abnormal situations. A contributing factor to this phenomenon is that the sophistication of the user support technologies has not kept pace with the task demands imposed by abnormal situations. Thus, the focus of this program is to develop collaborative decision support technologies that significantly improve abnormal situation management practices.

Decision making model for Operations Intervention Activities

To provide a decision making model with which to structure solutions to existing ASM problems, the ASM Consortium adopted a modified version of a model proposed by the Chemical Manufacturers Association show in Figure 1. This model provides a description of how a plant operations team responds to intervene in an abnormal situation. The left side of the figure represents the occurrence of an external abnormal situation. The operations team, represented by the dashed box, then will intervene to stabilize the process.
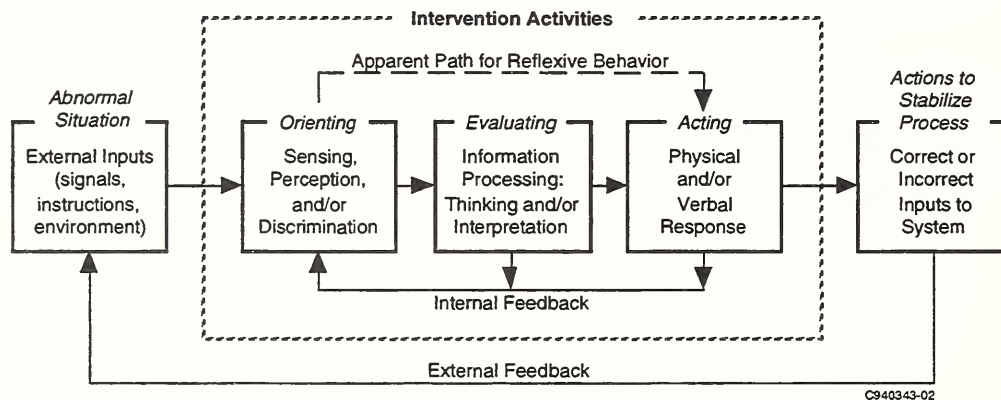
Figure 1. Decision-Making Model for Operations Intervention Activities

- **Orienting**—In the first stage of intervention, a disturbance in the process is detected. In the Evaluating stage, the operations team develops hypotheses regarding the cause of the anomalous operating conditions.

- **Evaluating**—As denoted by the dashed line, which goes directly from the first to the third stages of intervention, responses to certain disturbances may be so well-rehearsed that it may appear that the Evaluating stage is skipped.

- **Acting**—Finally, the operations team must make compensatory and/or corrective action, including actions to determine if the hypotheses are correct.

All preventable abnormal situations can be traced to a breakdown in the intervention activities (For detailed discussion, see Cochran & Nimmo, 1997). For example, external inputs pertaining to process upsets can be missing, erroneous, or masked—which leads to incorrect orienting and subsequent behaviors. The fundamental goal in developing AEGIS is to improve ASM by enhancing the accuracy, completeness, and speed of the human activities in Orienting, Evaluating, and Acting.

Plant Operations Roles

There are various operational and nonoperational personnel that impact the operation of a process unit in a plant complex.

The plant personnel fulfilling the operations shift team roles are expected to be the primary users of the AEGIS functionality. A prototypical operations shift team consists of a shift leader, a console operator, and 2-5 field operators. In general, shift teams work together, unless someone is out sick or on vacation, in which case another shift team member comes in to cover for him. Shift teams rotate from working nights to days. Most shifts are 12 hour periods. During the weekdays, many maintenance projects are going on, and the engineers, craftsmen, management personnel are all available to interact with the shift team. The night, or "graveyard" shift is a completely different story. Very few people are around and the goal is to "survive the shift", i.e., to keep everything going until morning. If an upset happens, there are no backup people available, so the shift leader has to call in help if there is time.

- **Shift Leader**—The shift leader role (also known as the head operator or stillman) is responsible for overseeing the field and console operators in the detailed monitoring of the process and ensuring the execution of the relevant preventative maintenance (e.g., daily routine duties) and abnormal situation management responses. The shift leader is a senior operations staff member, who may or may not be qualified as a console operator (although they typically are qualified). The shift leader is in charge out in the field. He typically makes "rounds" in the field as a follow-up to the field operator's rounds, observing the process equipment and making minor adjustments, noting potential equipment problems, and verifying sensor readings. He is also responsible for filling out the "Shift Log Book", which is a written log located in the shed, describing any significant activities that have taken place during his shift. During an upset, the shift leader becomes another hand in the field, generally playing a secondary role to the console operator who is in control. If there is sufficient time, the shift leader will get on the phone to call for backups.

- **Console Operator**—The console operator is responsible for controlling the process via the distributed control system. He/she has the task of monitoring and making "moves" that maintain and change the state of the process. In addition, the console operator has the responsibility of coordinating the actions of field operators and keeping abreast of the maintenance activities in the field. The console operator is the focal point of communication between the various distributed operations personnel throughout the complex because he has the central view and control of the process via the DCS. A console operator must first be certified as an field operator in all field areas.

- **Field Operator**—Depending on the size of a given unit within a plant, there will be one or more field operators. Each operator is responsible for their own area but many are qualified on more than one area (they may rotate from one area to the next, and they monitor each other's work, by keeping an eye on the other areas as they conduct their rounds, and sometimes

helping another operator when performing maintenance and/or troubleshooting tasks). Some of the field operators are also console operator qualified, and may rotate from working in the field to working in the control room every other day. The field operator has the equipment view of the process and serves as the "human sensor" that checks the status and validates the correctness of the sensor and instrumentation readings in the field to ensure that the view of the process is accurate. In addition, field operators use their senses (e.g., eyes, ears, and nose) to monitor the process, constantly identifying potential problems (a funny smell or sound) with the process equipment and initiating preventative maintenance, adjustment, and repair activities. Field operators take periodic readings of indicators not available to the console operator, take periodic product samples for laboratory analysis by the process engineers, and perform manual operations such as operating soot blowers or operating manual valves. Field operators prepare and warm up equipment such as steam pumps and place them on line in a controlled manner to avoid sudden surges, cavitation, or loss of suction. They are responsible for directing maintenance personnel to the appropriate worksite and for receiving deliveries of products (such as catalyst) to the unit. In an upset situation, they are the first on the scene and provide a critical diagnostic and mitigating response role in abnormal situation management by assessing the situation (e.g., confirming/refuting DCS data) from a process equipment perspective and by taking actions which are outside of the scope of DCS control either by design (e.g., fire fighting) or by circumstance (e.g., slide valve stuck, controller broken). In an upset situation, the console operator may need assistance, and one of the field operators working in the field who is also console-qualified will go to the control room as quickly as possible. The other field operators and the head operator must cover in the field in his absence.

Other plant personnel influence the operation shift teams intervention activities and at times are directly involved in the response to an imminent or existing abnormal condition. Brief descriptions of these roles are provided to give the reader an indication of extent of collaborations among diverse roles in the process operations environment. These other roles include:

- **Shift Coordinator**—The shift coordinator (often referred to as shift supervisor). plays the role of operations team coordinator and management interface between the operations superintendent and the operations staff. He is responsible for monitoring the process and its impact on the unit's daily production targets. In addition, he is responsible for high level coordination of the console and field operations staff in meeting production goals, completing preventative maintenance, and coordinating abnormal situation responses as well as working with maintenance and instrumentation staff in prioritizing and executing lock out, tag out (LOTO) and maintenance activities.

- **Operations Superintendent**—The operations superintendent is responsible for the productive and safe operation of the complex for which he is responsible (i.e., a complex is typically run by multiple shift teams). The areas of responsibility include monitoring and reporting of the budget and costs associated with complex operations, safety reporting and documentation, environmental compliance and incident reporting, training, and production reporting to upper plant management. He is responsible for tracking high level plant operational goals and is ultimately responsible for meeting those goals

- **Site Planner**—The operations planner in some cases may represent a team of people where each is responsible for different areas of plant. A planner is responsible for tracking possible market opportunities (e.g., high demand, high price, scheduled shipments, weather conditions) that may arise along with planning for expected maintenance and turnarounds (I&T).

- **Process Engineer**—The process engineer(s) is responsible for generating the daily production orders for each process unit within a plant which are developed by the site planner. The process engineer troubleshoots process unit problems and determines why they are not making plan from a process (as opposed to an equipment) perspective

- **Control Engineer**—The control engineer is process and economics knowledgeable. S/he maintains control tuning, control objectives, and develops and implements improved control. The control engineer often troubleshoots process and control related problems after operations has stabilized the process. This troubleshooting often depends heavily on historical databases and anecdotal information from operators.

- **Maintenance Coordinator**—The maintenance coordinator (also referred to as supervisor or foreman) is responsible for coordination of maintenance activities for the plant units. The maintenance coordinator schedules periodic preventive maintenance, and maintenance requests put in by the operations team. S/he will determine the resource requirements, order any required materials, and determine if any contractors need to be hired to perform the job.

- **Maintenance Technician**—Instrument or mechanical technicians are responsible for maintaining and repairing all the process equipment.

Collaborative Decision Support Vision

The purpose of AEGIS is to support the plant operations team in achieving the four high-level objectives of ASM:

- Keep the process operating normally;

- *Failing this*, restore the process to normal operation;

- *Failing this*, bring the process to a safe state;

- *Failing this*, minimize severity of any accident.

To support operations in achieving these objectives, AEGIS needs to work with the operations staff to recognize abnormal situations, correctly diagnose problems, plan the best course of action, and carry out the plan correctly and efficiently.

Six Functional Roles. The AEGIS collaborative decision support vision is presented in Figure 2. The vision depicts AEGIS as six major functional roles referred to as: *State Estimator, Goal Setter, Planner, Executor, Communicator,* and *Monitor*. Briefly, the State Estimator role provides a concise dynamic estimate of what is actually happening in the plant, including trend information that may be used to predict future states. The Goal Setter examines the result and proposes a prioritized agenda of high-level goals to pursue. The Planner creates a plan to achieve those goals given the current plant state. The Executor carries out the plan, ensuring that each step is executed at the right time. The Communicator manages the interface between the other Roles and the plant personnel. The Monitor examines the operation of the other Roles and regulates their activities accordingly.

Figure 2. AEGIS Functional Roles
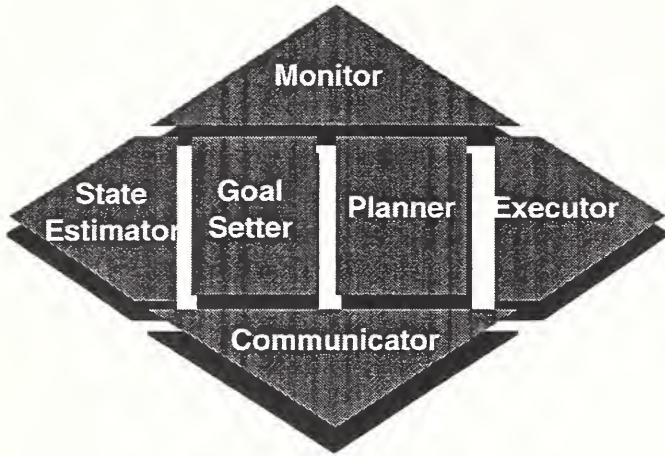
The following six sections describe AEGIS operation by discussing the individual functional roles. For each role, we discuss what it does and why this is important to solve the ASM problem. We also discuss how the role works and how it interacts with the other Roles and users.

The State Estimator Role

What it Does. The State Estimator provides a concise and dynamic estimate of what is actually happening in the plant. It reduces the torrent of sensor data available on the DCS every second to a narrow stream of explanatory states we refer to as *situations*. Situations attempt to capture the necessary and sufficient information to assess the health of the process and the equipment in the plant. The scope of the estimator encompasses the complete process area supervised by a single console operator along with major interactions with upstream and downstream process areas. This would include the mechanical condition of the process equipment, the maintenance activity within the area, the process control applications and systems, and the refining/chemical process itself. The estimator receives sensor inputs and historical information directly. Other relevant information needed as input (e.g., maintenance logs, operator observations, lab reports) can be accessed either from the communicator module (if these are available in an electronic form in different computers) or manually entered by the operations personnel.

The goal of the State Estimator is to explain as many plant events as possible with the smallest set of *situations*. A situation may represent a root cause of an abnormal situation, such as the failure of a piece of equipment or sensor, a process disturbance forcing control changes, a trend in process variables, or an incorrect manual intervention.
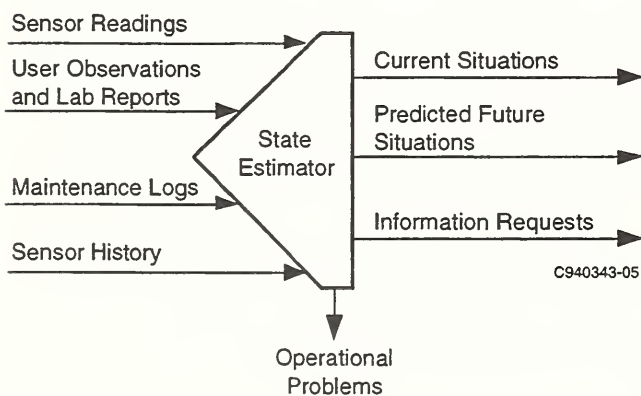
Situations are really a generalization of the notion of diagnosis, covering a wider range of explanatory levels. Some are quite specific. For example, the situation "Valve-32 is stuck in the 24% open position" is recognizable as a specific diagnosis. Other situations represent greater degrees of uncertainty, such as "Product quality from Tower-2 has been declining for the last hour." This is better described as a high-level symptom than as a diagnosis, but it may be as much as can be deduced at an early stage of an abnormal situation. All situations derived by the state estimator are qualified by estimates of their certainty.

402

The State Estimator's vocabulary of situations is defined by the needs of the other AEGIS modules. In essence, the situations of interest to AEGIS and to plant operators are those that may make a difference in plant operations. For example, instead of continuously presenting sensor measurements from the operation of a distillation tower, the State Estimator might condense sensor information to conclude that the liquid level in the tower was "too high to support X" when (and only when) X was a planned operation—and would pass along pertinent data to other AEGIS modules for planning and communication purposes. Generally, pertinent situations such as "pressure relief valve about to release" would always be monitored.

The State Estimator has several outputs. It is primarily responsible for deriving current situations—the instantaneous position of the process and equipment. However, with enough knowledge, predictions of future situations are also possible. Predictions are likewise qualified with certainty and time frame. The state estimator may be able to identify specific information it needs to make a better assessment; these are passed on as information requests. Finally, the State Estimator may be able to detect when it is failing. Such operational problems are also a form of output and are of interest to the Monitor Role.

How it Helps. The State Estimator impact on the ASM problem is made through the speed and accuracy of identification and diagnosis of existing or impending abnormal situations. The preventive detection and diagnosis of sources of abnormal situations will increase the time available for performing compensatory and/or corrective operational activities. Quick identification of abnormal situations and predictions of future abnormal situations allow operators to take compensatory actions before the problems escalate. A later explanation may contain the root cause description needed for a corrective action.

The AEGIS environment provides plant situation assessments from the State Estimator to the operations team by way of the Communicator module. Suggested remedial actions also are forwarded by the Goal Setter, Planner, and Executor.

How it Works. The AEGIS State Estimator incorporates a synthesis of some of the most successful approaches to sensor-based diagnosis and sensor validation, as demanded by the domain. In particular, we incorporate elements of heuristic methods as well as those based on models. Heuristic methods use the "compiled" knowledge of domain experts. Model-based methods use a model of the device or process being diagnosed along with a more generic reasoning method to derive a diagnosis.

Heuristic methods are fast and do not require a plant model, but are comparatively brittle—they cannot handle situations that were not explicitly anticipated. Model-based techniques are less brittle, but pose other problems. Most industrial chemical processes are unique. Consequently, it is difficult to build high-fidelity first-principles models of these processes and very difficult to anticipate what abnormal situations are common. Very few processes are amenable to theoretical modeling.

A complicating factor is that the sensors and actuators used to control the plant also fail, and not infrequently. One approach to sensor failure is to use redundant sensors and select a reading based on a majority vote or some combination. This technique has been widely used in the aerospace industry. Industrial process control operates under different economic constraints, and the application of triply redundant sensors is usually unrealistic in today's plants or in designs for the near future. In the absence of physical sensor redundancy, analytical redundancy (i.e., observer-based schemes) can be used to monitor the entire system—actuators, sensors and equipment—and capitalize on redundant information available in collections of different sensors.

One of the technical innovations we provide in the AEGIS State Estimator architecture is a framework that enables several of these promising techniques to work effectively together.

How it Interacts with Other Roles. The State Estimator provides fundamental information about plant health that is used by all of the other Roles. Although it receives its primary input from plant sensors (via the DCS) and users, it also accepts instructions in the form of requests for information or situation assessment from all of the other AEGIS modules. These may involve controlling its adaptation or allocating its resources.

Through the Communicator, the operations team members can interact with the State Estimator in various ways, including:

- Inspecting state explanations.

- Entering information not available through the sensor network.

- Directing an analysis of some specific data.

- Defining new state explanations for particular data patterns or relations.

The Goal Setter Role

What it Does. The Goal Setter examines the situations provided by the State Estimator and proposes a prioritized agenda of high-level goals to pursue. The goals are passed to the Planner to find the best way to achieve them. Because not all goals are simultaneously compatible or appropriate at a given time, the goal-setting process involves making some executive decisions about priorities and resources. These are based on external considerations in the context of the projected impact on process, plant equipment, and safety. For example, safety-related goals may cause production-related goals to be deferred in some situations.

Performance Feedback
Goal-Setting Problems
Top-Level Goals to Planner
Current Situations
Goal Setter
Goal Prioritization
Predicted Future Situations
Planning Problems
C940343-06
Manual Intervention
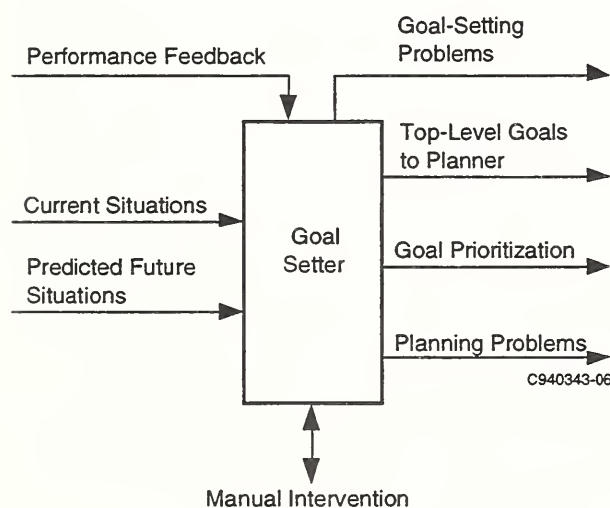
Key functions of the Goal Setter include:

- Identifying desired changes to the process state (target or safe situation to achieve).

- Finding alternative situations (in case the optimal one is unattainable).

- Assessing the future risks of the present situation.

- Evaluating the relevance of currently active goals.

- Setting the agenda for the Planner.

- Prioritizing goals and setting deadlines.

How it Helps. Critical problems confronting operations teams are communicating current operational goals, identifying the impact of goal modifications on operational objectives, and making consistent tradeoff decisions given multiple objectives.

The explicit representation and prioritization of goals in the AEGIS Goal Setter improve the consistency and efficiency of operations. Plant personnel can track tradeoff decisions and receive feedback on the impact of goals on abnormal situations. Allowing operations staff to interact with the goal-setting process helps their understanding of the plan and enables them to work better as a team.

How it Works. The Goal Setter must convert a set of identified situations into a set of goals to achieve. In the simplest case, a situation corresponds to a unique and certain diagnosis of a known problem. For example, the situation might be "Runaway reaction in vessel 7." The possible goals addressing this situation may be retrieved from a situation-indexed data/knowledge base. In the example, the only sensible goal may be "Quench runaway reaction." Information may also be retrieved about the expected costs and resources required for dealing with the situation, and the time course and consequences of not dealing with the situation. Continuing the example, it may be noted that quenching the reaction might take the process off line for anywhere from 2 to 4 hours, and that failure to quench the exothermic reaction in the next 15 minutes could ruin the $100,000 reactor.

As noted, goals may interact in a positive or negative way. In some cases, the interaction between goals may be known to the Goal Setter. The "quenching" goal might be known to be incompatible with a goal to increase production from this vessel. In other cases, the interaction could be more subtle, and the Goal Setter would rely on the Planner and the operator to detect negative interactions between proposed goals. For example, part of the quenching procedure may demand cooling water that is urgently needed elsewhere in the plant. The planner communicates these as "planning problems" to be resolved by the Goal Setter.

Part of the Goal-Setter's job is managing uncertainty. Uncertainty enters from the assessment of the current plant situation and from forecasts of future events. Suppose, for instance, the situation "Runaway reaction in vessel 7" was diagnosed by the State Estimator with likelihood 0.125. Given these odds, should the Goal Setter launch a "Quench runaway reaction" goal?

Another Goal Setter responsibility is noting when the motivation for a goal disappears. Some problems are corrected without the completion of a planned remedial action. This occurs since AEGIS cannot know all possible causal relations in the plant; the situation can be resolved for unknown reasons, or the operator may have had other means to solve the problem. In such a case, the goal setter needs to establish a goal to safely abandon a plan that may be under way. This may entail "clean-up" actions (e.g., restart halted equipment).

How it Interacts with Other Roles. As primary communications links, the Goal Setter receives situation information from the State Estimator via situations posted on the shared AEGIS blackboard. It provides an agenda to the Planner. The goal agenda can contain new situations to be monitored and these may need to be conveyed to human operators for approval or revision. Thus, as secondary communications links, pertinent aspects of the goal agenda are conveyed to the State Estimator and to the Communicator. The operations team members are able to interact with the Goal Setter by:

- Viewing the current prioritized agenda of goals.

- Asserting, modifying, or retracting individual goals or priorities if so authorized by plant management.

- Receiving feedback on the impact of prioritized goals on plant productivity.

The Planner Role

What it Does. The Planner takes the high-level goal agenda from the Goal Setter and generates a plan to satisfy it. The plan is elaborated as a partially ordered set of simple actions to be carried out by the Executor. The Planner uses special reactive techniques, since speed is essential to react to urgent situations.
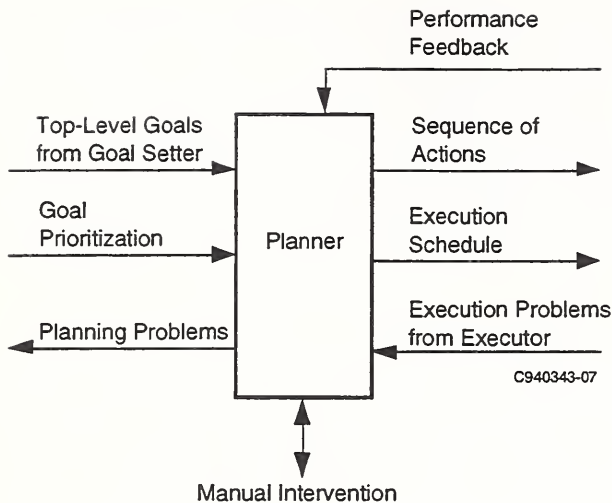
Key functions for the Planner include:

- Finding and scheduling a sequence of actions that satisfy all goals on the agenda.

- Detecting negative interactions between plan steps.

- Resolving choices between alternative courses of action (feasibility, risk, cost).

- Calling for diagnostic tests or other information required for planning,

- Determining key variables to monitor (especially process constraints) and their time profile while trying to execute goal sequences.

- Verifying plan steps to help monitor the plan success.

How it Helps. Abnormal situations are often caused or worsened by the application of inappropriate procedures. The Planner generates action plans, appropriate to the current situation, which can be modified as appropriate by the operations team. Over time, the system adapts to the specific planning needs of the operations team and improve the adequacy of procedures to ensure better responses to abnormal situations.

How it Works. The Planner adapts the best of emerging technologies to fit the real-time demands of this domain. Although there are many approaches to plan generation, including table lookup, task reduction, macro-operators, and planning from first principles, the real-time nature of the ASM planning problem and the need for replanning as the situations change imposes special constraints. Traditional task reduction (hierarchical planning) would require enormous knowledge-engineering task effort. The consequences of integrating different schemas in real-time plans are difficult to anticipate. Planning from first principles requires modeling of individual, atomic actions and makes it harder to encode effective guidance on how to construct good plans. These problems can be eased significantly if we can work with partial models and rely on users to critique plans of action.

From analysis of existing plant procedures, we believe that successful planning in this domain can exploit a *procedure library*. In some cases, this could be as simple as a database of procedures indexed by situation. The procedures can be parameterized to handle variations in the equipment or resources involved (a limited form of context sensitivity), or could consist of fragments to be combined into a plan or hierarchy of task "schemas" describing how to accomplish tasks under certain conditions. Each of these approaches has limitations, but good balance may be achieved via a combination of complete procedures for the majority of conditions, with additional fragments incorporated as necessary. All of the plans are parameterized, and the Planner can incorporate operator assistance in constructing and critiquing new plans.

How it Interacts with Other Roles. The Planner receives a top-level agenda from the Goal Setter. It may also get deadlines for producing plans from the Goal Setter, forcing it to react in a timely fashion. The plans produced are passed to the Executor to be put into action. The Planner uses situation information from, and posts monitoring requests to, the State Estimator. The Planner may need to ask the Communicator for human authorization or modifications to suggested plans, and it may need information about the availability and whereabouts of plant personnel. When the Planner is incapable of producing a plan for a goal it has been given, the problem is reported to the Monitor or the Goal Setter as appropriate.

Operations team members are able to interact with the Planner to:

- View the current partially ordered action plan.

- Authorize, command, forbid, or halt individual plans if so authorized by plant management.

- Assert, modify, or retract actions or the order of actions in the plan.

- View/modify the scope of responsibility for individual resources within the plant.

- Receive feedback on impact of the action plan on plant productivity and abnormal situations.

- Define new action plans for specific goals in terms of alternative actions and constraints on use.

The Executor Role

What it Does. The Executor supports carrying out plans. It issues the lowest level commands that actually cause changes in plant systems: to plant actuators, to other automated systems via the DCS, and to plant operations personnel, such as field operators, via the Communicator module. The Executor, in conjunction with the Planner, constructs a detailed schedule of actions and, in conjunction with the State Estimator, monitors their performance to ensure that each step is carried out at the right time.
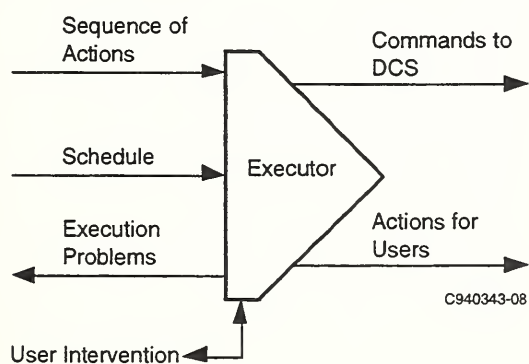
How it Helps. In addition to monitoring for failure of plant equipment to perform as expected, the Executor addresses a major source of ASM costs, failure to follow procedures. Human operators under stress can omit steps, add steps, execute steps out of sequence, enter wrong values without confirmation, and fail to confirm actions completed by another team member. When procedures are carried out incorrectly, remedial actions are required. Some simple remedial steps may be stored in the Executor itself; others may require more complex replanning, or revisions to goals involving the Goal Setter.

How it Works. The Executor requires detailed knowledge of procedures—how to perform plan steps in the operating environment of the specific plant. This must include knowledge about plant

personnel and their capabilities and duties, and information about resources required to execute these steps, including time of personnel.

Execution error monitoring is more demanding. The Executor matches observed states to expectations. There are two general approaches. Reactive error trappers catch either bad states or unexpected actions once they happen. Proactive error trappers monitor for series of actions that lead to an "error pattern" and try to halt them along the way. The Planner may have anticipated certain failure possibilities and inserted explicit "plan-checking" steps.

How it Interacts with Other Roles. The operations team members are able to interact with this Role in the following ways in accordance with their individual privileges and responsibilities:

- Initiate and/or confirm action command execution.

- Modify allocation of actions to resources including between system and humans.

- Indicate status of actions requiring manual response outside of system observations.

- View status of action commands including verification of command reception.

- View command queue for individual resources within the plant.

- Prompt or remind operations team member on need to perform manual task.

- Verify appropriateness of action.

The Monitor Role

What it Does. The Monitor evaluates the overall effectiveness of the coordinated activities of the other Roles. The object of the Monitor is the rest of AEGIS, not the plant per se. It identifies bottlenecks and inefficiencies in overall system performance and provides feedback to the other Roles to allow AEGIS to automatically adapt and improve.
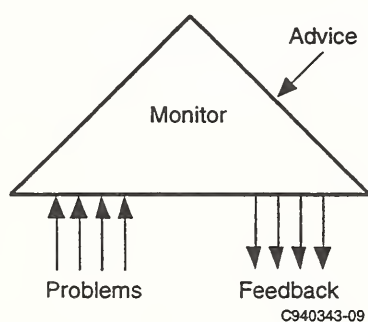
How it Helps. Some of the shortcomings in current ASM practices stem from a lack of reflection on how similar situations were handled (or mishandled) in the past. The AEGIS Monitor is a tool for evaluating the effectiveness of ASM practices. Through the tracking and documentation of ASM activities, operations personnel and AEGIS Roles are able to adapt and respond in a more effective manner to future situations.

A simple example of the Monitor in action involves improving the knowledge base of the Planner. Suppose the Planner assumes that pressure in a particular boiler can be reduced at the average rate of 50 psi/minute. The Monitor may notice that the last three times this procedure was attempted, the achieved rate was only 30 psi/minute. The Monitor could present such discrepant observations to plant personnel. Engineering might decide to the change the Planner or look for some other cause for the slower pressure release rate.

How it Works. The Monitor tracks activities of AEGIS in interaction with human operators by monitoring problems reported by AEGIS modules. It also monitors the statistics of normal inter-module communications. From this trace of plant operations, the Monitor attempts to detect internal problems and isolate patterns of suboptimal performance.

A key activity of the Monitor is to track success and failure rates for goals, plans, and actions. Plans and actions that routinely fail are inspected in more depth for either remediation or removal. The same sort of scrutiny applies to top-level goals that are routinely rejected by operations. A particularly interesting case exists when goals are met without a known plan succeeding. This implies either that an intermittent, self-repairing problem is present (an important trigger for planning and executing diagnostic activities) or, of more interest for the Monitor, that plant operators are using a plan or activities that AEGIS does not know about. In the latter case, machine-learning or knowledge-acquisition techniques can be used to acquire this novel plan and integrate it into the AEGIS knowledge bases.

We anticipate that the bulk of the Monitor's assessment and remediation activities will not occur in real time. Instead, the Monitor does its work off line, and remediation recommendations are provided when the Monitor has sufficient cause.

How it Interacts with Other Roles. The Monitor "reads" messages passed back and forth between all of the AEGIS modules. It interacts with the Communicator to notify users of detected problems and remedial actions, and with the Knowledge Customization Toolkit to facilitate the acquisition of new knowledge and the appropriate incorporation of that knowledge into AEGIS. It provides automatic feedback to those other Roles capable of accepting it. Operations team members are able to:

- View system self-evaluation of overall and individual Role performance.

- Enter user evaluations of AEGIS actions to improve future performance.

The Communicator Role

What it Does. Each of the AEGIS Roles needs to interact with users at various times. Moreover, they are all running simultaneously. To give each a separate interface would lead to confusion. The Communicator module provides a common user interface to AEGIS. It also serves to isolate the modules from the details of particular display and communications hardware that may vary from one installation to another.
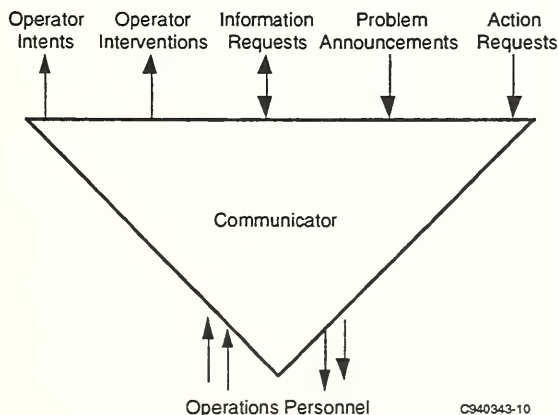
The Communicator manages interface devices to make the best use of users' attention. Messages arrive asynchronously from other modules; some may be much more critical than others. The Communicator brings the critical items to the attention of the appropriate users. For example, if the next planned action demands an operator adjustment to a rarely used valve, the Communicator can summon the controls for that valve instantly. Each member of the operations team can quickly receive information on plant state, goals, and context-appropriate actions that are pertinent to their activities, improving speed of

response and eliminating sources of human   error.

How it Helps. Poor access and presentation of information has been identified as a serious bottleneck in abnormal situations. Both problems are addressed by the Communicator role. Because AEGIS is always aware of the current state of the plant and the immediate plans for control, presentations and access to displays can be made context-sensitive. The Communicator is aware of the operator's information load and capabilities, and the amount of information on a given display surface can be regulated to avoid confusion and excessive workload, such as the flood of alarms described in the earlier steam plant operations scenario.

Communications problems are another major contributing factor in abnormal situations. The issue is often the speed and accuracy of information transfer. The increased noise level in the control room, from reaching key personnel and handling low-priority radio traffic, makes conversations difficult. In addition, finding out who to call and how to reach them adds valuable time to performing interventions. The Communicator can improve this situation by automating some of the message transmission tasks.

How it Works. The Communicator makes use of existing, human-computer interface devices by deciding when and where information is needed to best support operator tasks and then presenting that information in the best format available. The Communicator makes use of familiar formats (e.g., those available in the DCS environment) whenever possible, and makes use of novel formats via existing equipment (e.g., novel plan scheduling screens, vocal instructions, or advice via telephone).

The Communicator also keeps track of the users' current active interface devices. The Communicator uses this information to route appropriate information to the relevant devices. In some cases, those devices might be radios or telephones; the Communicator generates or forwards voice messages as needed. For example, it knows that Jane Smith is the current field operator, that she was last sent to Holding Tank 81, and that her radio monitors channel D. An emergency near Tank 81 may trigger AEGIS to plan special actions for all field operatives in the area. The actions are passed to the Communicator, which radios the message to Jane Smith on channel D.

How it Interacts with Other Roles. The other modules forward their user interaction needs to the Communicator. The Communicator selects among information available and transactions requested according to priorities, resources, and preferences of the individual human operator in his or her current context (equipment available, ongoing goals and plans, concurrent activities and workload, etc.). Requests for information to be displayed or collected are addressed to individuals serving a given job function. The Communicator knows who those people are and keeps track of where they can be reached at any time. Some of the modules may keep open channels with the Communicator so that users can provide unsolicited information and commands to AEGIS.

If the Communicator is unable to complete a transaction within the allotted time, it signals a problem to the Monitor. User intent information is sent to the Planner and Goal Setter. The users may request desired information directly via the Communicator, which may in turn need to post these information requests to the State Estimator (or other Roles) in order to comply.

Key Technical Challenges to Delivering Effective Solutions

All operations personnel must have better decision support from the DCS if abnormal situations are to be efficiently and effectively managed. The AEGIS collaborative decision support solution requires that technical challenges be overcome in three strategic areas:

- **Human-machine interaction**—A comprehensive approach to the design of the human-machine system interaction is needed so that a single user interface environment provides operations personnel with information appropriate to their needs, while at the same time

supporting the collaboration of appropriate members of the operations staff in solving the problem as a team.

- **System architecture**—To provide accurate, timely support in abnormal situations, a system architecture is needed that supports multiple processing modules, data bases and knowledge bases. These various software modules must communicate their conclusions with each other in real time and must remain coordinated among themselves and with human operators. Many past efforts have failed because this problem alone is so challenging.

- **System customization**—A major practical challenge in collaborative decision support technologies is configuring their capabilities to the idiosyncratic and dynamic nature of the plant processes and operations. Aspects of the software modules need to be customized with specific knowledge about the operations, equipment, personnel, and procedures of a specific site. Acceptable solutions need to be self-adaptive or easily customized by plant personnel.

Collaborative Human-Machine Interactions

AEGIS's functions are distributed over a set of six modular roles as described in the AEGIS vision. These specific roles were identified to provide operations assistance in those critical areas where recurrent behaviors occur in ASM practices (Reason, 1986, Cochran and Nimmo, 1997). Collectively, these modules provide automated assistance for the human activities of Orienting, Evaluating, and Acting, along with a mechanism for communication with plant personnel and a mechanism for modifying and improving AEGIS behavior. The AEGIS concept emphasizes human-system *collaborative* interaction; thus AEGIS will continuously interact with the plant operations team by configuring information as appropriate for the task of individual human users to facilitate the coordination of their operations. AEGIS will allocate tasks between people and computers based on the strengths of each. For example, AEGIS will improve Orienting by summarizing large amounts of data, Evaluating by resolving complex problems under pressure and formulating schedules, and Acting by managing or replacing human actions in time-critical situations. The development of AEGIS will restructure the ASM tasks to use the strengths of the computer (automated data analysis, rapid computation, etc.) to augment the strengths of plant personnel (flexibility, creativity, common sense). An important aspect of our technical approach is to recognize that ASM is an integrated activity. Hence, we have developed a user interface framework that provides the operations team a single window on operations, a task-centric layout of views, and mixed-initiative approach to human-machine interactions.

Single Window on Operations. The demands of process control, and in particular the need to interact with hundreds of instruments without adverse impact on the operators' awareness of the overall state of the plant, led the designers of distributed control systems to develop the "single window to the process" concept. In the AEGIS program, this concept has been extended to include all computer-based interactions supporting operations team members' ASM activities. This functionality includes access to current DCS operating display functionality, interactions with AEGIS diagnostic and decision support applications, logbooks, plant procedures, and other plant documentation. This design principle requires us to ensure that all interaction with the process take place in a unified, consistent, and comprehensive user interface. As new capabilities are added to the process control system and AEGIS components, they are required to be integrated into the existing user interface environment.

Task-Centric Layout. The AEGIS view library provides a suite of display views that comprise a set of information and interaction objects that covers the range of normal and abnormal situation management activities. Together, the suite provides a task centric structure for organizing custom graphic views. Based on an analysis of the kinds of activities operators engage in for both normal and abnormal operations, we defined a set of view types to support activity clusters (See Table 1 for brief description of each task view.

Mixed Initiative Interactions. For a variety of reasons, many decision support systems have been criticized for being to brittle as well increasing the possiblity of human error (Smith et. al., 1995). Consequently, we have taken a mixed initiative approach to the design of interactions with AEGIS components to enable a more collaborative human-machine interaction. The mixed initiative functionality enables the users to influence the behavior of the machine, and at the same time, the machine can influence the behavior of the user. For example, in the traditional alarm notification systems, the engineers preconfigure the system to detect a set of process conditions in which the operator needs to be alerted because some important action is required to prevent the occurrence of undesirable situation. The condition under which the machine sends a notification to the user is fairly static. Moreover, it is difficult to accurately anticipate all situational needs for notifications. This makes the preconfiguration model less than adequate for ASM. In practice, operators may get too many notifications or too few for a given plant situation. In reality, the conditions under which notifications are needed are quite dynamic. Users need the ability to tell the notification systems what conditions are important at for the current situation and intervention tasks. Hence, we have created a mixed initiative approach. Where the user can take initiative to define a notification object to help them perform a particular task. Moreover, the AEGIS system can determine a need for a notification and create a new notification monitor.

Table 1. Set of Task Views designed to support Operator Interactions with DCS, AEGIS, and Plant Information Systems.

View Type	View Description
Status	Operators get a summary view of the status of plant processes that enables orientation to the location and priority of disturbances and unexpected process behaviors. This view spans the scope of control of a single operator as well as the process units immediately upstream and downstream.
Operation	Operators access a broad range of functionality in a single display that is the equivalent functionality found in today's schematic, group and detail displays. The equivalent of a hierarchy of views on the plant exist in a single display with different kinds of information and interactions emphasized at each level.
Diagnosis	Operators can diagnose the cause and impacts of disturbances indicated in the Status and Operations Task Windows. With the advent of diagnostic applications and smart devices that can assist the operator in diagnosing abnormal conditions, this view allows the user to see the evidence associated with diagnostic and alarm notifications, evaluate and rank the highest priority threats (disturbances), and identify the source of threats in terms of root causes.
Notification	This display integrates the current process alarm notification events (found in current DCS alarm summary displays) with other types of notification events supporting diagnostic, decision support and collaborative team communications.
Trends	This process monitoring view allows the user to configure a set of real-time trends to proactively assess status and behavior of the plant. Although, trends are available in other task windows, the trend functionality is tailored to the specific kinds of activity supported in those windows.
Decision Support	Operators can evaluate the significance of current threats to operational goals, establish plans to respond to imminent or current abnormal conditions and track progress in executing intervention activities.
Logs	Operations team members record and browse significant plant changes or activities. Moreover, this log reporting system is integrated with an incident/near-miss reporting system.
Documents	In addition, operators dynamically assemble information through access to the myriad of documents in the plant information systems. These documents include drawings, equipment specifications, procedures, incident reports and operating instructions.

System Architecture

A key innovation of AEGIS is a system architecture that supports the collaboration of a set of independent diagnostic, planning, and operator support applications. Previous approaches to solving this problem have attempted to build comprehensive, monolithic applications, which inevitably have become large and difficult to maintain. AEGIS, on the other hand, is designed to permit a variety of specialized applications to work together to identify problems and aid in their resolution. In this section, we describe to aspects of the architecture that address problems associated with previous monolithic approaches.

Nonmonolithic—Distributed, modular architecture. Each of the six AEGIS roles is implemented as a set of software application modules (See Figure 3). The AEGIS modules must be capable of distributed and partially independent action due to the variable nature of data flow and processing demands on each module. On the other hand, all modules must cooperate to produce effective overall behaviors—this requires communication of results and needs of information between the modules. In order to address these potentially conflicting needs, we developed the Plant Operations and Control Language (POCL) and instantiate it in a blackboard-like communications architecture which serves as a common area for exchanging messages and information posted there by each of the separate AEGIS modules.

These modular software applications work together to determine the current state of the plant, decide upon the most appropriate goals to pursue, develop plans for pursuing those goals, and for executing those plans and monitoring the execution process. In addition, applications are responsible for communicating with plant personnel and for monitoring AEGIS itself.
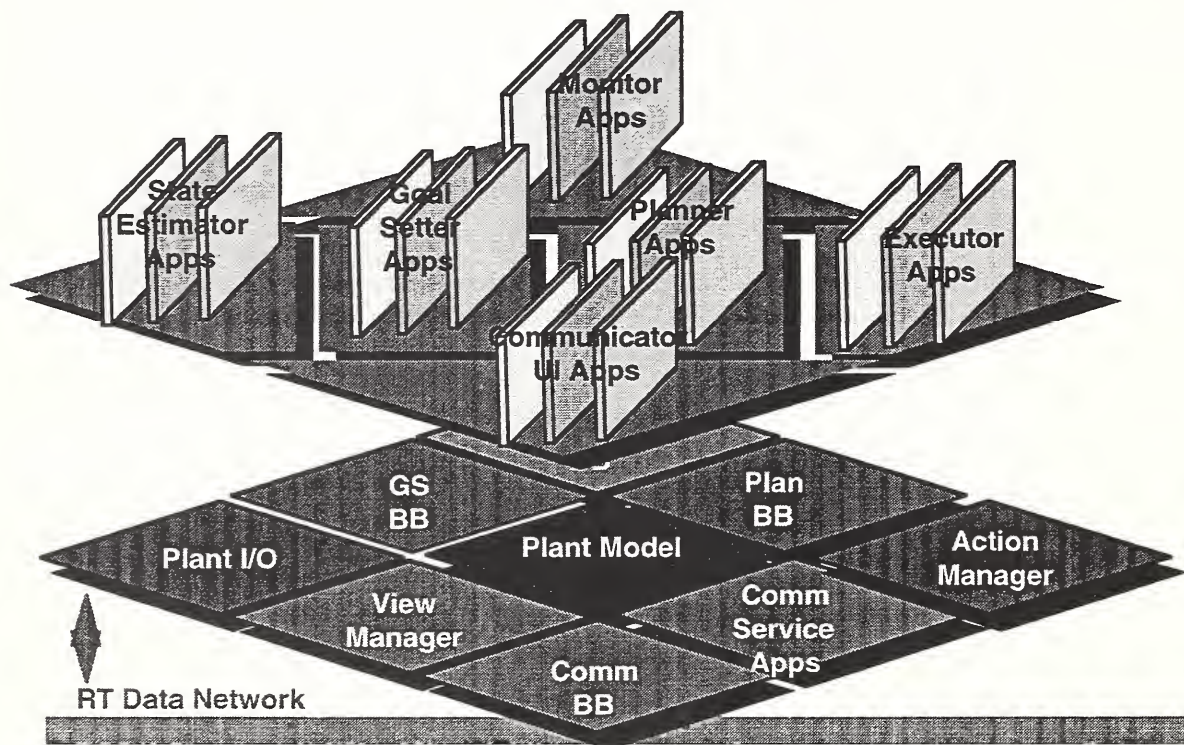
Figure 3. Distributed Modular Architecture with Multiple Component Applications for each of the Functional Roles

Extensible—Open API for Application Developers. The AEGIS prototype is being developed in a layered architecture based upon an open standard, and so will run on any DCS which supports that standard. The software consists of an infrastructure supporting many different kinds of applications, all sharing a common Plant Operations and Control Language (POCL) and a common blackboard on which messages in that language are shared.

The Plant Operations Communication Language (POCL) is built around human-level concepts in the operations domain to facilitate knowledge acquisition and later knowledge revision. The POCL specifies the major conceptual classes such as situations, *plans, goals, actions* as well as others. Each of these classes contains a hierarchical set of entities or objects that, in effect, comprise the common *vocabulary* by which AEGIS modules communicate with each other. The internal reasoning a module performs is in any format convenient for that module, but the module must express its conclusions and requests for information in the vocabulary of the POCL. Conclusions

and requests are posted to a central blackboard or shared memory area, from which they can be accessed by all other modules. Specific messages posted to the blackboard by the various modules are generally concerned with the activation or deactivation of a specific conceptual entity (situation, goal, plan, or action) at some time and for some expected duration. For example, a message from the State Estimator might be: ACTIVATE SITUATION: VALVE-STUCK at TIME 20:14:53 with parameters 106 (the specific value referenced) and 76% (the point at which the valve is stuck). Figure 4 shows the operation of the POCL by illustrating the flow of information between AEGIS modules via a common message area, the *blackboard*.
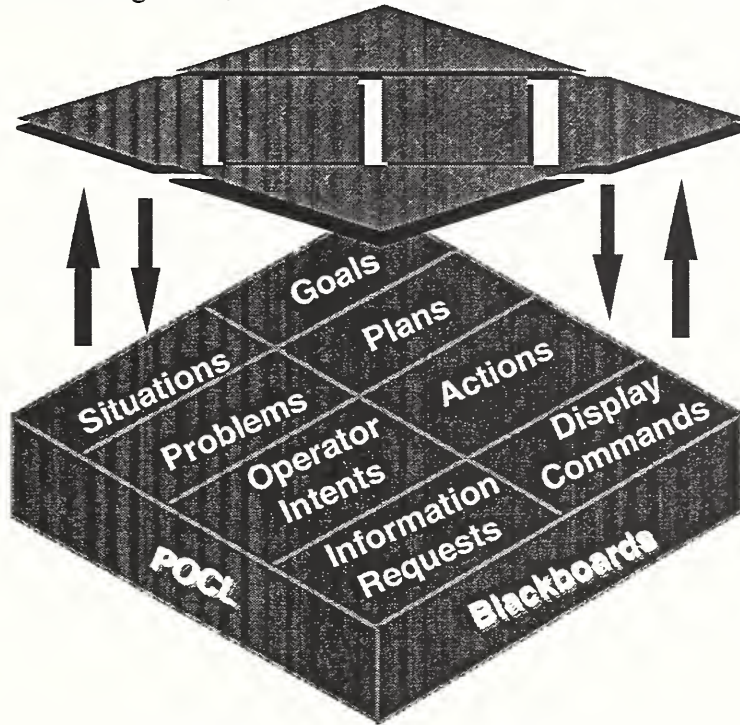
Figure 4. Information Flow between AEGIS Modules Using the Vocabulary of the Plant Operations Communication Language

There are only two general classes of alternatives to a POCL and blackboard approach to AEGIS system communications. The first is to construct a unified (non-distributed) AEGIS system that cycles through its modular functions, passing the results of one module (e.g., State Estimator) to the next module in line in a serial fashion. The problems with this approach are (1) the required delays while the system cycles through its modules (and coordinating incoming data from multiple sources during those cycles), and (2) the fact that the messages passed from one modular function to another must still be stipulated in some vocabulary in advance. The second alternative is to stipulate direct and dedicated message-passing protocols between each module without resorting to a blackboard (or similar) architecture. Although this approach still requires the creation of a vocabulary of acceptable messages, it does permit a more distributed or parallel processing approach that facilitates real-time operations.

While dedicated, direct message passing between AEGIS modules might be possible and might improve speed of system operations, we believe the modularity of the POCL and its blackboard architecture has advantages that outweigh these. Modularity in the POCL makes it easier to modify plant operating procedures, to revise or "grow" these procedures when equipment modifications are made, and to monitor and revise AEGIS operations via human assistance. For example, if a new set of diagnostic procedures improves the reliability of the State Estimator's performance, no

changes need be made to AEGIS knowledge bases as long as the State Estimator can continue to communicate the same vocabulary of situations to the other AEGIS modules. If the improved diagnostic capability means that new plans, goals, and actions are now possible for ASM (or that new information is needed by operators), then only small, isolated portions of each AEGIS module's knowledge bases need to be changed to handle these new capabilities—and the hierarchical nature of the POCL facilitates the tracing of affected components. Finally, the fact that the entire representation is closely tied to human operators' concepts of plant operations serves to make the capture of new knowledge straightforward.

Customization Tools

One of the challenges in making AEGIS a cost-effective addition to future control systems is to minimize the effort needed to customize the AEGIS component applications to the specifics of a given plant. In future phases of the program, we will develop these customization tools. One promising approach is to develop a single model of the plant that can be shared with the all AEGIS roles and the DCS. This model contains generic plant equipment descriptions that can be tailored to the specifics of a given plant. Another element of a tools solution will be to provide tools to rapidly exploit existing plant information containing process flow diagrams, plant piping and instrumentation diagrams, and operating procedures.

Summary of AEGIS Development Progress

In Phase I of the program (completed March, 1996), the first AEGIS prototype was developed which comprised the software infrastructure of AEGIS, a suite of State Estimation applications, and early prototypes of the AEGIS console operator's interface. The prototype was linked to a simulated refinery unit and tested on a variety of induced problems. AEGIS was able to consistently identify the problems to which it was exposed, and the AEGIS architecture was qualified for further development in Phase II of the program.

In Phase II (completed March, 1997), the functionality of the AEGIS prototype was significantly enhanced to incorporate a more comprehensive process model and a more advanced prototype of the AEGIS console operator's interface. The new interface was designed for operations during normal and abnormal situations, and included a consistent user interface model and comprehensive support for notifications, navigation, diagnosis, and monitoring. The process model was expanded to incorporate a Vapor Recovery Unit (VRU) in addition to the FCCU modeled in Phase I; the model's complexity has increased by more than a factor of two. The scope of knowledge of AEGIS was expanded to include over 700 items of equipment in the process simulation, over 400 operational goals, and over 4500 possible malfunctions. AEGIS was able to consistently identify the problems to which it was exposed, and present a coordinated view of the problem to the operator. The AEGIS architecture was qualified for further development in Phase III of the program, which will focus on refining the operator interface applications and developing methods of collaboratively generating, executing, and monitoring recovery plans.

AEGIS has been designed as a very modular, very distributed system, intended for use in an open, multi-vendor environment. The core of the second AEGIS prototype consisted of infrastructure for providing system services, acquiring process data, and providing blackboards for inter-application transactions. A suite of state estimation applications employing a variety of sensor validation and process diagnosis technologies (including knowledge-based systems, sign-directed graphs, and quantitative and qualitative methods) worked to identify problems in the simulated process. A suite of communication applications worked to present process data to the console operator. These applications, collectively called the Communicator, presented a view of the process that is sensitive to the operator's workload, the process conditions, and the operator's current knowledge. The Communicator integrated functions required for understanding and responding to process upsets with those required for monitoring and controlling the process during normal operations. The

Communicator also provided assistance to the operator in diagnosing process problems. The prototype was implemented using Gensym's G2, C and C++, Perl, Microsoft's Visual Basic, Java, and a few other specialized applications and languages.

The AEGIS testbed now consists of:

- two simulated refinery units (FCCU and VRU),

- a Honeywell TDC3000 Distributed Control System,

- a variety of networked workstations (UNIX and Windows NT OSs) running 25 component applications that comprise AEGIS itself,

- a prototype operator console, and

- a variety of supporting hardware and software.

The simulated refinery unit consists of a Fluid Catalytic Cracking Unit (FCCU) and its associated fractionator, feed and air heaters and blowers, waste heat boiler, and compressors as well as a Vapor Recovery Unit consisting of five towers and associated heat exchangers, pumps, reboilers, and other equipment. The simulation has been constructed by combining and extending first-principles simulations used in Honeywell Hi-Spec Solutions' SACDA Trainer. It provides real time data about, and access to, over 450 sensors and actuators. The simulation runs on a DEC Alpha, and is accessed by the TDC via a dedicated network link.

The TDC provides for traditional DCS control capability for the simulated FCCU, including a user station providing schematic, group, point, and alarm displays of the unit, and regulatory, supervisory, and advanced control algorithms for the unit itself.

The AEGIS prototype runs on a variety of Sun , HP, and Pentium hardware using Solaris, HP-UX, and Windows NT operating systems. AEGIS acquires its data via a link to an AxM module on the TDC 3000, and displays its conclusions via an operator's console consisting of several networked Windows NT and Honeywell GUS systems connected to 21-inch monitors equipped with touchscreens, as well as a 40" Seufert Systems Overview display.

The first AEGIS prototype was formally evaluated in 22 scenarios including sudden and unexpected malfunctions, problems that developed over the course of an hour, and problems originating in process equipment, process control devices, and the process itself. Fast upsets involved a deviation of a process parameter of up to 15% of range over a one minute period; slow upsets took 60 minutes to reach the same level of deviation. In all of the test scenarios, the AEGIS prototype was able to take advantage of the diverse strengths of its variety of state estimators to consistently identify the broad variety of problems that were presented. Moreover, AEGIS was able to identify the slowly developing problems well before they became significant (on average before drifts reached 5% of scale).

The second AEGIS prototype is currently being evaluated by operations personnel from Consortium Member companies. The evaluation protocol aims to assess the usability of the AEGIS Communicator, and to empirically determine whether operators using AEGIS are more capable in responding to a variety of process upsets than operators using current control systems technology.

A key feature of AEGIS is an architecture that supports the collaboration of a large number of independent diagnostic, planning, and operator support applications, all working together to solve the problem and keep the operator in control. In Phase I, this approach permitted AEGIS to combine diagnostic approaches which are powerful in some areas but less so in others to enable an overall diagnostic ability that is more accurate than any single application could be. In Phase II, we

extended this approach to user interface functions, permitting specialized user interface applications to work together to present the operator with information that would otherwise have to be laboriously extracted from multiple independent sources. This architecture will ultimately support the interaction of process operators with each other, with field operators, and with other plant personnel.

AEGIS has been developed using an open architecture so that state estimators, user interface applications, and ultimately planning, modeling, documentation, and data management applications can be developed and incorporated by Honeywell, Consortium members, and independent parties. The current prototype was designed and developed by user interface specialists, process engineers, and software engineers from the Honeywell Technology Center, Honeywell Industrial Automation and Control, and Honeywell Hi-Spec Solutions, with significant assistance from Amoco, Exxon Gensym, and Texaco.

Issues remain to be addressed in several areas, and work will continue in Phase III: State estimators need to be refined; tools for easy installation and application of AEGIS Roles need to be developed; User interface functionality needs to be extended and enhanced; and the entire realm of planning and response needs to be integrated.

The ASM Consortium remains committed to this effort, and believes that its continued investment and that of the NIST Advanced Technology Program will ultimately lead to a solution resulting in a reduction in the costs of petroleum processing and chemical manufacturing and a significant benefit for U.S. economic growth.

References

Cochran, E. and Nimmo, I. (1997). Managing abnormal situations in the process industries I: automation, people, culture. *Proceedings of the MVMT Workshop*, Ann Arbor Michigan.

Nimmo, I. (1995) Abnormal situation management. *Process and Control Engineering*, **49**, 5, 8, November, 1995.

Nimmo, I. and Cochran, E. (1997). The future of supervisory control systems in the process industries: Lessons for discrete manufacturing. *Proceedings of the MVMT Workshop*, Ann Arbor Michigan.

Reason, J. (1986). Recurrent errors in process environments: Some implications for the design of intelligent decision support systems. In E. Hollnagel (Ed.), *Intelligent Decision Support in Process Environments*, Berlin: Springer-Verlag.

Smith, P.J., Guerlain, S., Smith, J.W., Denning, R. McCoy, C.E., and Layton, C.F. (1995). *Theory and practice of human-machine interfaces*. Paper presented at the Intelligent Systems in Process Engineering Conference, Snomass Village, Colorado.

Future of Supervisory Systems in Process Industries:
Lessons for Discrete Manufacturing

Ian Nimmo
Honeywell Industrial Automation and Control

Edward L. Cochran
Honeywell Technology Center

This paper provides a brief review of the history of process control and automated systems and an overview of what we can expect in the future, together with some lessons that may be applicable to the development and operation of highly automated discrete manufacturing systems.

Supervisory systems in the past

During the early 1970's—the beginning of the computer control era—memory and disk space were very limited, and the interface to the computer was primitive. The initial control capability provided what we describe today as Supervisory control. The computer had limited output capability, but could influence instrumentation systems by changing the setpoint to PID electronic controllers.

Early applications

Starting up a continuous polymer process is challenging: The characteristics of the product change as operating conditions change, and they also change as a function of plant throughput. These changes require different process control setpoints. In those days the operator used to manually tweak the controller setpoints but would often overcompensate and introduce new disturbances to the process. The early computers helped by automatically changing the setpoints to the controllers. Within a couple of years technology development enabled the computer to carry out full PID control using a mathematical model of a PID controller. This allowed not only automatic changes to the setpoint but changes to the tuning constants of the control loops as well. This was especially useful for non-linear processes.

However, the user interface was still very crude: Plant supervisors used to enter data by identifying the address of a variable and entering data by first entering the address and then the data, using a set of 16 piano-style keys. The process was very dangerous, because the opportunity for error was so high, especially when we consider that the data was not entered in normal base 10 numbering but in octal (base 8) or binary formats.

Soon more powerful monolithic control computers were introduced but were limited by small amounts of memory and little or no disk storage capability. The Supervisory Control Computer System held the main program and instructions for the "target" control systems. The Supervisory system downloaded the software to the control system and the supervisor initiated a booting or

---

startup sequence. During plant operations, the target system ran its PID control algorithm and fed information about the process back to the supervisory computer system. Other information, such as batch records, historical plant data, and any abnormal condition reports was recorded manually.

The control system user interface improved slightly, allowing operators to change the control computer setpoints and alarm values directly at the control computer system. However, this required regular updates back to the Supervisory system to allow easier synchronization should the computer shut down and require reloading of the currently running program and data. This meant that the Supervisory Control Computers required several versions of the controller software, including an original or "virgin" image, a copy of the original image which had been modified during operations and which held the latest setpoints and other data parameters, and any development or test versions of the software.

During the late 1970's the Supervisory system became a very powerful system, especially on batch processes, and as Digital Corporation's VAX computers and commercial VMS operating system became available and cost effective, the use of off-the-shelf database products improved the historical data capture and replay capabilities.

As Supervisory systems continued to evolve, Honeywell introduced a more powerful Distributed Process Control Computer System (DCS) that could work independently of Supervisory systems. The DCS had better human interface, historical data capture and replay capabilities, and powerful alarm management systems. Supervisory Systems then became a place to store large amounts of historical data and the optimization or mathematical modeling environment.

Today, the DCS works independently of any supervisory system, although some responsibility for advanced control remains delegated to more powerful supervisory systems. As new open system technologies are developed for control, it is becoming difficult to tell were the control system starts and ends and if a separate supervisory system as we know it is required anymore.

Adopting (and adapting to) new technology

The evolution of any technology has consequences for those who the technology is intended to help. The development path is rarely straight, and unexpected consequences must often be dealt with. Control systems are no exception: During the evolution of control systems, plant management and operations personnel have been required to continually reassess their abilities, progress, and needs, and adopt the technology as it becomes available. There are human and institutional consequences to this, and often the technology must be adapted to by the users as much as it is adapted for the application. The lessons learned during this process transcend the particular process and control technology involved, and provide the only hints we have available to us about what the issues of the future will be. In the sections that follow, we describe a number of examples of the sometimes unexpected consequences of the evolution of automation technology.

The ICI shrink-wrapping robot

The development of more capable computers allowed mechanical systems such as robots to be used for many highly labor-intensive and perhaps dangerous or unpleasant activities that required little in the way of human cognitive skills. Initially the robot was used for repetitive tasks that people added no value to, and for safety reasons people and robots would never work together. The robots were put into large safety cages and isolated from all contact with people. The safety concerns, although based upon legitimate concerns, often had deeper origins, as the following experience by the senior author illustrates.

> I remember my first robotics application within Imperial Chemical Industries (ICI) in the UK. The company was keen to identify and use time-saving and cost-effective automation. A group of engineers with ideas for applications of this technology was formed and several trial applications were investigated. As the

technology was tested it became obvious that there were many reasons why people should not trust this technology. In particular, the British Government had identified high risks to people from the robots themselves as they were very powerful and could seriously hurt or even kill a person them should they get in the way of the robot's movements. The initial solution was simple: Isolate the robots from people by placing them in high security prisons. Should a person need to enter the room were the robot was working elaborate isolation and safety interlock equipment was employed to secure the robots and make them safe. Unfortunately, this solution made robots expensive and no longer cost effective for a variety of applications.

My first robot was no exception to this rule. I had the great idea to use my robot as a shrink wrapping system for 300 Kg bales of staple fiber. These bales looked like very large cotton pillows, and were wrapped in polyethylene which needed to be shrunk by heating the polyethylene at very high temperatures. The existing system consisted of a very large open electric oven with a conveyor through its middle. The conveyor carried the bale from the wrapper to the oven and then to labeling and the warehouse for transportation to customer sites. The electric oven was very inefficient (large amounts of heat escaped out the open ends) and was costing millions of dollars in energy usage; further, the wrapping was poor due to the large side ears that resulted in the wrapping and shrinking process.

My idea was to train a robot to hold a hot air gun and effectively shrink the bale by blowing hot air in a motion similar to spraying a car. I had witnessed robots performing this activity very effectively in the automobile industry. The first step was to develop a gun suitable for this activity and a local company soon came up with an effective prototype which we could use with the robot. By chance, another part of the company had been trying to use robots in the manufacturing of explosives and after an unfortunate accident the explosive was re-classified and no electrical equipment was allowed during the manufacture of the product. This allowed me to acquire the perfect robot for my application, and one that had already been paid for by my company. With minimal negotiating the robot was transferred from Scotland to England and was on its way to save the company a lot of money and produce a better package for our customers. What could go wrong?

The robot arrived in bits but with the aid of some untrained personnel and in true engineering tradition we figured out how to re-assemble the robot. Soon the basic computer was connected and, after a little experimentation, the robot was programmed in assembly language and soon responded to my every request. The sight of such new and exciting technology set the political wheels in motion. The craftsmen who would have to maintain it saw it as a large threat; the operators who were to interface with it could not make their minds up, and senior management needed to be convinced. So my boss decided that we should gather all the interested parties together and demonstrate of the capabilities of this new technology.

Given 48 hours, without all of the equipment for the final application and no time to do a correct installation in any event, we needed a demonstration task. We came up with the idea to put a box in front of the robot and write the name of a the senior executive who would be present to demonstrate the flexibility and precision of the equipment. Programming was easy and soon we had a working demo and we had confidence we could win everyone over to this new idea and effectively use this technology. At the end of the demo I was to give a short presentation on the real application and the benefits and savings.

On the day of the demo one of the mechanical craftsmen noticed a small hydraulic oil leak and without asking anyone he decided to just go ahead and fix it as he would any other piece of plant equipment. However, the very small leak was not fixed by a couple of turns of his wrench, so he took the pipe off and went to find new seals for the connection. He was not aware that the piping was designed for very high oil pressure and required that specific maintenance procedures be followed. A simple oversight?

After the audience assembled, and after a few brief words, the demo was started and to my horror the robot went crazy: Instead of writing the executive's name on the carton it used the felt pen as a weapon and destroyed the carton with several lethal blows. Then, with one effortless blow the robot knocked the three-foot-square carton towards the amazed audience and then repeatedly drove the pen into the ground as if it were a bayonet into a dummy until all semblance of the pen was totally gone.

By which time I was able to make a safe path to the emergency button on the console.

Without any opportunity for discussion, I was told to get rid of the machine, and to never use the word "robot" in any discussion with anyone ever again.

Lessons

There are lessons to be learned from this story on multiple levels.

Configuration management as root cause

Traditionally speaking, the root cause of this incident was inadequate configuration management. In fact, there were a host of organizational culture issues that needed to be resolved, involving work permitting, skill certification, project planning, and safety.

This was a large set back for technology at ICI, and from the experience the senior author learned to never do anything by half measures: Every piece of equipment introduced should have the protection of competent people working on it, professional installation, clear operating procedures, and management of change control. If we had had the system identified as a plant asset, it would have been put on the maintenance system, and the craftsman would not have worked on the robot without a work-order and a permit. The work would have been reviewed and the correct replacement parts would have been identified from the manufacturers recommendations. As we assessed the incident we discovered that the craftsman was also in a dangerous situation during his maintenance activity — he was not aware of the high pressures involved; the system had not been isolated correctly, and he could have been seriously injured. The only isolation was that the computer program was not running and traditional electrical and hydraulic isolation which was not done would probably not have been sufficient based on the capability of this beast.

Why did this one event temporarily stop progress in the application of automation?

This one incident brought the introduction of automation into this company to a screeching halt. It had a disproportionate impact across the whole workforce because it reinforced preexisting doubts, fears, and insecurities, and it confirmed negative beliefs not only about robots but about technology in general. Every robot became a bad thing despite success in other parts of the company. People wanted to believe that the technology was unnecessary and dangerous and the old way of doing things was the best. After all it had been that way for hundreds of years why change? This incident confirmed their suspicions.

The demonstration was the wrong thing to do. The engineers involved wanted to demonstrate impressive technology, but the audience wanted a demonstration of safety and job security. The engineers thought that the goals being demonstrated addressed important underlying values, in this case efficiency and cost savings. However, the audience was interested in a different set of values: The impact on jobs, the change in responsibilities and training required to be competent on the new technology , and the impact of the new technology on the value of the existing workforce. The engineers involved still truly believe that this was the right solution, and that it could have saved large amounts of money and been very efficient and produced a better product. The organization is still trying to achieve other objectives in addition to efficiency.

ICI Robot Automated Guided Vehicles

Some years after having survived the debacle of the first demonstration of robot technology, the senior author discovered another problem that could be solved with automation, and has hope for the redemption of robot technology:

> Well after the demonstration of the first robot , another opportunity arose. The same plant was having difficulties transferring the same bales of fiber from the baling machines to the wrapper and to two different storage locations. The existing transportation was via computerized hydraulic hoist. The hydraulics were worn out and the leaking oil contaminated the bales. More importantly, the hoist periodically dropped the bales on the floor and an operator would have to drive the hoist under manual control and pick the bales of the floor and transport them to the wrapper. In the process, the unique identification of the bale would be lost and the automatic labeling system would get out of step and put wrong labels on the bales. This often would cause a major customer relations problem because large batches of clothes would be ruined, resulting in hundreds of thousands of dollars in losses.

> Conveyors didn't work very well with this product because fibers would get into the bearings and seize the conveyor in a short time. The hoist system's inherent problems with hydraulic systems and reliability made it a poor choice for replacement.

> The best solution from an engineering point of view was a robot truck, called an Automatic Guided Vehicle (AGV). AGVs are used in many warehouse applications and can be designed especially for carrying problem loads. An additional advantage of the AGV in this application was that it could maintain the orientation of the bale out of the box of the baler guaranteeing perfect wrapping and perfect orientation for autoamtic labeling every time.

> All we had to do was write a capital expenditure request to get some money and do it, except for the one minor problem that I was not allowed to use the "R" word wiithin 20 miles of my management team!

Past lessons applied

The senior author, presented with an uncommon opportunity to push a favorite but disfavored technology using learnings from his past attempts to overcome the nontechnical barriers to the effort, takes up the quest:

> Well this time before I mentioned "robots" I meet with the Unions and the operations folks and identified and confirmed the existing problems and issues. We all agreed that we could not accept the current way of working and we needed a better solution. And, in an almost fully automated plant, the most unpleasant place remaining requiring the operators to interact with the process was the bale handling and labeling area.

During the discussions we reviewed all the options and identified conveyor transportation issues but agreed we would investigate and get quotations for the best available conveyors. After the initial investigation we arranged for sample conveyors to be tested and, as in the past the loose fiber found its way into the works and the conveyors stopped working. The union and the operators, concerned that we were not keeping up to date with progress, demanded that we find other alternatives—surely technology has progressed so that we could move a 300 kg bale of staple fiber a couple of hundred yards and maintain orientation for wrapping and labeling? So my management demanded we come up with other alternatives to the equipment we had so effectively ruled out.

Well this was the opportunity I had been waiting for to introduce the "R" word, but there were cultural issues still outstanding from the first failure. So this time I worked with the manufacturers to identify all the issues and to demonstrate what other industries were doing about them. With the aid of films the team was able to review what other folks do. We showed a frozen food warehouse with extremely cold areas where people did not want to work and how AGVs were being used in that environment. We showed how AGVs interface with and behave around people. We showed how full size automatic forklifts were handling difficult products and how reliable and accurate they were. We demonstrated how they could maintain bale orientation because they followed guided wires buried in the floor. We also presented an estimate of the proposed system which was significantly less than the proposed conveyor system. Finally, we described how we could maintain the trucks by working with the existing maintenance group.

It was the Union that recommended that management make a case to justify this work. The union would support the successful implementation of this system as it was the only sound alternative to a major problem that was now a bottleneck to the whole business. The management decision was not difficult; the whole work force was united behind the AGV solution.

The fully computer controlled system was installed and worked from day one without any problems. It was interesting that the operators took pride in the AGVs from the early days and demonstrated their acceptance by naming them and creating faces on the glass shields of the fork. Each truck had its own personality which the workers identified by watching the way they worked. The final acceptance was expressed in the form of a cartoon where the AGVs were depicted as a team members taking a break and playing dominos, and the plant manager pulling his hair out as he had recently had a campaign to keep people on the job and out of the mess room.

Initial success enables more

The initial success of the AGV system led to extensions. The robot truck system was later fully controlled by a supervisory system which tracked bales and handled grade changes and handover to the warehouse for putting into campaigns for customers delivery. The supervisory control computer was an important part of the system but most people probably did even realize that it existed. The only time people interfaced to the system was so request information about the product and where it was in the plant or to call a truck to do something unusual such as store and retrieve bales with a unique identification during wrapper maintenance or other outages.

What culture lessons were learned

It is not very often that case studies on technology implementation can be found with so much in common: The automation attempts were made in the same company, the same plant, and even in the same area. The cultural issues to be solved were the same. The engineering group was the same. What was different was the approach: Technology was introduced as a solution to a problem that people were having. The solution to the problem was consistent with the values of

the people involved. Concerns based upon misimpressions were allayed through early and thorough exposure to the problem and the possible solutions.

Unlike engineers, most people are not interested in technology per se—only how it impacts their way of life. They are more interested in what they will see and how they will deal with problems and failures. Modern industrial plants are introducing more and more technology in every operation and people are often stretched and usually not considered partners or co-workers. The AGV system demonstrated how people and equipment could work in harmony and be one team. Preparation for change is important and winning sponsors is extremely important.

People do not like to face the unknown they like to see well managed and professionally implemented projects. Having a supervisory system was not important—it was the functionality and how people identified and worked with it that mattered.

Often in automation projects, the need to communicate with the affected users is identified, but as an end in itself instead of as a means to an end. Involving people in automation changes is the only way to understand all of the obstacles to the success of the technology, most of which are not technical in nature.

Formentor

Formentor is a research program conducted in Europe with funding by ESPRIT. Participants include by Cap Gemini Sogeti, the Joint Research Centre of the Commission of the European Communities and Aerospatiale Protection Systems. The goal of this team was to develop a risk management solution for complex systems, in the form of an "intelligent watchdog". Formentor is a decision support system that provides the process operator in charge of process supervision with:

- monitoring,
- situation assessment,
- diagnosis support,
- reactive planning, and
- prediction capabilities

Formentor was intended as an added-value system to the existing control systems. It was to provide a global and permanent overview of the process state, and help react to deteriorated or disturbed operation. It was to allow the anticipation of severe process states by detection of precursor signs and malfunctions, and measure the impact of actions on the process behavior, thus limiting plant shutdowns and improving quality of the products.

The Formentor team have completed two industrial prototypes: one at the BP Grangemouth chemical plant in Scotland, and the other at the Total Refinery in LeHavre, France. Both of the systems have met many of their technical goals.

The impact of the technology life cycle

Both of the prototype applications have been abandoned by their industrial sponsors due to the high cost and required knowledge skill set to maintain the system, and the lack of benefit achieved during actual use.

Formentor systems need to reflect the actual plant in order to provide technically accurate information. The system is initially accurate, but plants are always being maintained, upgraded, or expanded, and small changes in the plant equipment or the process operation that occur over time require specialist computer experts to make corresponding changes to Formentor. This turned out to require resources beyond the ability of the plants to provide.

The model-based approach used by Formentor (and some other efforts with similar objectives) is technically very promising. However, the same monolithic approach that enables good diagnoses of problems limits the accuracy if the models are not accurate. Thus the approach may require more maintenance effort in practice than can be expected.

Multiple applications for each problem

Another issue with monolithic approaches is that not all problems are best diagnosed using the same approaches. For subtle imbalances in processes, the performance of good models is hard to beat. However, other kinds of problems may be better diagnosed with empirical (e.g., statistical or neural network), knowledge-based, qualitative, or fuzzy techniques. Aside from diagnostic accuracy, an advantage to using multiple techniques is that the installed system can be developed incrementally—the user gets benefits from each new diagnostic capability that is added—which helps get over barriers raised by the large initial cost to install a comprehensive system.

The value of current practice

Had Formentor provided significant benefit in actual operations, maintenance costs might not have been a significant issue, but it turns out that Formentor was not used during process upsets as much as it could have been. This was due to the fact that the Formentor display was not incorporated within the system used by operators, but on a separate, independent system. In the event of an upset, this required operators to leave their primary view of the process—and their only means of process control. History has repeatedly demonstrated that operators will not do this regardless of the extra information that may be available. The robot examples showed the need to take into consideration the users' practice of values, but we also need to take into consideration the value of practice.

For a diagnostic or decision support tool to work in the real world, it must seamlessly fit into the way people work with the existing control system. More generally, for any improvement to an existing system to succeed, it must accommodate the ways users work with the existing system.

The Advanced Automation System for air traffic control

Beginning in 1984, the FAA attempted to replace large parts of its obsolescent air traffic control system with better technology. After a rigorous design competition, a $2.5 Billion contract was issued in 1988 calling for initial deliveries in 1992. The schedule quickly slipped and costs escalated. At the peak of the program in 1992, more than 1000 software engineers were working on the system, and the estimated cost to complete the system had risen to $7.6 Billion. With no end in sight, most of the program was canceled in 1994.

Lessons revisited

Every one of the issues raised in our examples so far played a role here as well. There was inadequate attention to configuration management. There was insufficient attention paid to the culture of air traffic controllers and to their operations practices [In fact, at least some informed observers lay the ultimate failure of the program to intractable problems associated with automating the racks of flight progress strips that controllers physically manipulate to help them stay oriented to the progress of flights in their airspace.] Inadequate attention was paid to lifecycle costs, and the solution being built was too monolithic [Calling for the consolidation of 200 air traffic control centers into 20, for example].

The system also serves as an example of two new issues that frequently arise in the evolution of complex systems, which will be encountered more and more frequently, and which may represent the largest challenge in the future: The program attempted to apply the wrong kind of technology, and it did not address the right problem, in the first place.

New technology for new solutions?

In 1984, the Advanced Automation System attempted to modernize air traffic control through the development of a system using newly developed parallel-processing UNIX workstations running hundreds of thousands of lines of Ada code and with a seven nines availability requirement (allowable downtime: 3 seconds per year). To the best of our knowledge, none of those elements have yet been commercially demonstrated, five years after the system was supposed to have been fielded. This was not a system depending on well tested off the shelf components!

Conversely, petrochemical companies, manufacturing industries, and even NASA often have the opposite problem: It is not uncommon to find key control software running on the equivalent of IBM XT technology—if it is computerized at all. What is lost in capability is more than made up for in reliability: A desktop PC might crash once or twice a week (or even once or twice a day); a control PC is not expected to crash during years of continuous operation. In these industries, even upgraded technology that seems old by desktop standards.

The challenge to those introducing automation is to select technology that is new enough to be useful, but old enough for the risks be understood.

What is the problem here?

The FAA attempted to better automate the existing air traffic control system, and in particular to make it better at doing the things it already did. Thus, designers went after throughputs, and planes per controller, and other measurements of the system currently in use.

Fundamental goals—the reason the system exists in the first place—were given short shrift. Thus, the problem left unaddressed was something like, "How can we have more planes in the same airspace with more safety and less cost?".

It would be unfair to hold up this particular project to ridicule as a particularly bad example, because the issue is endemic, and very difficult to overcome. Most of engineering is incremental— we attempt to do the same things, only better. The engineers at ICI did not think much if at all about whether there was a better way to move fiber to customers than in bundled and labeled bales; they instead tried to bundle, label, and move bales better.

The real advances—the breakthroughs in the application of technology—come not from solving the same problems in better ways, but from solving new problems in new ways. All of us, even the most inventive, have to watch out for this one. [To take just one example, the Wright Brothers spent years trying to improve the aerodynamics of their front-mounted horizontal stabilizer, even after their competitors had put it in back.]

At this point in the history of control systems, it may be time to look away from efforts to gain another .5% of efficiency, and instead focus on something that can give us 10%.

Integration Support for Large JV ATP Projects: Why? How? So What?

Van Parunak, Ray VanderBok, Mark Brown
Industrial Technology Institute
Ann Arbor, Michigan

Abstract

The Near-Zero Stamping and Intelligent Resistance Welding ATP projects under the AutoBody Consortium have established a Systems Integration and Multidisciplinary Planning Team to help project participants see the forest for the trees, in other words to be able to form their project activities within the framework of a deliverable and within the context of the program's efforts as a whole. This paper will outline the requirements that this team's activities address, describe the techniques that have been applied to meet those requirements, and reflect on the successes and lessons for future JV's, which can be drawn from these experiences.

Special Requirements of Consortium Projects

Team-based projects can be very effective. They allow a wide variety of participation; which brings a number of disciplines together to solve a problem. Each member contributes from their own experience and expertise. In this way, multiple aspects of a problem can be solved concurrently. Team projects often facilitate deployment of the solutions. When the team includes end-users and vendors, the team itself becomes the vehicle for deployment of results.

For many projects, the team includes participants from multiple organizations. Team projects enable people who do not normally work together to do so. Each organization contributes staff to the project. A consortium is formed that is a new organization designed to work on a specific issue. The consortium project is organized with recognized leaders and roles for participants.

The benefit of a consortium project does not come without cost. Members of the consortium will not initially know the other members and their capabilities. Where a close-knit project team that has worked together for years may be able to respond to a problem like a disciplined pit-crew, a new consortium team will spend time to develop a common understanding of a problem, make assignments, make mistakes, and complete the work. Consortium projects require additional time and support beyond traditional projects to facilitate this common understanding.

The additional support required for consortium projects is very practical:
- With whom am I workingand how do I communicate with them?
- How do administrative functions (like scheduling a meeting) work within the consortium?
- Where do I put things we are working on, so others can get at them?

As a manager of a consortium project, there are additional needs, including building support and commitment among the participants, tracking progress, and capturing results. Running a consortium project is often complicated by the "on-loan" status of the individual participants. A consortium staff member is still an employee of his or her own company. As such, they have to balance consortium work with responsibilities "back at the office."

The issues raised above exist for traditional consortia and joint ventures. In general, the need to respond explicitly to the special needs of the consortium increases as the number of participants increases. Addressing the needs of the consortium will increase its effectiveness. In this way the consortium can focus on reaching the goals and the solving the problems that brought it together, and a support function can work to make the consortium operate effectively.

The Intelligent Resistance Welding (IRW) and Near Zero Stamping (NZS) joint venture ATP programs managed under the Auto Body Consortium (ABC) have a Systems Integration and Multidisciplinary Planning project that specifically supports the working of the consortium. ABC is leveraging the lessons learned and integration tools developed under these projects in its ongoing initiatives, including the Advanced Chassis Structures (ACS) ATP proposal.

Techniques and Solutions

The Systems Integration and Multidisciplinary Planning projects of IRW and NZS are led by the Industrial Technology Institute (ITI) and staffed additionally by Rockwell Automation, General Motors, and Autodesk. This Integration Team is involved across NZS, IRW, and the proposed ACS ATP programs. The objective of the Integration Team is to make the research teams more effective by improving communications within and between research teams, sharing common tools and solutions, and facilitating deployment of the results.

Communications within and between Research Teams

The Integration Team has built a set of communications services for the consortium based on Internet technology. Its first activity was to assist companies and individuals to connect to the Internet. This activity provides participants the basic capabilities of email and access to the World Wide Web so that they can access the Internet-based services provided to the consortium.

Because of the proprietary nature of program activities, a secure Web server was configured to support program services. This secure server provides user identification and password protection for access and encryption of communications with participants. In addition, system operators monitor use-logs to identify unauthorized access attempts.

To facilitate interoperability among participants, the Integration Team worked with members to standardize a common set and version of computer applications. These common applications include MS Word 6.0, MS Excel 4.0, MS Project 4.0, MS PowerPoint 4.0, Netscape 2.0, and AutoDesk Mechanical Desktop Release 12. In addition, the Integration Team has developed a suite of custom tools where no commercial tools were appropriate. The general strategy is to develop tools quickly, see how useful they were in practice, and further develop them as necessary.

Internet Connections

The communications strategy is based on providing access to a set of services over the Internet, while continuing to support the more commonplace means of facsimile. It is important to get everyone connected to the Internet. Use of the Internet is growing because of its own intrinsic value. The consortium leverages this change in business communications. Not surprisingly, there is wide variance in the level of connection and familiarization with the technology.

Some users have to start with Internet connections at home because their organization is not ready for Internet-connected computers in the workplace. Some users have stand-alone computers with modem connections to Internet Service Providers, a low-cost but effective way to provide Internet connections for project members. The best connections are those where the Internet connection is built into the company's network, so Internet connections are broadly available within the company and use of the Internet is a common part of doing business.

Each organization is unique, so the Integration Team provides individual support to assist participants to establish Internet connections. Often this involves meeting with the computer support staff at the user's site.

Some common needs can be solved collectively. For example, the Integration Team provides hands-on training sessions for users who have no previous experience with using the Web.

Email

Once connected, participants can send email to each other. Email supports communications that are more complex than voice mail. The Integration Team maintains many project email distribution lists making it easy to send email to all participants, all leaders, or all the members of a project team.

Internet Discussion Groups

Participants who are familiar with the Internet have asked for threaded discussion groups. Much like Internet News Groups, this discussion area will allow participants to post messages and respond. This is a good way to open topics for wide-range input from the program participants. Because the messages are linked together by topic, it is easy for members to understand the history of the discussion. To ensure the security of the program information, the discussion group is configured under the program's secure Web service.

Program Status Synopsis

The Program Status Synopsis provides a short description of the current state of the projects, so that participants can track progress in tasks other than those in which they are involved in detail.

People Finder

The People Finder is a WWW-based service to find people related to projects, tasks, or companies. The People Finder provides a searchable index database over a collection of profiles, each with a person's name, address, phone, email, photograph, work experience, and task roles, thus helping people identify those consortium participants with the required skills or company contacts.
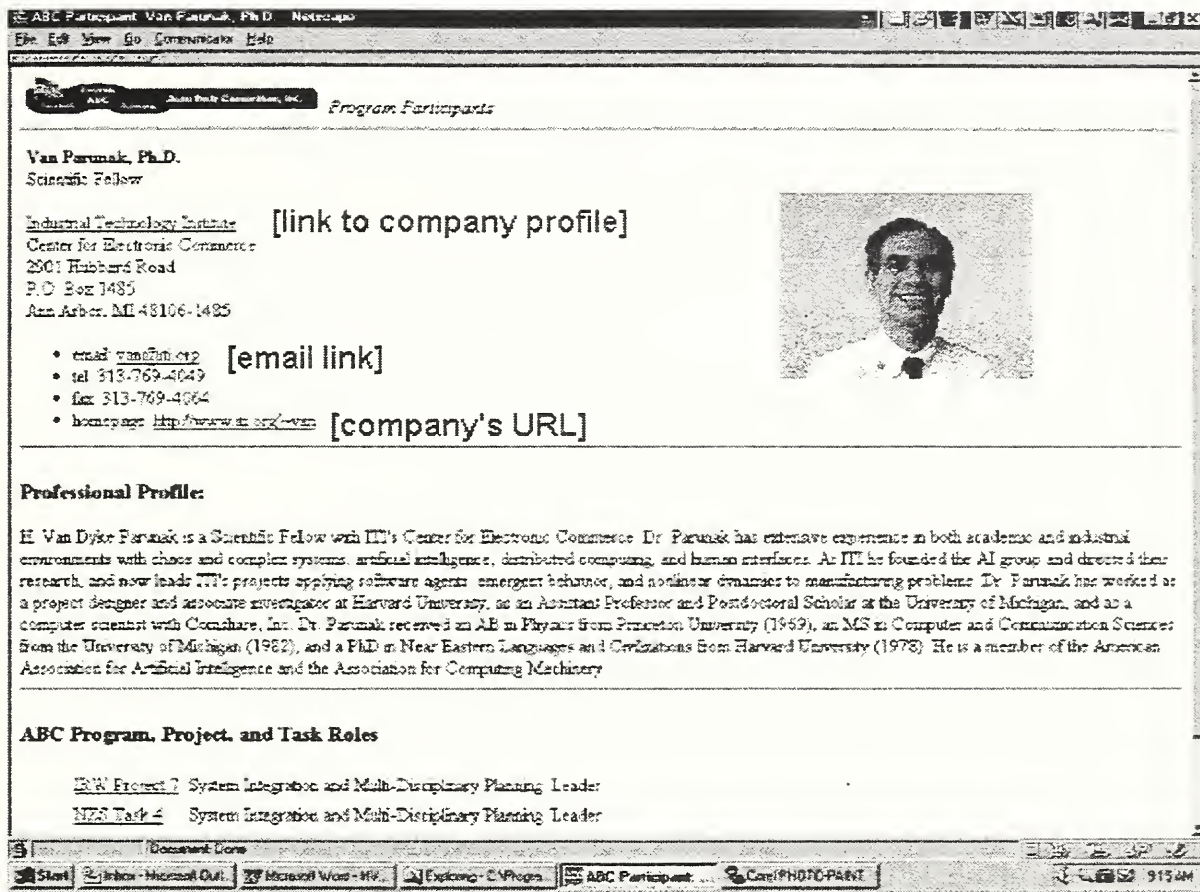
An Example of a Participant's Profile Page from the People Finder

File Download Area

The file download area provides a central location where reports, document templates, and procedures can be kept. Participants can be assured of getting the most current version of these documents. Often these files can be viewed on the Web, in addition to being available for download in their native format.

Calendar

The Program Calendar is a WWW based service that describes events for the program. A simple chronologically ordered list introduces the events. This list is a series of links to meeting notices. Each meeting notice includes a map to the location, contact information, and the meeting agenda.

When a meeting is completed, the minutes are posted on the same list as the meeting itself. At times attachments to the meeting such as a to-do list are also posted within the list of meetings. The Integration Team converts the documents (notice, minutes, handouts) to html format so they can be viewed directly from the Web.

Leadership
ABC
Teamwork
Technology
Auto Body Consortium, Inc.

Meeting Notice

Development & Testing of a Communicator Loop
(Communication between weld controller & machine)

Convener: W. Patel

Contact Person: W. Patel
tel. 810-497-6518
fax 810-497-6302
E-mail wpatel@lambtech.com **[email link to RSVP]**

Date & Time: Thursday July 17, 1997
3:30pm - 4:30pm
Following Quarterly Review Meeting

Location: Lamb Technicon
5663 E Nine Mile Road
Warren, MI
Map To Lamb Technicon Engineering Ctr **[link to print map]**

Required Attendees:

Dan Carnahan, Allen-Bradley **[agenda follows Required Attendees]**

An Example of a Meeting Notice

Glossary of Terms
The Glossary of Terms provides the program participants with a common vocabulary. The Glossary provides a clear definition of the program's "official" term and links to other common terms. Links are also provided for the source of the term and related terms or graphics.

Links to Related Sites
The program Web site provides a set of links to related project sites. These include government programs, trade associations, and vendor product information.

Search the Site
A search service is provided to search for text in any document on the Web site.

FTP
An FTP area allows participants to place large files in an area where others can access them. The IRW project 2 group at Battelle is using this FTP site to pass large FEA results to IRW project 3 at UM. The data is available for a week on the FTP site, then it is archived to CD-ROM. If an archived data set is needed, it can be placed back on the FTP site for general project access.

Sharing Common Tools and Solutions

Common solutions are needed to support collaborative research. With the number of people and organizations involved, it is likely that duplication of effort would occur. The Integration Team works to eliminate this possibility by developing common solutions to meet the needs of multiple researchers.

Common solutions and information strategies ensure that data can effectively pass between systems. For example, the IRW program developed a common ASCII data format that can be used to pass data feature values between project teams. This data format is general enough to support plant floor destructive test data collection, dynamic test data, and FEA test results. As a result, only a single interface was required to interface this data into the experimental test database.

Common Solutions

Many of the research activities involve experimental testing. The research teams need access to the status and results of the experiments. These experiments can be physical experiments carried out in a laboratory or virtual experiments carried out in a computer. A relational database was developed to track the experiment process. The database contains the definition of the experiment, data regarding the test conditions, data regarding the test material, and pointers to resulting data sets. The database ties information from several sources into a common tool to facilitate researchers access to experimental test results. In the future, researches will be able to check the status of a test via the Web.

IRW Experiment Information Flow
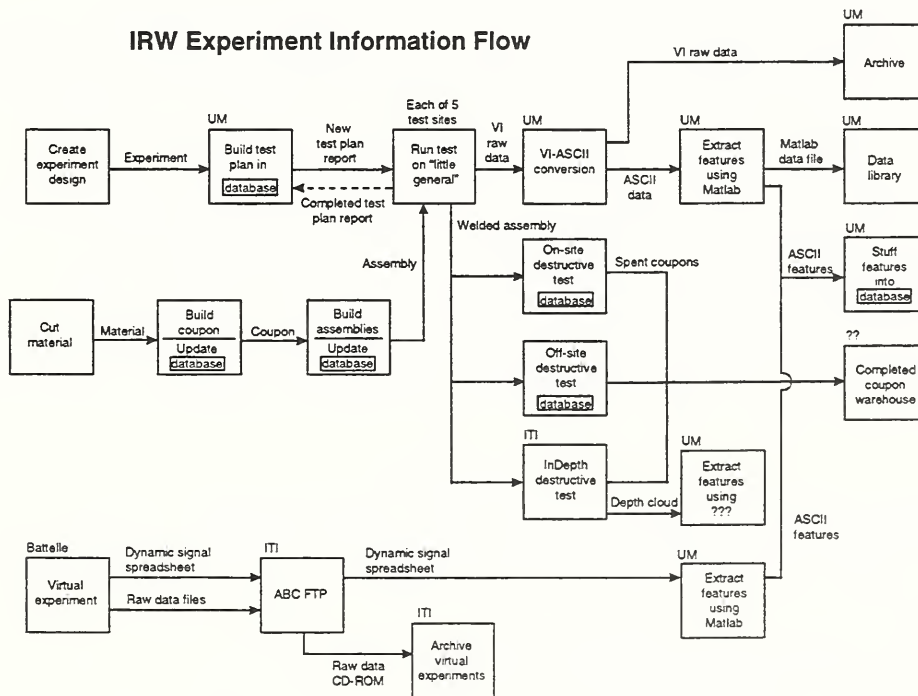
As part of their participation in the Integration Team, General Motors provided access to their MARS dimensional analysis software for NZS and IRW participants. The MARS system is designed to collect and report dimensional quality information. The software supports statistical analysis of data from coordinate-measuring machines and gauges of many different brands and types. The collected data can be compared to CAD data. Because the software was built on a common Microsoft Access database, the system can be used to collect data from various sources and present the data in a common format for researcher's custom analysis. MARS is planned for use in NZS for designed experiments regarding critical process parameters.

The Internet is also becoming a common solution for deployment of research results. Some research takes the form of software that is used to support the manufacturing process. By building software to be accessed over the Internet, it can be easily updated and evaluated by many members. The Web browser provides a platform-independent method to present text, graphics, and linked documents for use by researchers and end-users. The IRW is building a resistance welding diagnostic tool for use by plant engineers, which is accessible on the Web and on CD-ROM.

Solutions that Support Deployment

In addition to challenging research problems, ATP programs must ensure that the results are deployed such that the benefit is realized. It is easy for researchers to focus on the technical challenges of the research. A valuable function for the Integration Team is to facilitate deployment of results. The Integration Team reminds the researchers of the need for deployment and the realities of manufacturing businesses.

Manufacturing engineers understand how a new technology impacts many aspects of the manufacturing operations and set requirements on technology to ensure that it integrates with other parts of the plant floor. The Integration Team has identified the following issues as requirements to the researchers to facilitate deployment of the research results:

- The research must be deployed on a platform consistent with near future plans. It is not realistic to expect a UNIX workstation to be used at a welding re-work station.
- The deployment must be consistent with workers' jobs on the plant floor. The use requirements need to match skill levels and work loads.
- The resulting technology is expected to work reliably. The manufacturing plant floor is a harsh environment. Shop floor workers have high expectations and a low tolerance for failure.

The Integration Team works provides researchers with platforms and requirements so they will be successful. For example; Allen-Bradley provides new machine controllers for use in NZS research prototypes.

Prior to deployment, managers must be able to understand the research results and the benefit to their organization. The Integration Team will work with the researchers and engineers in the project teams to identify and describe specific technologies and results. These descriptions will be used to introduce the new solutions to managers. They will capture the tools and manufacturing methods that will result from these ATP programs.

Lessons for the Future

The work of the Integration Team over the past two years has yielded some lessons for current and future programs. The integration effort has required significant resources and still remains an inexact science, but we are learning.

Access to the Internet

Good Internet connectivity is extremely important. Increasingly, the Web is being used to help people work together, manage the project, and deploy the results. Getting good Internet access for participants is a time-consuming effort. Interestingly, some of the most intractable problems come from the security policies of the largest companies.

Use of the Internet to support the program depends on the ability to access quality content. The more reasons a person has to visit the site, the more of the site will be used. It requires dedicated operations staff to maintain and build the content of the program Web sites.

Distance Meetings

The NZS and IRW programs have participants spread across the Great Lakes states. The time and cost associated with travel to meetings represents expending resources on non–value-added activities and limiting team interactions. The Integration Team is planning to utilize communications technology to facilitate meetings spread across distant locations. In addition to common telephone conference calls, video conferencing and Internet collaboration software will be used.

Prior to use we will determine how these methods can be successfully used. For example, we may find that a video conference is a good way to present results, but a bad way to aggregate ideas for project planning.

Project Management Support

The Integration Team has been careful not to overlap with the role of program management. We have found, however, that we can provide services to support program management. For example; the IRW program has requested that the Integration Team provide a common system for tracking action items across the project teams. In addition, systems are planned to capture resource forecasts (company staff by month and activity).

Summary

Collaborative pre-competitive R&D is inherently difficult, due to the centrifugal forces of business competition, disciplinary specialization, and the cultural divide between the research world and industry. ATP Joint Venture projects provide an important venue for bringing together participants whom these forces would otherwise keep apart, and experience shows that addressing these forces explicitly can increase the productivity and effectiveness of the ATP opportunity. The techniques described in this paper have proven effective in the Auto Body Consortium's projects under the MVMT ATP program, and are being refined and extended to support ongoing and future initiatives.