Title	ICMA 2002: proceedings of the International Conference on Manufacturing Automation : Rapid Response Solutions to Product Development, December 10-12, 2002, Hong Kong, China
Other Contributor(s)	Hua zhong gong xue yuan. School of Mechanical Engineering.; National University of Singapore. Dept. of Mechanical Engineering.; University of Hong Kong. Dept. of Mechanical Engineering.
Author(s)	Chen, Ye-Hwa; Gibson, I. (Ian); Tan, S. T.
Citation	
Issued Date	2003
URL	http://hdl.handle.net/10722/54871
Rights	Creative Commons: Attribution 3.0 Hong Kong License

INTERNATIONAL CONFERENCE ON MANUFACTURING AUTOMATION

Rapid Response Solutions to Product Development

ICMA 2002

Edited by: S T Tan, I Gibson, and Y H Chen

International Conference on Manufacturing Automation

Organizing Committee

Co-chairmen

A Y C Nee (NUS)   S T Tan (HKU)
Y L Xiong   (HUST)

Program Chairmen

Y H Chen (HKU)   J Y H Fuh (NUS)
I Gibson (HKU)   P G Li (HUST)

Members

K W Chan (HKU)   K Z Chen (HKU)
S H Choi (HKU)   H Ding (HUST)
H J Gan (HUST)   G Q Huang (HKU)
S Q Li (HUST)   H T Loh (NUS)
J Qin (HUST)   W S Sze (HKU)
W Wang (HKU)   T N Wong (HKU)
Y S Wong (NUS)   S N Yang (HUST)
J C Yuan   (HUST)

International Scientific Committee

D Bennett, UK   H Meerkamm, Germany
A Bernard, France   B Mieritz, Denmark
D Bourell, USA   J Mo, Australia
A Bowyer, UK   K Pawar, UK
R L Brown, USA   D T Pham, UK
L M Camarinha-Matos, Portugal   M Pratt, UK
R I Campbell, UK   A Valenzano, Italy
J S Chen, Taiwan   J S M Vergeest, The Netherlands
Z Dong, Canada   J S Wang, China
N N Z Gindy, UK   X K Wang, China
P H Gu, Canada   T Wohlers, USA
I Horvath, The Netherlands   J R Woodwark, UK
S Kumara, USA   C Wu, China
A Kusiak, USA   Y B Xie, China
J Lee, USA   J Q Yan, China
K Lee, Korea   Y Yan, China
S Y Lee, Taiwan   M Yeung, Canada
Y T Lee, Singapore   M F Yuen, HK
Y Z Lei, China   H C Zhang, USA
G Lin, Australia   S Zhang, China
F Liu, China   J Zhou, China
R Martin, UK

ICMA 2002

Proceedings of the International Conference on
Manufacturing Automation
Rapid Response Solutions to Product Development

December 10–12, 2002

Hong Kong, China

Edited by

Professor S T Tan, Dr I Gibson, and **Dr Y H Chen**

Organized by

Department of Mechanical Engineering
The University of Hong Kong (HKU), China

School of Mechanical Engineering
Huazhong University of Science and Technology (HUST), China

Department of Mechanical Engineering
National University of Singapore (NUS), Singapore

Sponsored by

K C Wong Education Foundation

William M W Mong Engineering Research Fund, Faculty of Engineering,
The University of Hong Kong, China

National Natural Science Foundation of China (NSFC), China

Supported by

The Global Alliance of Rapid Prototyping Associations (GARPA)

The Hong Kong Institution of Engineers, Manufacturing and Industrial Engineering
Division, China

Related Titles of Interest

Title	Editor/Author	ISBN
Advances in Manufacturing Technology XVI (NCMR)	K Cheng	1 86058 378 4
Advances in Manufacturing Technology XV (NCMR)	D T Pham, S S Dimov, and V O'Hagan	1 86058 325 3
Computer-aided Production Engineering (CAPE 2001)	H Bin	1 86058 367 9
Design and Manufacture for Sustainable Development	B Hon	1 86058 396 2
IMechE Engineers' Data Book – Second Edition	C Matthews	1 86058 248 6
Managing Engineering Knowledge MOKA: Methodology for Knowledge Based Engineering Applications	M Stokes	1 86058 295 8
Rapid Prototyping Casebook	J A McDonald, C J Ryall, and D I Wimpenny	1 86058 076 9
Rapid Prototyping, Rapid Tooling, and Rapid Manufacturing	A Rennie, C Bocking, and D M Jacobson	1 86058 374 1
Software Solutions for Rapid Prototyping	I Gibson	1 86058 360 1
Virtual Design and Manufacture	IMechE Seminar	1 86058 314 8

For the full range of titles published by Professional Engineering Publishing contact:

Marketing Department
Professional Engineering Publishing Limited
Northgate Avenue
Bury St Edmunds
Suffolk
IP32 6BW
UK

Tel: +44 (0)1284 724384
Fax: +44 (0)1284 718692
e-mail: orders@pepublishing.com
Website: www.pepublishing.com

Keynote Speaker Biographies

Professor Nam Pyo Suh

Professor Nam P Suh is the Ralph E & Eloise F Cross Professor and Director of the Manufacturing Institute at MIT. He served as Head of the Department of Mechanical Engineering at MIT for ten years, from 1991 to 2001. In October 1984, Professor Suh accepted a Presidential Appointment (by Ronald Reagan) at the National Science Foundation where he was in charge of engineering.

His contributions to the field of tribology include the delamination theory of wear, the solution wear theory, a theory on the genesis of friction, coated cutting tools, and the use of undulated surfaces to lower friction and wear. His paper on the delamination theory of wear was chosen as the citation classic by the Institute of Scientific Information (ISI). In the field of design, he developed the subject of axiomatic design. In the field of polymer processing, he invented many important processes and devices, including microcellular plastics (commercially known as MuCell) and many others. In metal processing, he is the inventor of a new metal processing technique called the Mixalloy Process.

He is the author of more than 280 papers and five books, holds about fifty patents, and edited several books. Among the books he has authored are *Elements of the Mechanical Behavior of Solids* (with A P L Turner), *Tribophysics, The Principles of Design, The Delamination Theory of Wear*, and *Axiomatic Design: Advances and Applications*. Professor Suh is a Series Editor for the Advanced Manufacturing Series and the MIT-Pappalardo Series of Oxford University Press. He was also the Founding Co-Editor-in-Chief of the International Journal, *Robotics and Computer-Integrated Manufacturing*, and also serves on many editorial boards.

Professor Suh has received three honorary doctoral degrees: Doctor of Humane Letters from the Univ. of Massachusetts-Lowell in 1988, Doctor of Engineering from Worcester Polytechnic Inst. in 1986, and Honorary Doctor (Tekn. Hedersdoktor) from Royal Inst. of Technology (KTH), Stockholm, Sweden, in 2000. He also received the Gustus L Larson Memorial Award, the Blackall Award, the Best Tribology Paper Award, and the William T Ennor Manufacturing Technology Award from ASME; the F W Taylor Research Award of SME; an SPE Best Paper Award; Federal (NSF) Engineer of the Year Award from NSPE; and the American Society for Engineering Education Centennial Medallion. He was also awarded the NSF's Distinguished Service Award. In 1994, he was awarded the KBS Korean Compatriot Award for Scholarly Achievements. He is also the winner of the 1997 Ho-Am Prize for Engineering. In 2000, he was the recipient of the Mensforth International Gold Medal of the Institution of Electrical Engineers of the United Kingdom. In 2001, he received the Hills Millennium Award from the Institution of Engineering Designers of the United Kingdom.

He is a Fellow of ASME and a member of ASEE, SPE, and AAAS. He is also a Foreign member of the Royal Swedish Academy of Engineering Science (IVA), a member of Collége International pour l'Etude Scientifique des Techniques de Production Mécanique (CIRP), and the Life Fellow of the Korean Academy of Science and Technology.

Professor Suh is an honorary professor at Yanbian Univ. of Science and Technology, China, and an Eminent Visiting Professor at the Korea Advanced Inst. of Science and Technology, Korea. He has been on visiting committees of Georgia Inst. of Technology, Stanford University, the University of Michigan, the Texas A&M University, and the Univ. of California – Berkeley. He was a member of the Visiting Committee for the National Inst. of Standards and Technology (a statutory committee) and served on a research award committee of ASEE. Professor Suh also served on advisory committees of the Lawrence Livermore National Laboratory, the Idaho National Engineering Laboratory, and Alcan Aluminum Corp.

Professor Suh has served as a consultant for many international and national industrial firms, including the Lawrence Livermore National Laboratory and the Korea Electric Power Research Inst. Professor Suh was the key architect of the Five-Year (1980-85) Economic Development Plan of the Republic of Korea.

Dr Les Piegl

Dr Les A Piegl is professor of computer science and engineering at the University of South Florida. His research interests center around geometric computing including such broad fields as Computer-aided Design and Manufacturing (CAD/CAM), Computer Graphics, and Software Engineering. His speciality is NURBS (Non-Uniform Rational B-splines) in which he does basic research as well as commercial grade software development.

In 1985 Dr Piegl received an *Alexander von Humboldt Research Award* from the Alexander von Humboldt Foundation, Bon-Bad Godesberg, Germany. He also received an *Outstanding Undergraduate Teaching Award* and a *Teaching Incentive Program Award*, both from the University of South Florida, in 1991 and 1995, respectively.

Dr Piegl is the co-author of the text The NURBS Book, published by Springer-Verlag, New York, 1995, 1997. The second edition is in its second printing.

Dr Piegl serves as Editor-in-Chief for Computer-Aided Design, an international journal on computer supported design in engineering. Founded in 1968, the journal is now published by Elsevier Science, Amsterdam, The Netherlands.

Professor D T Pham

Duc Truong Pham, BE, PhD, DEng (Canterbury, NZ), FREng, FIEE, is Professor of Computer-Controlled Manufacture (appointed in October 1988), Head of the Systems Engineering Division, and Founder Director of the award-winning WDA Centre of Excellence in Manufacturing Engineering in the School of Engineering at the University of Wales, Cardiff. He obtained his Bachelor of Engineering (Mechanical) degree with First-Class Honours, Doctor of Philosophy degree and Doctor of Engineering degree from the University of Canterbury in Christchurch. Between 1979–1988 he was a lecturer in Control Engineering at the University of Birmingham. His work at Birmingham focused on robotics and automation. At Cardiff, his research encompasses the wider areas of intelligent systems and advanced manufacturing engineering. His Intelligent Systems Laboratory and Manufacturing Engineering Centre have over 80 researchers and support personnel conducting externally funded projects on intelligent systems and advanced manufacturing technology. These cover knowledge-based systems for manufacturing design and quality control, robot vision and automated visual inspection systems, neural networks for non-linear systems identification and control, rapid prototyping and tooling, virtual reality, and multi-media intelligent product manuals. The amount of external research grants and contracts he has won for his Laboratory and Centre currently totals £20M. He has produced over 50 PhD graduates, written more than 200 technical papers, co-authored four books and edited four other books in his research areas.

His higher doctorate was obtained for published work in the field of systems engineering in the period 1979–1993. He has lectured extensively abroad on his research and has acted as a consultant to several major companies. He is a member of the editorial boards and panels of seven international journals and the editor of the Springer-Verlag Advanced Manufacturing book series. He is the recipient of several prizes including the Sir Joseph Whitworth prize awarded by the Institution of Mechanical Engineers in 1996 and 2001 and the Institution's Thomas Stephens Group Prize in 2001. He is a Fellow of the Royal Academy of Engineering and a Fellow of the Institution of Electrical Engineers.

Foreword

It is with great pleasure that I present to you the proceedings for the International Conference on Manufacturing Automation. This is the third time that the ICMA has been held. It is somewhat coincidental that the ICMA has been held every 5 years since it was first held. Each conference has had its own character and it has been interesting to note in particular the change of emphasis on different themes. One thing to note is that the ICMA has always had a healthy mixture of topics, bringing together experts from different fields, encouraging cross-fertilization of interests and ideas.

ICMA 2002 has also developed its own character. This time we chose to have everyone submit his or her papers electronically and in full. This way, papers have been reviewed from the very start to ensure high quality and effective development of papers throughout the process. An incentive is that a number of journals have expressed interest in the conference proceedings and selected papers will be prepared for them. The result is around 60 top quality papers from no less than 13 different countries on a variety of topics that contribute to our theme of *Rapid Response Solutions to Product Development*. I hope you agree with me that this conference has been very worthwhile and that the proceedings represent a significant contribution to research in this area.

Another key factor in this conference has been the participation of our co-organizers at Huazhong University of Science and Technology, and The National University of Singapore. The organizing committee at The University of Hong Kong have been grateful of the advice and support that has been provided by our colleagues and friends at these other institutions and we look forward to future ICMAs being held, perhaps in less than 5 years' time.

The Organizing Committee would like to acknowledge the very generous support of the sponsors of this conference. Thanks are due to all members of the International Scientific Committee and Keynote Speakers for helping to make this a rich, rewarding and truly international event. The assistance of all session chairs, student helpers, and secretaries are all gratefully appreciated, as is the support of the Department of Mechanical Engineering, The University of Hong Kong.

Ian Gibson
University of Hong Kong

Contents

Machining Technology

Intelligent Systems

Technology Management

Rapid Product Development

Internet-based Systems

Authors' Index

Rapid manufacturing – technologies and applications

D T PHAM and **S S DIMOV**
Manufacturing Engineering Centre, Cardiff University, UK

ABSTRACT

The adoption of rapid manufacturing technologies can greatly enhance the competitiveness of a company. A key enabler of rapid manufacturing is rapid prototyping (RP). This paper categorises existing RP processes according to the mechanism used to transfer CAD data into physical structures. The paper then outlines the most common RP processes, gives examples of their applications and discusses future developments in rapid manufacturing.

Key words: rapid prototyping, rapid tooling, rapid manufacturing

1 INTRODUCTION

Rapid manufacturing is a new mode of operation that has the potential to enable companies to handle the pressures of increasing competition and the drive towards mass customisation. Rapid manufacturing is based on layer manufacturing which is also known as rapid prototyping (RP). RP typically allows small numbers of functional parts to be manufactured directly from CAD data without requiring tooling or fixtures. Large batches of components can be replicated using patterns or tools created by RP methods. In both cases, RP compresses manufacturing lead times and increases the responsiveness and therefore competitiveness of firms adopting this technology.

This paper briefly reviews RP processes, giving examples of their applications in rapid manufacturing and discussing the technical challenges to be overcome for rapid manufacturing to achieve its full potential.

2 RAPID PROTOTYPING

RP processes may be divided broadly into those involving the addition or the removal of material. According to Kruth [1], material accretion processes may be categorised by the state of the prototype material before part formation, namely, liquid, powder or solid sheets.

Liquid-based processes may entail the solidification of a resin on contact with a laser, the solidification of an electrosetting fluid, or the melting and subsequent solidification of the prototype material. Processes using powders (discrete particles) aggregate them either with a laser or by the selective application of binding agents. Those processes which employ solid sheets may be classified into two types depending on whether the sheets are bonded with light or with an adhesive.

Material accretion processes may also be clustered according to the mechanism employed for transferring data from the sliced 3D models into physical structures. Following this method of categorisation, the processes fall into one of four groups.

☐ 1D Channel. The first group of processes transfers data using 1D channels. These data channels may be realised in the form of a laser beam, an extrusion head, a jet of thermoplastic, a nozzle spraying a binder, a welding head, a cutter or a computerised knife.

☐ Multiple 1D Channels. A process in this category would employ multiple 1D channels working in parallel. Currently, there is only one process implementing this data transfer method with two independently controlled lasers. However, this multi-channel approach could be adopted for other processes in the first group to multiply productivity without introducing any changes to the fundamental working principles.

☐ Array of 1D Channels. The third group includes processes that utilise arrays of 1D channels to construct 3D structures. These may be arrays of nozzles or jets. Currently, RP systems with the highest build speeds all use this mechanism for data transfer.

☐ 2D Channel. The fourth group includes processes employing 2D channels, for example masks. At present, there are only a few processes using this mechanism although it offers significant productivity advantages over the other three approaches.

Fig. 1 shows a classification of RP processes that takes into account both the state of materials before part formation and the mechanism employed for data transfer. In the following section, material accretion processes are presented according to the build material used.

2.1. Material Accretion Processes

2.1.1 Liquid Polymer

Of the seven processes in this category, which all involve the solidification of a resin by applying electromagnetic radiation, two construct the part using points to build up the layers whilst the other four solidify entire layers or surfaces at once.

☐ *Stereolithography (SL).* This process relies on a photosensitive liquid resin which forms a solid polymer when exposed to ultraviolet (UV) light. SL systems consist of a build platform (substrate) which is mounted in a vat of resin and a UV Helium-Cadmium or Argon ion laser [2]. The first layer of the part is imaged on the resin surface by the laser using information obtained from the 3D solid CAD model. Once the contour of the layer has been scanned and the interior hatched, the platform is lowered and a new layer of resin is applied. The next layer may then be scanned. Once the part is completed, it is removed from the vat and the excess resin drained. The 'green' part is then placed in an UV oven to be postcured. To broaden the application area of SL, research and technology development efforts have been directed towards process optimisation [3, 4].

☐ *Liquid Thermal Polymerisation (LTP).* This process is similar to SL except that the resin is thermosetting and an infrared laser is used to create voxels (three-dimensional pixels). This means that the size of the voxels may be affected through heat dissipation, which can also cause unwanted distortion and shrinkage in the part [1]. The system is still being researched.
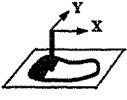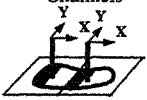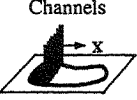
	1D Channel	Multiple 1D Channels	Array of 1D Channels	2D Channel
Liquid Polymer	SL, LTP		Objet	SGC, RMPD, HIS
Discrete Particles	SLS, LST, LENS, SDM, LCVD, SLRS GPD, SALD	LST	3DP	DPS
Molten Mat.	FDM, BPM, 3DW, PDM		MJM	
Solid Sheets	LOM, PLT			SFP
Electroset Fluids				ES

Fig. 1 Classification of rapid prototyping processes

Ballistic Particle Manufacture (BPM)
Direct Photo Shaping (DPS)
Electrosetting (ES)
Fused Deposition Modelling (FDM)
Gas Phase Deposition (GPD)
Holographic Interference Solidification (HIS)
Laminated Object Manufacturing (LOM)
Laser-Assisted Chemical Vapor Deposition (LCVD)
Laser Engineering Net Shaping (LENS™)
Laser-Sintering Technology (LST)
Liquid Thermal Polymerisation (LTP)
Multi Jet Modelling (MJM)
Objet Quadra Process (Objet)
Paper Lamination Technology (PLT)
Precision Droplet-Based Net-Form Manuf. (PDM)
Rapid Micro Product Development (RMPD)
Shape Deposition Manufacturing (SDM)
Selective Area Laser Deposition (SALD)
Selective Laser Reactive Sintering (SLRS)
Selective Laser Sintering (SLS)
Solid Foil Polymerisation (SFP)
Solid Ground Curing (SGC)
Stereolithography (SL)
Three-Dimensional Printing (3DP)
Three-Dimensional Welding (3DW)

☐ *Holographic Interference Solidification (HIS).* A holographic image is projected into the resin causing an entire surface to solidify. Data is still obtained from the CAD model, although not as slices [1]. There are no commercial systems available yet.

☐ *Solid Ground Curing (SGC).* This system again utilises photopolymerising resins and UV light [5]. Data from the CAD model is used to produce electrostatically a mask on a glass that is developed using a toner. Then, the mask is placed above the resin surface and the entire layer is illuminated with a powerful UV lamp. Once the layer has been cured, the excess resin is wiped away and any spaces are filled with wax. The wax is cooled with a chill plate, milled flat and the wax chips removed. A new layer of resin is applied and the process is repeated.

☐ *Rapid Micro Product Development (RMPD).* The RMPD process is a mask-based technology very similar to that of photolithography as used in microelectronics manufacture [6]. CAD data is employed to produce masks for laser polymerisation of a liquid photoresin in a layer-by-layer fashion. The process allows micro components to be built with a minimum layer thickness of 1 μm and X-Y resolution of 10 μm. In addition,

this process can be used to create complex micro systems that integrate electronics, optical and mechanical components.

☐ *Objet Quadra Process.* The process employs 1536 nozzles to build parts by spreading layers of photo sensitive resin that are then cured, layer by layer, using two UV lights. The intensity of the lights and the exposure are controlled so that models produced by the system do not require post-curing. To support overhanging areas and undercuts, Objet deposits a second material which can be separated easily from the model [7].

2.1.2 Molten Material

There are six processes which involve the melting and subsequent solidification of the part material. Of these, the first five deposit the material at discrete points whilst the sixth manufactures whole layers at once.

☐ *Ballistic Particle Manufacture (BPM).* The process builds parts by ejecting a stream of molten material from a nozzle. The stream separates into droplets which hit the substrate and immediately cold weld to form the part [8]. Commercial systems based on this process were available until 1998.

☐ Multi Jet Modelling (MJM). The process builds models using a technique similar to inkjet or phase-change printing but applied in three dimensions [9]. A "print" head comprising 352 jets forming a linear array builds models in successive layers, each individual jet depositing a specially developed thermo-polymer material only where necessary (Fig. 2). The MJM head shuttles back and forth along the X axis like a line printer. If the part is wider than the MJM head, the platform repositions itself (Y-axis) to continue building the layer. When a layer is completed, the platform is moved away from the head (Z-axis) which begins to create the next layer. At the end of the build, support structures are brushed off to finish the model.

Material Supply
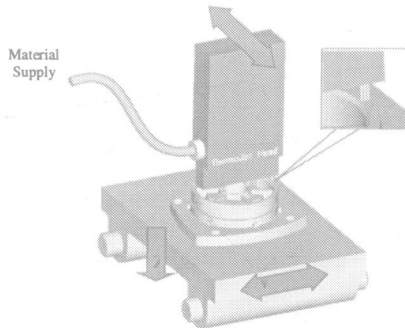
Fig. 2 Multi Jet Modelling Head

☐ *Fused Deposition Modelling (FDM).* FDM systems consist of two movable heads (one for building the part and one for the supports) which deposit threads of molten material onto a substrate. The material is heated just above its melting point so that it solidifies immediately after extrusion and cold welds to the previous layers [10].

☐ *Three-Dimensional Welding (3DW).* This experimental system uses an arc-welding robot to deposit material on a platform as simple shapes which may then be built into more complex structures [11]. Unlike most RP processes, the prototypes are built using CNC programs generated directly from the CAD files instead of employing slice data. Another

experimental system deposits the weld material in layers. Feedback control is established by means of thermocouples which monitor the temperature and operate an on-line water cooling system. A grit blasting nozzle minimises the oxidisation of the part and a suction pump and vacuum nozzle remove excess water vapours and grit [12].

☐ *Precision Droplet-Based Net-Form Manufacturing (PDM).* This is a droplet-based net-forming manufacturing technique [13]. The process exploits the capillary instability phenomenon of liquid jets for producing uniform liquid metal droplets. The thermal state and mass flux of the droplets can be controlled to tailor the microstructure of the deposit. There is no commercial system based on this process.

☐ *Shape Deposition Manufacturing (SDM).* Still experimental, this layer-by-layer process involves spraying molten metal in near net shape onto a substrate, then removing unwanted material via NC operations [14]. Support material is added in the same way either before or after the prototype material depending on whether the layer contains undercut features (Fig. 3). The added material bolsters subsequent layers. If the layer is complex, support material may need to be added both before and after the prototype material. Each layer is then shot-peened to remove residual stresses. The prototype is transferred from station to station using a robotised pallet system. To date, stainless steel parts supported with copper have been produced. The copper may then be removed by immersion in nitric acid. These prototypes have the same structure as cast or welded parts and the accuracy of NC milled components.

☐
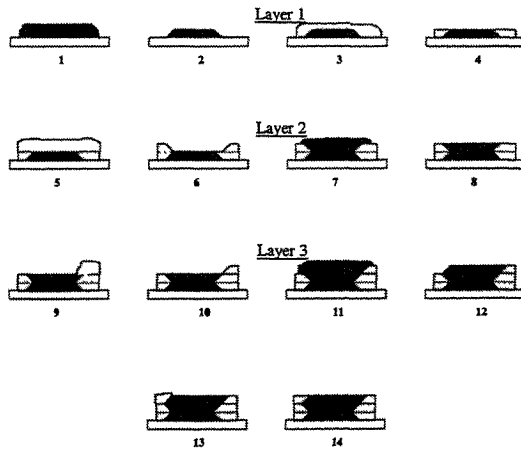
Fig. 3 Shape Deposition Manufacturing. The construction of the first 3 layers of a part is shown [14].

2.1.3 Processes Involving Discrete Particles

These processes build the part by joining powder grains together using either a laser or a separate binding material. The main processes in this category are described briefly below.
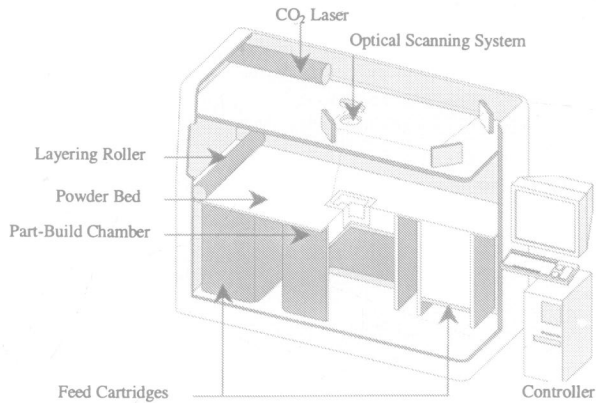
Fig. 4 Selective Laser Sintering
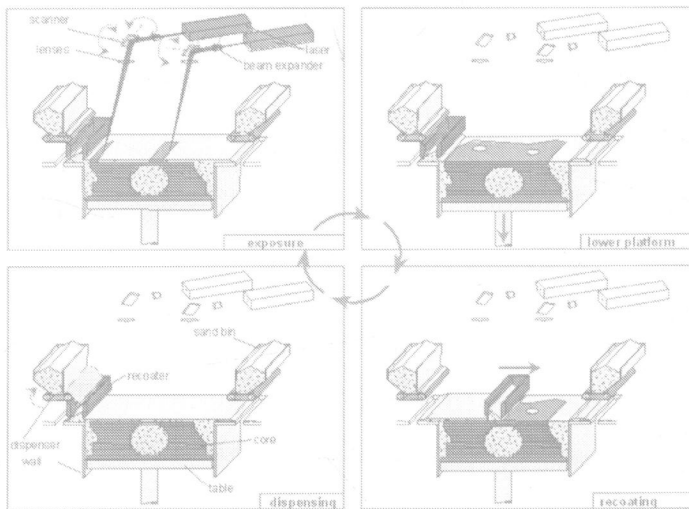
Fig. 5 A dual laser LST system (Courtesy of EOS GmbH)

☐ *Selective Laser Sintering (SLS)*. SLS uses a fine powder which is heated with a CO_2 laser so that the surface tension of the particles is overcome and they fuse together. Before the powder is sintered, the entire bed is heated to just below the melting point of the material in order to minimize thermal distortion and facilitate fusion to the previous layer [15]. The laser is modulated such that only those grains which are in direct contact with the beam are affected. A layer is drawn on the powder bed using the laser to sinter the material. The bed is then lowered and the powder-feed cartridge raised so that a covering of powder can be spread evenly over the build area by a counter-rotating roller. The sintered material forms the part whilst the unsintered powder remains in place to support the structure and may be cleaned away and recycled once the build is complete (Fig. 4). There is another process, *Laser Sintering Technology (LST)* that employs the same

physical principles. Fig. 5 shows a LST system equipped with two laser beams working in parallel. Currently such dual-laser systems are available for processing thermoplastics and sand. Significant development efforts have been directed towards process optimisation [2, 15, 16, 17, 18] to widen the range of applications of SLS and LST.
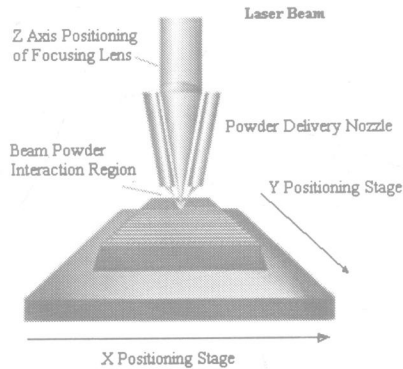
Fig. 6 LENS™ process (Courtesy of Optomec Design Co)

☐ *Laser Engineering Net Shaping (LENS™)*. The LENS process involves feeding powder through a nozzle onto the part bed whilst simultaneously fusing it with a laser (Fig. 6) [19]. The powder nozzle may be on one side of the bed or coaxial with the laser beam. If it is to a side, a constant orientation to the part creation direction must be maintained to prevent solidified sections from shadowing areas to be built. When the powder feeder is coaxial, there may be inaccuracies in the geometry of the part and the layer thickness if the beam and the powder feeder move out of alignment. Because the stream of powder is heated by the laser, fusion to the previous layer is facilitated. Other systems have also been developed based on the same principle, in particular, *Direct Metal Deposition* (DMD) [20] and *AeroMet Laser Additive Manufacturing* [21].

☐ *Gas Phase Deposition (GPD)*. In this process, the molecules of a reactive gas are decomposed using a laser to generate a solid [2]. The resulting solid then adheres to the substrate to form the part. Three slightly different methods of constructing the part have been investigated. With the first method, called *SALD (Selective Area Laser Deposition)*, the solid component of the decomposed gas is all that is used to form the part. It is possible to construct parts made from carbon, silicon, carbides and silicon nitrides in this way. The second method, *LCVD (Laser-Assisted Chemical Vapor Deposition)*, spreads a thin covering of powder for each layer and the decomposed solids fill in the spaces between the grains. With the third method, *SLRS (Selective Laser Reactive Sintering)*, the laser initiates a reaction between the gas and the layer of powder to form a solid part of silicon carbide or silicon nitride. There are no commercial GPD systems available yet.

☐ *Direct Photo Shaping (DPS)*. The process employs a Digital Micromirror Device (DMD™) array [22] as a mask to photocure selectively layer-by-layer polymerisable compositions. The DMD array integrates more than 500,000 microscopic mirrors that can

be electronically tilted to reflect visible light onto the photocurable slurry [23]. No commercial DPS systems are available yet.

☐ *Three-Dimensional Printing (3DP)*. The process builds parts by first applying layers of powder to a substrate and then selectively joining the particles using a binder sprayed through a nozzle [24]. Once the build is completed, the excess powder, which was supporting the model, is removed leaving the fabricated part. Since there is no state change involved in this process, distortion is reduced [25].

2.1.4 Solid Sheets

There are three different processes that employ foils to form the part, namely:

☐ *Laminated Object Manufacturing (LOM)*. The build material is applied to the part from a roll, then bonded to the previous layers using a hot roller which activates a heat-sensitive adhesive [26]. The contour of each layer is cut with a CO_2 laser that is carefully modulated to penetrate to the exact depth of one layer. Unwanted material is trimmed into rectangles to facilitate its later removal but remains in place during the build to act as supports. Separating LOM models from the surrounding excess material can still be a lengthy and tedious task. A method that speeds up and simplifies it has recently been developed [27].

☐ *Paper Lamination Technology (PLT)*. The PLT process is very similar to LOM. The main differences between the LOM and PLT processes are in the material used and the methods employed for cutting the contours of the part cross-sections, which in the case of the PLT process is a computerised knife. The PLT process prints the cross section of the part onto a sheet of paper, which is then applied to the work-in-progress and bonded using a hot roller [28].
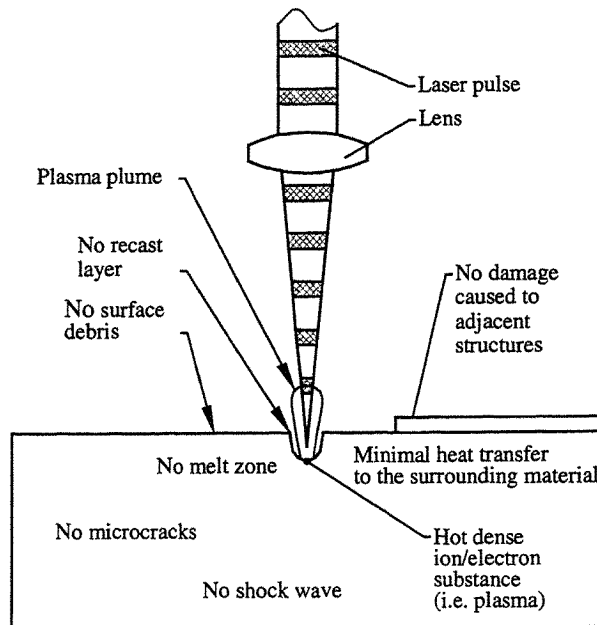
Fig. 7 Femto and picosecond laser ablation

☐ *Solid Foil Polymerisation (SFP)*. The part is built up using semi-polymerised foils which are soluble in monomer resin. On exposure to UV light, the foil solidifies and bonds to the previous layer. It also becomes insoluble. Once the cross-section has been illuminated, a new foil can be applied. The areas of foil which do not constitute the eventual part are used to support it during the build process, but remain soluble and so are easy to remove [1, 29]. No commercial systems are available yet.

2.1.5 Electroset Fluid: Electrosetting (ES)

Electrodes are printed onto a conductive material such as aluminium. Once all the layers have been printed, they are stacked, immersed in a bath of electrosetting fluid and energised. The fluid which is between the electrodes then solidifies to form the part. Once the composite has been removed and drained, the unwanted aluminium may be trimmed from the part.

Advantages of this technology are that the part density, compressibility, hardness and adhesion may be controlled by adjusting the voltage and current applied to the aluminium. Parts may be made from silicon rubber, polyester, polyurethane or epoxy. The hardware for such a system may be inexpensively bought off the shelf [12].

2.2 Material Removal Processes

This category includes two processes, *Desktop Milling (DM)* and *Laser Milling (LM)*. DM is a process which removes material from the workpiece as in traditional machining instead of creating the part by gradual material build up [30]. Prototypes can be made with a high degree of accuracy because they do not deform after they have been completed.

LM is a new process for fabricating relatively small prototype components in advanced engineering materials such as ceramics, titanium and nickel alloys. This process removes material as a result of interaction between a laser beam and a workpiece. Several mechanisms exist for material removal, depending on the laser pulse duration and some material-specific time parameters [31-33]. The laser ablation mechanisms for femtosecond and picosecond pulses are alike and can be regarded as direct solid-vapour transition (sublimation), with negligible thermal conduction into the substrate and almost no heat affected zone (Fig. 7) [33-37]. For nanosecond and longer pulses, the absorbed energy from the laser pulse melts the material and heats it to the vaporisation temperature (Fig. 8). There is enough time for a thermal wave to propagate into the material. Evaporation occurs from the liquid material. The molten material is partially ejected from the cavity by the vapour and plasma pressure, but a part of it remains near the surface, held by surface tension forces. After the end of a pulse, the heat quickly dissipates into the bulk of the material and a recast layer is formed.

A number of techniques for LM have been developed. They differ from one another in the applied laser source, the relative beam-workpiece movements and the laser spot characteristics. A common feature of all LM techniques is that the final part geometry is created in a layer-by-layer fashion by generating overlapping craters. Within an individual layer, these simple volumes are arranged in such a manner that each slice has a uniform thickness. Through relative movements of the laser beam and the workpiece, the micro craters produced by individual laser pulses sequentially cover complete layers of the part. The LM process is flexible and can be employed in a wide range of applications from one-off part production to the manufacture of small batches [38-40].
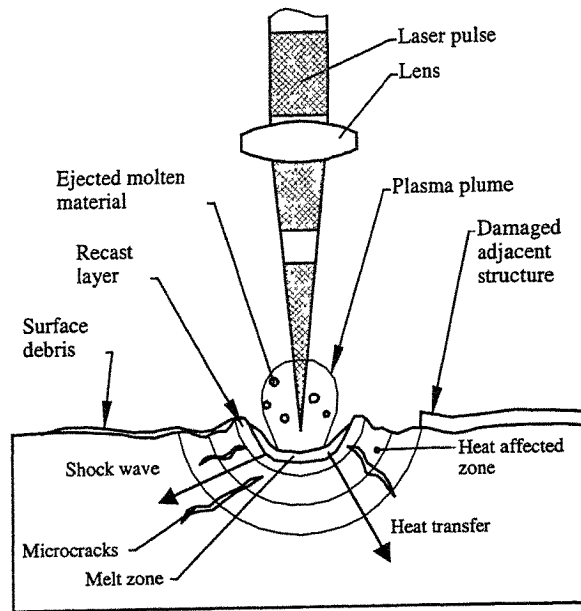
Fig. 8 Nanosecond and longer pulse laser ablation

3 APPLICATIONS

Initially conceived for design approval and verification, RP now meets the needs of a wide range of applications from directly building small batches of functional parts to fabricating tools for replicating them in larger quantities. Examples of different applications are described in this section.

3.1 Functional Parts
One of the RP processes that is widely used for producing functional Polyamide-based models is SLS. The SLS production of Polyamide parts is generally cost effective when a small number (1-5) of parts is required.

The housing in Fig. 9 is a test part built in glass-filled Polyamide (a blend of 50% by weight Polyamide powder with a mean particle size of 50μm and 50% by weight spherical glass beads with an average diameter of 35μm) because it is required to withstand harsh test conditions including temperatures of about 100°C. As a base part for mounting precision components, it has to keep its dimensions within close limits.

Due to its overall dimensions (190x50x250 mm), the part was constructed vertically to fit within the build area (⌀305x410 mm) of the SLS machine used (DTM Sinterstation 2000). The first part manufactured suffered from much distortion: there was vertical growth and "wash out" (loss of definition and rounding of edges) on the downward facing surfaces and the external dimensions of the sidewalls varied by more than 1mm. This problem was solved

by making the wall thickness uniform and reducing it to 2mm. Furthermore, 2mm non-functional ribs were added across the housing to stiffen it. Two ribs were positioned vertically and two others horizontally as shown in Fig. 9. The number and size of the ribs were determined empirically to constrain post process distortion in the X and Y directions without adding too much build time. The ribs were also located so that they could easily be removed by machining after completing the build. Subsequently manufactured parts had much better dimensional accuracy. The errors in 90% of all functional dimensions for the modified part were between + 0.35 and − 0.31mm.
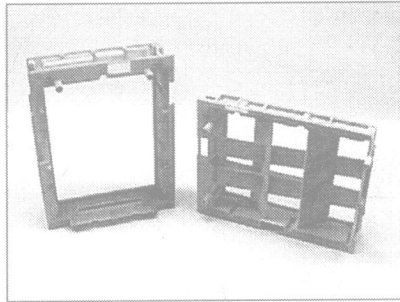
Fig. 9 Composite Nylon housing: Without ribs (left) – With ribs (right)

3.2 Patterns for Investment and Vacuum Casting

RP technologies are widely used for building patterns for investment and vacuum casting. For example, models built in SLA, SLS and FDM can be employed as patterns for both casting processes. The example discussed below illustrates the use of the SLS process to build investment casting patterns in CastForm [2]. CastForm is a polystyrene-based powder that gives a low ash content and is compatible with standard foundry practices. Processing CastForm creates porous low-density parts that have to be subsequently infiltrated with a low-ash foundry wax to yield patterns containing 45% polystyrene and 55% wax.

The heat exchanger assembly of a Pratt & Whitney PW6000 engine shown in Fig. 10 was produced using CastForm patterns. The assembly includes three cast aluminium components that have to withstand high temperature and pressure. These complex castings are essentially pressure vessels with multiple portings, mountings and sensor pads. The largest component measures 600 mm in height and 325 mm in diameter (Fig. 10). Several sets of sacrificial casting patterns were built using the SLS process. The errors in 90% of all functional dimensions were between ± 0.25 mm. The accuracy of the patterns was highly dependent on their size, the largest errors being found on the largest dimensions. However, although some dimensions were out of the required general tolerances (± 0.125 mm), the aluminium castings were fully satisfactory as any deviations were corrected when some of the features were machine-finished afterwards.

The main benefit of employing the SLS process was that the design team was able to incorporate major and minor modifications into the CAD models between the builds. There was no need to freeze the design before proceeding to manufacture. The prototype heat exchangers underwent stringent testing before the design was approved. As a relatively small number of exchangers was required per year, the SLS process was approved as a production

method for the fabrication of the required casting patterns. In general, RP patterns are a cost effective alternative when a small number of parts, say up to 50, of complex design are required and the cost of a mould tool for wax patterns is prohibitive.
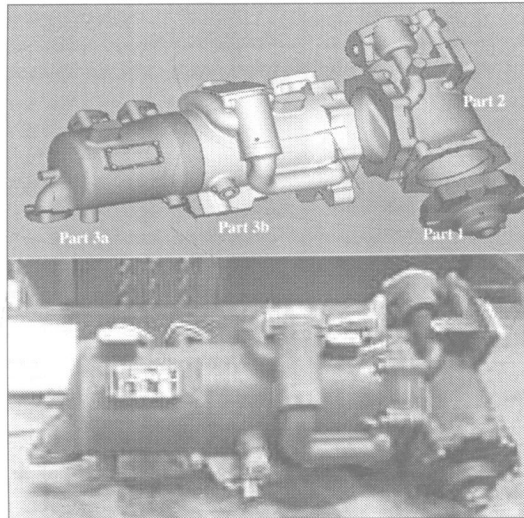
Fig. 10 Heat exchanger for a Pratt&Whitney PW6000 engine

3.3 Rapid Tooling

As RP becomes more mature, material properties, accuracy, cost and lead-time have improved to permit it to be employed for the rapid production of tools. Conventional tool making methods based on the replication of models have been adapted, as illustrated in the previous section. This section gives examples of tools fabricated directly by RP. The process adopted for producing the tools was *RapidTool*^{TM} which ultilises SLS to produce metal inserts for mould tools.

The insert material was *LaserForm*^{TM} [2]. This is a powder made up of 420-stainless-steel-based particles, coated with a thermoplastic binder. The processing of LaserForm can be broken down into two main stages. During the first stage (the "green" stage) tooling inserts are built layer by layer through fusion of the binder in a SLS machine. In the second stage (oven cycle), the green part is converted into a fully dense metal part by infiltration with molten bronze. During the oven cycle, at between 450 and 650°C the polymer evaporates and at 700°C the sintering of the remaining steel powder begins. Then, the inserts are heated up to 1070°C where bronze infiltration occurs driven by capillary action. To avoid oxidation of the steel surfaces, all processing is done in a nitrogen atmosphere. The final LaserForm inserts are 60% stainless steel 40% bronze fully dense parts which can be finished by any technique including surface grinding, milling, drilling, wire erosion, EDM, polishing and surface plating.

Die casting inserts. To evaluate the applicability of *RapidTool*^{TM} to aluminium gravity die-casting, inserts for a windscreen wiper arm were built (Fig. 11). The inserts were finished following the steps described in [2]. One of the bosses at the end of the wiper was used as a reference feature to achieve good matching of the two halves of the tool. The tool was used to

cast parts in LM6 aluminium alloy. After producing 250 castings in four separate runs, no degradation signs were visible on the inserts surfaces or on the cast parts. The tests showed that *RapidTool*^{TM} dies can be utilised for production of low to medium size batches of castings. Given the quality of the die material, it is estimated that over five thousand castings could easily be produced from the dies.

Injection moulding inserts. This example illustrates the capability of the *RapidTool*^{TM} process for fabrication of injection moulding inserts. An insert was manufactured for moulding the cap for a nose hair trimmer. While the external surface of the part is relatively simple, its internal features are much more complex. The internal surface of the cap consists of a cone that transforms progressively into a square hole. The hybrid approach adopted was to machine the mould conventionally from steel and to make the core using the *RapidTool*^{TM} process. As shown in Fig. 12, the *RapidTool*^{TM} core was built without the protruding pin. This feature was judged to be too small and weak to be reproduced reliably by the *RapidTool*^{TM} process. Because of its simple shape, the pin was machined from steel and added later to the insert. The tool thus manufactured by combining the capabilities of conventional tooling techniques and the *RapidTool*^{TM} process was successfully used to produce several hundred mouldings in ABS.
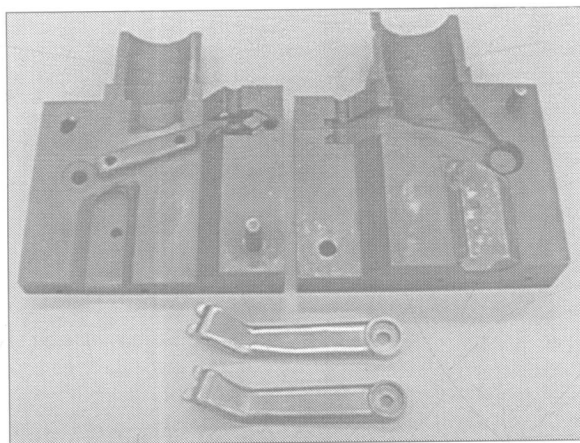
Fig. 11 Die casting tool produced from *RapidTool*^{TM} inserts

4 FUTURE TRENDS

Research in the field of RP is just over 10 years old but in spite of this significant progress has been made in widening the use of this technology and in the development of new processes and materials. To achieve long-term growth in this field and realise fully the potential of RP as a means to rapid manufacturing, a number of challenges remain. These challenges could be grouped under the following categories:
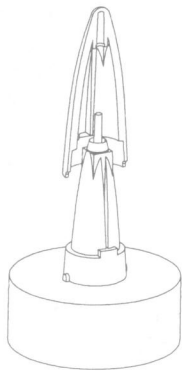
(a) Core and Cavity
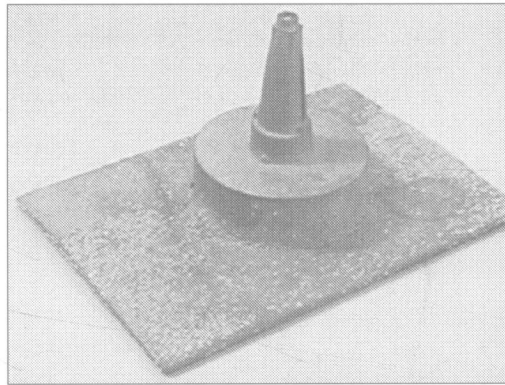
(b) RapidTool core
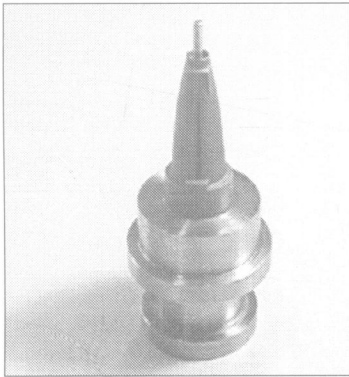
(c) Finished Core
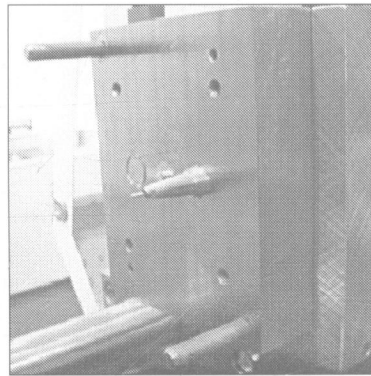
(d) Core mounted in injection moulding bolster

Fig. 12 Building stages of the cap insert (3D design, infiltration, finishing, integration)

☐ *Materials.* One of the main limitations of RP processes is the limited variety of materials and their properties, and also their relatively high cost. Significant research efforts are focused on the development of a broader range of materials that simulate very closely the properties of the most commonly used engineering plastics. In particular, much research is being conducted on the development of new materials with high rigidity, high impact strength and high tensile elongation at breaking. Also, a range of materials for fabrication of investment casting patterns with low ash content, high impact strength and good surface finish are currently under development. Recently, the fabrication of multi-materials and heterogeneous objects has attracted the attention of the research community. This is quite understandable because RP is well suited to building such objects. Functionally gradient components could be manufactured from different constituent materials exhibiting continuously varying composition and/or micro structure. Developments in this area will make possible the fabrication of objects with multiple and

conflicting functionality. Progress in the area is directly linked with the development of new CAD tools that are suitable for designing heterogeneous objects.

❑ *Process Planning*. Although process plans for building complex RP parts are reduced to containing only three operations (which most often are building parts, inspection and finishing which can include painting) compared to the many steps required by conventional material removal processes, the process planning tasks associated with layer manufacturing requires special attention. These tasks include selecting the part orientation, identifying the required support structures, slicing and deposition path planning and the specification of process parameters. Existing approaches to addressing these problems fall into two categories: algorithmic and decision-support solutions [42]. The algorithmic approach relies on geometrical reasoning mechanisms to find solutions for these tasks. For example, this approach is used to: determine the part orientation in respect to some user defined criteria (minimisation of the support structures required, avoidance of trapped volumes, improving part quality and engineering properties), study the influence of different deposition patterns and process parameters on part properties, identify overhanging features requiring support structures utilising STL file facets, solid models or slice data, and develop new techniques for slicing (adaptive slicing and slicing of heterogeneous objects). The second approach employs decision-support methods to perform tasks that require the trade-offs between competing goals to be quantified. Such process planning methods employ multi-criteria optimisation techniques, analytical models and heuristics [43]. With the increase of part complexity and the range of available RP materials and RP machines, there is a need for more advanced process planning tools, in particular, tools that could relate process variables to part quality characteristics and address the process-specific requirements associated with the fabrication of parts from heterogeneous materials.

❑ *RP Data Formats and Design Tools*. The STereoLithography format (STL) was introduced in the early years of RP technology and is considered a de facto standard for interfacing CAD and RP systems. The STL format has a number of drawbacks [44] that are inherent in the representation scheme employed. The use of other standard formats for product data exchange such as IGES, HPGL, STEP and VRML has been considered in place of STL but as problems remain these alternative formats are not widely accepted. Work on the development of new formats continues in order to address the growing requirements of RP and RT applications for more precise methods of data representation. Also, in recent years with the emergence of RP processes for fabrication of heterogeneous objects, there is an increasing interest within the research community in developing new CAD tools that enable objects with varying material composition and/or microstructure to be designed [42]. Currently, a number of CAD systems for constructing such objects are under development employing: voxel-based methods [45], generalised cellular decomposition [46], finite element based methods [46, 47] and constructive methods [48]. As already mentioned, advances in this area are directly linked to research and development in technologies capable of producing materially graded structures.

5 CONCLUSION

In the current state of the art, RP enables the economic *direct* rapid manufacture of small batches of functional parts in various materials including polymers, metals and ceramics. As seen in this paper, there are also *indirect* rapid manufacturing processes employing RP to

create tooling to replicate components in large quantities. However, such indirect processes incur additional time and expenditure compared to direct rapid manufacturing. With intense efforts being spent on the development of new RP machines, processes and materials, as exemplified by the 208 RP-related patents filed in the past two years [49], it should be possible to overcome the previously mentioned technical challenges in the foreseeable future. This would then make RP appropriate as a route to direct rapid manufacture for all but mass production applications.

ACKNOWLEDGEMENTS

REFERENCES

1. **Kruth J.P.** Material Incress Manufacturing by Rapid Prototyping Technologies, *CIRP Annals*, 1991, Vol. 40, 2, pp 603-614.

2. **Pham D.T. and Dimov S.S.** Rapid Manufacturing: The Technologies and Applications of Rapid Prototyping and Rapid Tooling, 2001, *Springer Verlag*, London.

3. **Pham D. T. and Ji C.** Design for Stereolithography, *Proc. IMechE, Part C Journal of Mechanical Engineering Science*, 2000, Vol. 214, no. 5, pp. 635-640

4. **Onuh S. O and Hon K. K.** Application of the Taguchi Method and New Hatch Styles for Quality Improvement in Stereolithography, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 1998, Vol. 212, no. 6, pp. 461-472

5. **Cubital Web page,** 2001, Cubital Ltd., 13 Ha'Sadna St., Industrial Zone North, Ra'anana, 43650, Israel, http://www.cubital.com/

6. **MicroTEC Web page,** 2001, Gesellschaft für Mikrotechnologie GmbH, Bismarckstraße 142b, D-47057 Duisburg, Germany, www.microtec-d.com

7. **Objet Web page,** 2000, Objet Geometries Ltd. Rehovot, Israel, *http://clients.tia.co.il/objet/inner/products.html.*

8. **Sachs E, Cima M, Williams P, Brancazio D and Cornie J** Three Dimensional Printing: Rapid Tooling and Prototyping Directly from a CAD Model, Transactions of ASME: *Journal of Engineering for Industry*, November 1992, Vol. 114, pp 481-488.

9. **3D Systems Press Release:** ThermoJet, 1998, *3D Systems*, Worldwide Corporation HQ, 26081 Avenue Hall, Valencia, California, USA.

10 **Stratasys Web page,** 2001, Stratasys, Inc. 14950 Martin Drive, Eden Prairie, MN 55344-2020 USA, www.stratasys.com.

11 **Spencer J. D., Dickens P. M. and Wykes C. M.** Rapid Prototyping of Metal Parts By Three-Dimensional Welding, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 1998, Vol 212, no. 3, pp. 175-182

12 **Anon,** State of the Art Review-93-01, 1993, *MTIAC*, 10 West 35 Street, Chicago, IL 60616, USA

13. **Liu Q. and Orme M.** On Precision Droplet-based Net-form Manufacturing Technology, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 2001, Vol. 215, no. 10, pp. 1333-1355

14. **Merz R., Prinz F.B., Ramaswami K., Terk M. and Weiss L.F.** Shape Deposition Manufacturing, Proceedings of the 5th Symposium on *Solid Freeform Fabrication*, 8-10 August 1994, Austin, Texas, pp 1-8.

15. **Pham D.T., Dimov S.S. and Lacan F.** Selective Laser Sintering: Applications and Technological Capabilities, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 1999, Vol 213, pp 435-449

16. **Pham D. T. and Wang X.** Prediction and Reduction of Build Times for the Selective Laser Sintering Process, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 2000, Vol. 214, no. 6, pp. 425-430

17. **Childs T. H, Berzins M., Ryder G. R, Tontowi A.** Selective Laser Sintering of an Amorphous Polymer-Simulations and Experiments, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 1999, Vol. 213, no. 4, pp. 333-349

18. **Kathuria Y. P.** Metal Rapid Prototyping Via a Laser Generating/Selective Sintering Process, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 2000, Vol. 214, no. 1, pp. 1-9

19. **Optomec Web page,** 2000, Optomec Design Company, 2701-D Pan American Freeway - Albuquerque, New Mexico - 87107, USA, http://www.optomec.com/.

20. **POM Web page,** 2001, Precision Optical Manufacturing, 44696 Helm Street, Plymouth Michigan 48170, *www.pom.net*

21. **AeroMet Web page,** 2001, AeroMet Corporation, 7623 Anagram Drive, Eden Prairie, Minnesota 55344, USA, *www.aerometcorp.com*

22. **Texas Instruments Web page,** Digital Light Processing, 2001, Texas Instruments, USA, *www.dlp.com*

23. **SRI Web page,** 2001, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025-3493, USA, *http://pguerit.sri.com/SriWeb/srihome.html*

24. **MIT Web page,** 1999, MIT, Three Dimensional Printing Group, http://me.mit.edu/groups/tdp/.

25. **Sachs E., Cornie J., Brancazio D., Bredt J., Curodeau A., Fan T., Khanuja S., Lauder A., Lee J. and Michaels S.** Three Dimensional Printing: the Physics and Implications of Additive Manufacturing, *CIRP Annals*, 1993, Vol. 42, 1, pp 257-260.

26. **Helisys Web page,** 2000, Helisys, Inc., 24015 Garnier Street, Torrance, California 90505-5319, USA, *http://helisys.com/*

27. **Karunakaran K. P., Dibbi S., Shanmuganathan P.V., Raju D. S, Kakaraparti S.** Optimal Stock Removal in Lom-Rp, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 2000, Vol. 214, no. 10, pp. 947-951

28. **KIRA Web page,** 2000, KIRA Corporation, Omiyoshishinden, Kira-Cho, Hazu-Gun, Aichi, Japan, *www.kiracorp.co.jp*

29. **Corbel S., Allanic A.L., Schaeffer P. and Andre J.C.** Computer-Aided Manufacture of Three-Dimensional Objects by Laser Space-Resolved Photopolymerization, *Journal of Intelligent and Robotic Systems*, 1994, Vol. 9, pp 310-312.

30. **Song Y. and Chen Y. H.** Feature-Based Robot Machining for Rapid Prototyping, *Proc. IMechE, Part B: Journal of Engineering Manufacture*, 1999, Vol. 213, no. 5, pp. 451-459

31. **Chichkov B. N., Momma C., Nolte S., von Alvensleben F. and Tuennermann A.** Femtosecond, Picosecond and Nanosecond Laser Ablation of Solids, *Applied Physics*, 1996, A63, pp 109-115

32. **Momma C., Nolte S., Chichkov B. N., von Alvensleben F. and Tünnermann A.** Precise Laser Ablation with Ultrashort Pulses, *Applied Surface Science*, 1997, Vol. 109-110, pp 15-19

33. **Pham D.T., Dimov S.S., Petkov P.P. and Petkov S.P.** Laser Milling, *Proc. IMechE, Part B: Journal of Engineering Manufacture,* 2001, submitted.

34. **Shirk M. D. and Molian P. A.** A Review of Ultrashort Pulsed Laser Ablation of Materials, *Journal of Laser Applications*, 1998, Vol. 10, No 1, pp 18-28

35. **Kautek W. and Krüger J.** Femtosecond Pulse Laser Ablation of Metallic, Semiconducting, Ceramic and Biological Materials, *Proceedings SPIE*, 1994, Vol. 2207, pp 600-610

36. **Preuss S., Demchuk A. and Stuke M.** Sub-picosecond UV Laser Ablation of Metals, *Applied Physics*, 1995, A61, pp 33-37

37. **von der Linde D. and Sokolowski-Tinten K.** The Physical Mechanisms of Short-Pulse Laser Ablation, *Applied Surface Science*, 2000, Vol. 154-155, pp 1-10

38. **Toenshoff H. K., von Alvensleben, Ostendorf A., Willmann G. and Wagner T.** Precision Machining Using UV and Ultrashort Pulse Laser, *SPIE Proceedings*, 1999, Vol. 3680, pp 536-545

39. **Mendes M., Oliveira V., Vilar R., Beinhorn F., Ihlemann J. and Conde O.** XeCl Laser Ablation of Al_2O_3-TiC Ceramics, *Applied Surface Science*, 2000, Vol. 154-155, pp 29-3

40. **Pham D.T., Dimov S.S., Petkov P.P. and Petkov S.P.** Rapid Manufacturing of Ceramic Parts, *Proc. of 17th National Conference of Manufacturing Research*, 2001, Professional Engineering Publishing, London, pp 211-216

41. **Birnnard M.** Design by Composition for Rapid Prototyping, 1st Ed., Kluwer Academic, 1999, Dordrecht

42. **Dutta D., Prinz F.B., Rosen D. and Weiss L.** Layer Manufacturing: Current Status and Future Trends, Transactions of the ASME, 2001, Vol. 1, pp 60-71

43. **West A.P., Sambu S. and Rosen D.W.** A Process Planning Method for Improving Build Performance in Stereolithography, Computer-Aided Design, 2001, Vol. 33, No 1, pp 65-80

44. **Kumar V. and Dutta D.** An Assessment of data Formats for Layer Manufacturing, 1997, Adv. Eng. Software. Vol. 28, No.3, pp. 151-164

45. **Wu Z.W., Soon S.H. and Feng L.** NURBS-based Volume Modelling, Int. Workshop on Volume Graphics, 1999, pp. 321-330

46. **Jackson T., Liu H., Patrikalakis N.M., Sachs E.M. and Cima M.J.** Modelling and Designing Functionally Graded Material Components for Fabrication with Local Composition Control, Materials and Design, 1999, Special Issue, Elsevier Science, Netherlands

47. **Kumar V.A. and Wood A.** Representation and Design of Heterogeneous Components, Proc. SFF Conference, 1999, Austin, TX

48. **Bhashyam S., Shin K.H. and Dutta D.** An Integrated CAD System for Design of Heterogeneous Objects, Rapid Prototyping Journal, Vol. 6, No. 2

49. **Wohlers T.** Wohlers Report 2000: Executive Summary, Time-Compression Technologies, 2000, Vol. 8, 4, pp 29-31.

Design Systems and
Methodologies

Risk mitigation investment in concurrent design process

S AMORNSAWADWATANA, A AHMED, B KAYIS, and H KAEBERNICK
School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney, Australia

ABSTRACT

Concurrent Engineering has been a proven methodology for developing products and processes especially with a design emphasis, where the design process is often critical to the success or failure of the project. Product specifications, costs, materials, and manufacturing processes are influenced by design. Failure in a design activity generates extra costs in the form of redesigning, reworking, rescheduling and resource reallocation and when extra iterations are encountered. Moreover, the company may also experience loss of opportunities and potential loss of customers. Thus, a continuity of the design process must generally be maintained. Success probability of a task indicates the success ratio of an individual design task and depends on behavior of risk factor(s) in a particular task. Risk mitigation and regular monitoring programs implemented in design can improve the success probability of a design process. However, a risk mitigation program may generate additional cost, time, and/or effort. This paper describes a risk management methodology based on IDEF3 system definition to describe the design process. Overall success probability of a design process was calculated based on the IDEF3 model and the system reliability was tested. An investment analysis was carried out to see whether or not it was worth investing in a mitigation and maintenance program. An appropriate decision can be made based on the benefits from the proposed investment program using further simulation as detailed in this paper.

Keywords: Concurrent engineering; Risk mitigation; Reliability; Annual worth method; IDEF3 Process modeling.

1 INTRODUCTION

Concurrent engineering (CE) is the parallel development of design and manufacturing processes [1,5,10]. Team members from different departments carry out tasks simultaneously in design. Information is shared among different departments to reach an overall understanding of the product and process, and information from implementation stage is shifted to the design stage. Consequently, product design is thoroughly understood and the right design can be achieved in an early stage of product life cycle compared to a sequential

design process. Thus, CE provides a capability to manufacture a high quality product with low development cost and a quick response to the market.

In CE, design is crucial to the success or failure of the project. The product can be manufactured in time with low cost and high quality by having an efficient design. Design influences large proportion of the product cost and quality than the manufacturing process [5,10]. The right design can reduce chances of making a mistake in the implementation phase because right materials, specifications, machines, and manufacturing processes are indicated in the design. Moreover, production planning and control tools such as inventory control, resource allocation, and job scheduling originate from the design stage itself. This clearly implies that failure in design usually generates extra costs. These costs are incurred from materials used, labor, time, resource reallocation, rescheduling and rework. Additionally, poor design generates unaccountable costs in the later stages of the product life cycle such as poor products, extra time, cost in manufacture and assembly, loss of customers and opportunities, etc. Thus, the design process must be monitored and controlled to allow only a small chance of failure.

Design process can be represented graphically by using an IDEF3 model [7-9]. An IDEF3 model is used to determine sequences of tasks and express their relationships into units of behavior (UOBs), links, and junctions. Graphical representation of a process model allows formulation of a method for measurement of the overall system performance. Initially, the performance of an individual task is calculated by its success probability. A success probability of a particular task, influenced by risk factor(s) in that task, is the probability that an individual task will be performed successfully as planned. Consequently, overall success probability is determined by calculating the effect of individual success probabilities in their relevant paths. However, a high success probability level of a system requires a systematic risk mitigation program. Mitigation programs, e.g. repairs, insurance, regular maintenance works, spare parts inventory control, and employee training, generates additional maintenance costs. Thus, investment in mitigation programs to enhance system success must be rational compared to the potential damages incurred when the system is found not to be reliable.

2 BACKGROUND

IDEF3 Process Description Capture Method is a graphical model used to represent a process definition [8,9]. Sequences of activities and relationships are formulated into UOBs, links, and junctions. A UOB represents a task, an activity, or a function in a process. A simple box refers to a UOB in an IDEF3 process model. Links are relationships between UOBs. Links, represented by arrows, connect UOBs together in processes. Junctions are logical branches and refer to multiple paths in the process, some of which are optional. *"Exclusive OR"*, *"OR"* and *"AND"* junctions can be represented by a small box containing X, O or &. IDEF3 has a capability to model design processes because of its flexibility in representing alternatives in design [7]. An example of a simple IDEF3 model is illustrated in Figure 1. There are 8 processes represented as UOBs starting from UOB 1 and finishing at UOB 8. Decision are made at fan-out junctions *a* and *b*. An *OR* junction (O) indicates an alternative path between UOB 2 or UOB 3. At this junction, either UOB 2 or UOB 3 or both can be performed. An *exclusive OR* junction (X) shows an alternative between the path of UOBs 5 and 6 and the path of UOB 7. Only one of these paths can be selected.
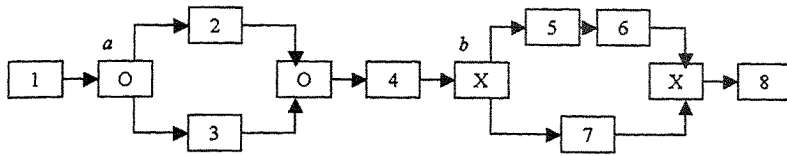
Figure 1: An example of a simple IDEF3 model.

3 SUCCESSS PROBABILITY OF A TASK

Success probability of a task is the probability that a particular task will be performed as planned. This success probability depends on behavior of risk factor(s) in that particular task. Risk factors are main sources of risk and cause failures in a particular task. Examples of risk factors are physical risk, resource risk, network risk, financial risk, organizational risk, etc. [4,7]. Ideally risk assessment should be conducted to estimate the success probability of an individual UOB. Expertise, experience, and attitude of assessors are very important in estimating the success probability. The success probability of a UOB can be generated from quantitative and/or qualitative assessments. Quantitative assessment refers to theoretical and statistical analysis based on historical data. Qualitative assessment refers to subjective assessments such as ranking, comparison, and quantitative measures of opinions. Some examples of subjective probability assessment can also be found in reference [2]. Both quantitative and qualitative assessments can quantify influence of risk factors in a particular UOB. Thus, the success probability of a particular UOB can be calculated from the equation:

$$R_i = f(x_1, x_2, x_3, , x_k) \tag{1}$$

Where, R_i is the success probability of a UOB$_i$ and x_k is the risk factor k of the UOB$_i$.

System reliability calculation is used to measure the overall performance of a process based on the success probability of UOBs in its path set. Individual success probability of a UOB is treated as reliability of a UOB. Thus, the overall system reliability is referred to as the overall success probability of a system. In an IDEF3 model, success probabilities are calculated differently at &, X, and O junctions [8]. At an *AND* (&) junction, overall success probability can be calculated by multiplying all success probabilities of UOBs connected to the & junction as illustrated in the equation:

$$R_{total} = R_1 \times R_2 \times \ldots \times R_n \tag{2}$$

Where, R_{total} is the total success probability of UOBs at the *AND* (&) junction, R_i is the success probability of a particular UOB$_i$ and n is the total number of UOBs at the junction under consideration.

For an *exclusive OR* (X) junction, the probability of only the selected UOB must be considered. Thus, the total success probability at junction X is calculated by the equation:

$$R_{total} = (P_1 \times R_1) + (P_2 \times R_2) + \ldots + (P_n \times R_n)$$

$$\text{Such that } \Sigma P_j = 1 \text{ for all } j \tag{3}$$

Where, P_j is the conditional probability of a particular UOB chosen at X junction and where no other UOBs are performed while j refers to all possible alternatives at an X junction.

In the overall system, a X junction implicitly means that only one UOB is actually chosen, thus the success probability of this junction comes from the success probability of the chosen

UOB multiplied by the probability that a particular UOB is chosen. Success probability calculation of an OR (O) junction is very complicated because there are 2^{n-1} combinations of UOBs that can be selected at this junction [8]. The probability that a particular UOB is chosen at junction O is independent and not conditional. For simplification purpose, only 2 UOBs are considered at junction O in this example. Success probability is calculated by union sets of both success probabilities [3] as in the equation:

$$R\ total = R_1 + R_2 - (R_1 \times R_2) \tag{4}$$

It is more complex to calculate system reliability if the number of UOBs and junctions increase. However, overall system reliability can be determined by calculating reliability in all possible paths [6,8]. A path set method was developed to find the overall system reliability. All possible paths were identified and the overall system reliability was calculated with relevance to the probability that a particular path is taken.

3.1 The example
In Figure 1, suppose all individual UOBs have a success probability of 80 %. Further, that the path through UOBs 5 and 6 has a probability of being chosen of 60% (Pr(5,6)) and the path through UOB 7 has a probability of being chosen of 40% (Pr(7)). Thus, the overall system success probability is calculated by Equations (2), (3), and (4) as follows

$$R_{23} = R_2 + R_3 - (R_2 \times R_3) = 0.96; \qquad R_{567} = 0.6 \times R_5 \times R_6 + 0.4 \times R_7 = 0.704$$
$$R_{total} = R_1 \times R_{23} \times R_4 \times R_{567} \times R_8 = 0.346$$

The overall success probability from direct calculation is 34.6%. However, when simulation runs were conducted for 10 replications with 100 runs in each replication. The result provided the average success probability for overall system at 36.3%. Moreover, 90 % confidence interval of the total system success probability is between 33.33% and 39.27%. Summary of this simulation is depicted as in Table 1.

Table 1: Summary of simulation analysis of the IDEF3 model in Figure 1.

Run	R_1	R_2	R_3	R_4	R_5	R_6	Pr(5,6)	R_7	Pr(7)	R_8	R_{total}
1	0.84	0.79	0.79	0.76	0.79	0.77	0.62	0.84	0.38	0.83	0.31
2	0.83	0.78	0.84	0.82	0.82	0.78	0.56	0.74	0.44	0.83	0.35
3	0.82	0.80	0.77	0.82	0.77	0.79	0.51	0.73	0.49	0.88	0.38
4	0.77	0.87	0.79	0.83	0.82	0.84	0.63	0.84	0.37	0.84	0.41
5	0.81	0.85	0.79	0.81	0.73	0.77	0.53	0.79	0.47	0.77	0.33
6	0.79	0.79	0.78	0.87	0.85	0.80	0.50	0.79	0.50	0.83	0.45
7	0.78	0.80	0.80	0.82	0.81	0.81	0.63	0.81	0.37	0.78	0.31
8	0.78	0.77	0.81	0.81	0.84	0.73	0.58	0.79	0.42	0.82	0.33
9	0.77	0.77	0.79	0.79	0.75	0.76	0.64	0.82	0.36	0.77	0.33
10	0.81	0.83	0.84	0.81	0.82	0.80	0.62	0.80	0.38	0.84	0.43
Mean	0.80	0.81	0.80	0 81	0.80	0.79	0.58	0.80	0.42	0.82	0.363
Stdv	0.025	0.034	0.024	0.028	0.039	0.030	0.054	0.037	0.054	0.035	0.051

3.2 Mitigation cost
Success probability indicates the performance of a system. A high success probability means that system probably performs successfully without failures or disruptions. However, in order to improve the success probability in design, additional cost must be invested because success probability in design process requires extra programs to be implemented in order to achieve a higher level of system success probability. The design process can be maintained at a high

success probability level by carrying out tasks that affect UOBs such as testing, resource outsourcing, training, regular monitoring, equipment maintenance, raw material inventory, documentation, etc. Generally if these actions are conducted regularly unexpected disruptions in the system are prevented. In this paper, mitigation cost describes the investment that intends to improve success probability of a system. Mitigation cost is an equal amount invested on a periodic basis during the project life. For example, resource outsourcing cost or maintenance cost per week, month, or year.

3.3 Damage cost

When system fails, the process can no longer perform and the system has to be revised until it is functional again. System failure generates unexpected consequences and damages to the project. Costs in design failure usually comes from materials and labor used, process reworks, machine's damage, loss of opportunity, reschedule, resource reallocation, set up and time necessary for a new set up and potential loss of customers. Damage cost is the total sum of money incurred when project fails. Thus, this total damage cost will be incurred in the project at mean time between failures in the project life. Generally, mitigation and damage costs are inversely proportional to each other. High investment in mitigation cost may prevent chances of a system failure and reduce possible damage from failures, while high damage cost is generated from systems with a low success probability. Thus, a proper investment decision has to be made in order to achieve a reasonable cost for a successful project completion.

4 INVESTMENT ANALYSIS

The mitigation program in only a particular UOB may be enough to improve the overall system success probability. Thus, the overall system success probability can be enhanced with reasonable costs. Initially, a simulation analysis is conducted to run scenarios from all possible alternatives in design after the mitigation have been taken into account in some UOBs. Simulation results can be used for making a decision by a project manager. In this paper, the traditional annual worth method [11] is used as an investment analysis tool.

Consider the cash flow diagrams of a project shown in Figure 2. The project has an initial investment P at the beginning of the project life. During the project life, mitigation cost (Rc) is incurred equally in each period until the end of project life n. When, an expected total damage cost (TDc) is incurred, the system fails. Thus, the time between two total damage costs is referred to as the mean time between failures (MTBF). For the investment evaluation, the interest rate per week is i % in each period. Figure 2(a) represents a system without sufficient mitigation plans where, system success probability is maintained at a low level. Thus, mitigation cost (Rc$_1$) of a system is low. The system may have frequent disruptions and may experience damage losses when it fails. Mean time between failures (MTBF$_1$) is the average time between two failures encountered in a system. When failure is encountered, repair and set up are required and total damage costs (TDc$_1$) are incurred. Figure 2(b) illustrates the project with a regular mitigation program. A higher amount of mitigation cost is invested to maintain the overall system success probability at a higher level (Rc$_2$ > Rc$_1$). Consequently, higher success probability results in longer mean time between failures (MTBF$_2$ > MTBF$_1$). Since discontinuity in system is mitigated, the total damage cost from failures encountered in the system occurs at a less frequent rate. Thus, the investment in reliability (Rc$_2$) saves expenses from potential damage when system does fail. The cash flow diagram 2(c) represents the difference between cash flow diagrams 2(a) and 2(b). The traditional annual worth method is applied in investment analysis to this scenario. A sinking

fund factor $(A/F, i, n)$ is used to change the total damage cost in to equal expense per period [11]. Thus, total damage costs (TDc_1, TDc_2) are changed into equal money in each period by multiplying $(A/F, i, MTBF_1)$ and $(A/F, i, MTBF_2)$ respectively as depicted in Figure 2 (d). Consequently, the equation below is used to evaluate the success probability investment.

$$TDc_1 (A/F, i, MTBF_1) = TDc_2 (A/F, i, MTBF_2) + Rc_2 - Rc_1 \qquad (5)$$
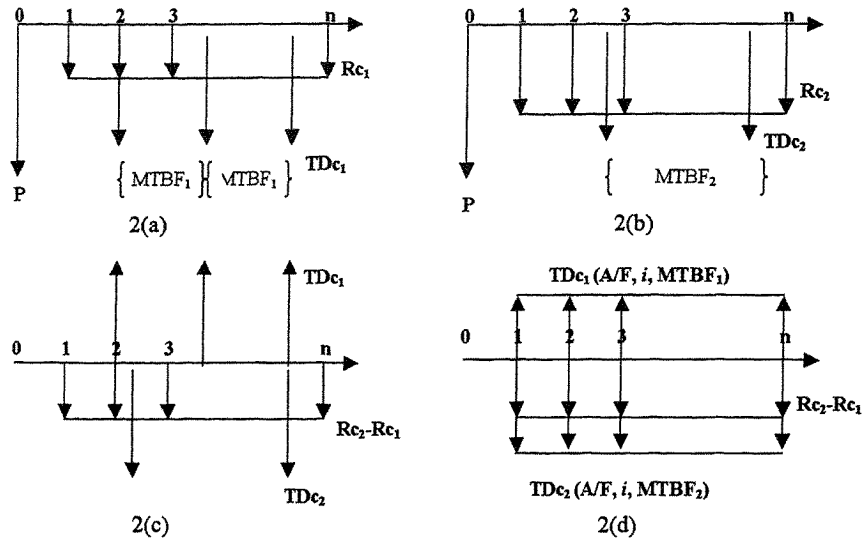
Figure 2. Cash flow diagrams for a typical project

The left hand side of Equation (5) refers to the benefits from mitigation investment while the right hand side of the equation represents cost after success probability improvement. In order to gain benefits from mitigation investment, the benefit from success probability improvement must be greater than cost incurred after improvement. $TDc_1 (A/F, i, MTBF_1)$ refers to the savings from success probability improvement, while $TDc_2 (A/F, i, MTBF_2)$ refers to the damage cost after improvement. $Rc_2 - Rc_1$ represents the additional amount invested in system success probability improvement. Hence, Equation (5) can be rewritten into a new form as shown in the equation:

$$TDc_1 (A/F, i, MTBF_1) - TDc_2 (A/F, i, MTBF_2) = Rc_2 - Rc_1 \qquad (6)$$

From Equation (6), the additional investment in success probability $(Rc_2 - Rc_1)$ must be less than the difference of expected total damage costs in order to justify the investment. Moreover, the Equation (6) can also be applied to reliability investment evaluation in traditional system reliability problems e.g. failures in machine breakdowns.

5 SIMULATION RESULTS

In the example discussed in section 3.1, UOBs have an individual reliability of 80%. The overall system success probability was calculated as 36 3%, as illustrated in the simulation analysis Suppose an individual UOB generates 1,000 monetary units every time it fails. An expected damage cost (Dc) in individual UOB is then equal to $1,000 \times (1-R_i)$ where R_i is the success probability of UOB i (Table 1). Thus, the total damage cost of the existing system

can be calculated from the simulation data as shown in Table 1 and depicted in Table 2. The total damage cost of the overall system calculated by this method is found to be 1,570 monetary units.

Table 2. Damage costs for the overall system

Run	UOB1	UOB2	UOB3	UOB4	UOB5	UOB6	UOB7	UOB8	Total
R_i	0.80	0.81	0.80	0.81	0.80	0.79	0.80	0.82	0.363
$1-R_i$	0.20	0.19	0.20	0.19	0.20	0.21	0.20	0.18	0.637
Dc	200	190	200	190	200	210	200	180	1570

However, from Figure 1, success probability improvement in UOBs 1, 4, and 8 can directly improve the overall system success probability. If investment in maintenance program at UOBs 1, 4 and 8 had been made to improved success probability to 90%, a direct calculation can be carried out as following:

$$R_{total} = R_1 \times R_{23} \times R_4 \times R_{567} \times R_8 = 0.4927$$

Table 3. Simulation after improvement in UOBs 1, 4, and 8

Run	R1	R2	R3	R4	R5	R6	Pr(5,6)	R7	Pr(7)	R8	R_{total}
1	0.87	0.87	0.76	0.94	0.74	0.78	0.59	0.76	0.41	0.87	0.46
2	0.89	0.80	0.83	0.90	0.76	0.80	0.60	0.76	0.40	0.90	0.44
3	0.91	0.83	0.73	0.88	0.80	0.76	0.67	0.76	0.33	0.89	0.49
4	0.87	0.79	0.73	0.95	0.83	0.86	0.55	0.80	0.45	0.85	0.50
5	0.96	0.83	0.83	0.92	0.77	0.79	0.55	0.79	0.45	0.89	0.49
6	0.91	0.78	0.75	0.93	0.81	0.84	0.62	0.75	0.38	0.92	0.53
7	0.92	0.85	0.79	0.94	0.81	0.79	0.60	0.81	0.40	0.87	0.54
8	0.90	0.86	0.78	0.86	0.84	0.83	0.58	0.75	0.42	0.89	0.53
9	0.91	0.79	0.82	0.92	0.80	0.73	0.60	0.80	0.40	0.89	0.44
10	0.91	0.82	0.83	0.85	0.79	0.84	0.60	0.81	0.40	0.95	0.58
Mean	0.91	0.82	0.79	0.91	0.80	0.80	0.60	0.78	0.40	0.89	0.500
Stdv	0.026	0.032	0.041	0.035	0.031	0.040	0.034	0.025	0.034	0.028	0.046

Table 4. Damage cost summary after improvement

Run	UOB1	UOB2	UOB3	UOB4	UOB5	UOB6	UOB7	UOB8	Total
Ri	0.91	0.82	0.79	0.91	0.80	0.80	0.78	0.89	0.500
1-Ri	0.09	0.18	0.21	0.09	0.20	0.20	0.22	0.11	0.500
Dc	90	180	210	90	200	200	220	110	1300

The simulation analysis is carried out to find the overall system success probability and expected damage cost in Tables 3 and 4 illustrates the simulation results from 10 replications with 100 runs in each replication. The overall system success probability then increases to 50% with 90% confidence interval from 47.35% to 52.65%. The total damage cost after improvement was found to be 1,300 monetary units. Suppose the mean time between failures before improvement (MTBF1) equals to 5 weeks and the mean time between failures after improvement (MTBF 2) equals to 10 weeks and the interest rate (i) is 0.25% per week. From equation 6, the investment can be estimated as:

Rc$_2$ -Rc$_1$ = 1,570 (A/F, 0.25%, 5) – 1,300 (A/F, 0.25%, 10) = 183.86

From this investment calculation, it can be seen that in order to improve the overall system success probability with a reasonable cost, an additional investment in success probability must be less than 184 monetary units per week.

6 CONCLUSION

In this paper, IDEF3 model was used to represent the design process. Then, reliability calculation was performed based on IDEF3/UOBs to evaluate the overall performance of the design process. It is generally realized that a mitigation program must be implemented in the design process in order to maintain the system performance without unexpected failures. However, improvement in system success probability (a mitigation program) incurs extra expenses to the project. Project evaluations based on the annual worth method was conducted to compare additional cost from success probability improvements to the expected savings from prevention of unexpected failures. The type of investment analysis, as outlined in this paper is expected provide project managers with options for different scenarios and help them identify the best investment plan suited to their needs for reasonable amount of acceptable risk in the project under consideration.

7 REFERENCES

1   **Caillaud, E., Gourc, D., Garcia, L. A., Crossland, R., and McMahon, C.** (1999) A framework for a Knowledge-Based System for Risk Management in Concurrent Engineering. In *Concurrent Engineering: Research and Applications*. Vol. 7(3), pp. 257-267.

2   **Clemen, R. T.** (1996) Making Hard Decisions: An introduction to decision analysis. Duxbury Press.

3   **Devore, J. L.** (1995) Probability and Statistics for Engineering and the Sciences. Duxbury Press.

4   **Jaafari, A.** (2001) Management of risks, uncertainties and opportunities on projects: time for a fundamental shift. In *International Journal of Project Management*. Vol. 19, pp. 89-101.

5   **Jo, H. H., Parsaei, H. R., and Sullivan, W. G.** (1993) Principles of concurrent engineering. In *Concurrent Engineering: contemporary issues and modern design tools*, **Parsaei, H. R., and Sullivan, W. G.** London, Chapman & Hall.

6   **Kusiak, A. and Zakarian, A.** (1996) Risk Assessment of process Models. In *Computers & Industrial Engineering*. Vol. 30(4), pp. 599-610.

7   **Larson, N. and Kusiak, A.** (1996) Managing Design Processes: A Risk Assessment Approach. In *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*. Vol. 26(6), pp. 749-759.

8   **Larson, N. and Kusiak, A.** (1996) System reliability methods for analysis of process models. In *Journal of Integrated Computer-Aided Engineering*. Vol. 3(4), pp. 279-290.

9   **Mayer, R. J., Menzel, C. P., Painter, M. K., deWitte, P. S., Blinn, T., and Perakath, B.** (1995) Information Integration for Concurrent Engineering (IICE): IDEF3 Process Description Capture Method Report. College Station.

10   **Salomone, T. A.** (1995) What every engineer should know about concurrent engineering. New York, Marcel Dekker.

11   **Sullivan, W. G., Bontadelli, J. A., and Wicks, E. M.** (2000) Engineering Economy. Prentice Hall International Inc.

Discrete adaptive mesh based on behaviour constrain of dynamic particles for three-dimensional reconstruction[1]

W YANG, W HU, and Y XIONG
School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China

ABSTRACT We present an adaptive mesh model for 3D reconstruction from unorganized 3D data points. Firstly, we construct the convex hull of data points as an initial mesh model. Vertexes of mesh are defined as Newton particles. Then, a data point categorization mechanism is designed. In each mesh deformation step, elastic potential energy is computed between the selected data point and mesh vertex. Based on elastic potential constraint, mesh deforms towards the local concave boundary. The final mesh model is constructed by adapting initial mesh to data points iteratively.

1. INTRODUCTION

Researches on surface reconstruction can be classified as static, geometric techniques and dynamic, energy-based techniques. Typically, Alpha shape [7] presented a static method to infer shape from unstructured 3D data set. Physically based deformable model is a dynamic method to recover geometric and topological shape. J. Montagnat [5] summarized the basic method of deformable surfaces geometry and topology. Andrew Witkin [1] presented a method to capture of 3D shape by elastically deforming an initially object into the desired shape given as a set of unorganized 3D points. Jacques-Olivier Lachaud [4] presented deformable meshes with automated topology changes. Their model changes its topology according to the classical Eulerian topological transformations of creation, deletion or inversion. Non-Eulerian topological transformations of closed and oriented surfaces that evolve in the Euclidean space are described also. David Love Tonnesen [2] presented a dynamically particle systems to reconstruct 3D model. The individual data points of the

unknown shape interactively exert forces upon each other to fit and recover surface and volume object. The basic components of particle-based surface extension algorithm are two heuristic rules, stretching and growing rules that control the addition of new particles. J.V. Miller [6] proposed a polygon-based model. The behavior of the model is determined by a local cost function associated with each model vertex.

2. ALGORITHM OVERVIEW

Our algorithm starts at a convex hull of data points. Then the initial mesh model adapts to data points based on physically based behavior function. We perform an iterative procedure of deformations, subdivision and model adapting. Each step of mesh adapting to data points is constrained by elastic potential energy between mesh and selected data points. At each deformation step, visibility is checked.

The initial triangulated mesh T_0 is the convex hull of all data points. A behavior function is defined at each vertex of the mesh based on particle dynamics. At each iterative step, mesh T_i replace with a new mesh T_{i+1}, and the data points are categorized as free and non-active one. The free points are those which have not yet been processed. The non-active points are those which lie on the mesh. Because of elastic potential energy constraints, mesh will deform towards those selected free points. Mesh T_i is subdivided to mesh T_{i+1} when mesh deform. Some new points are created on current mesh T_i, these new points are defined as active points which will take part in further computation at next step. As T progressively deforms, the triangular mesh $T(t)$ interpolates increasing number of free data points. Then, these new free data points are change to non-active points. The mesh $T(t)$ will finally adapt itself to object boundary step by step. The main procedure of mesh adaptive algorithm is shown as figure 1.

Initial mesh T (convex hull)

Dynamic particle mesh T

Deformation and check

Iterative mesh

Non-active mesh mechanism and Surface

Refinement of coarse mesh
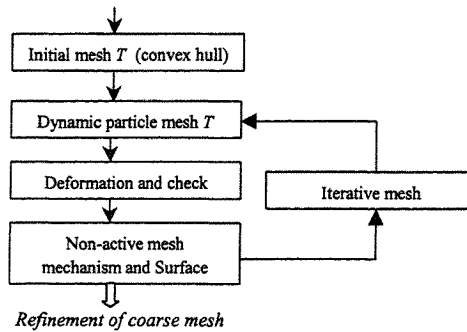
Figure 1. Overview of mesh adaptive algorithm

3 CONSTRAINT FUNTION

A behavior function says that two particles x_i and x_j should be in the same place is just $r = x_i$

$-x_j$. The elastic deformation potential function is defined as a function of separation distance r between a pair of particle,

$$\phi_{ij}(r) = \frac{k_s}{2}(r - r_0)^2 \tag{1}$$

where k_s is a spring constant. Since the force due to a scalar potential is minus the energy gradient, the force on particle x_i is

$$f_i = -\nabla_{x_i}\phi_{ij} = (-k_s r - k_d \dot{r})\frac{\partial r}{\partial x_i} \tag{2}$$

where f_i is the generalized spring forces on particle i, which attract the system to state that satisfy $r = 0$. r_0 is the rest length, k_d is a generalized damping constant, and $\dot{r}$ is the time derivative of r, is just $v_i - v_j$, the difference between the two particles' velocities.

When two particles are in equilibrium, the potential energy between them is minimal. The distance between two particles when in equilibrium is known as the equilibrium separation distance, r_0. If instead we want particle i and j to be distance r_0 apart, then a suitable behavior function is $r = |x_i - x_j| - r_0$. We assume that the sampling points of S do not leave unsampled holes of radius larger than ρ, and unknown surface S is a closed surface. We assume sample noisy scalar value is $\delta = 0$. Then we set equilibrium separation distance $r_0 = \rho - \delta$.

A Newton particle system is governed by the set of ordinary differential equations of motion

$$m_i \ddot{x}_i + \gamma_i \dot{x}_i + f_i^{int} = f_i^{ext}, \quad i = 1, \dots, N \tag{3}$$

where f_i^{int} is the sum of inter-particle forces and f_i^{ext} is the sum of external forces acting on the particle. When a deformable mesh T is subjected to external forces, it reaches a state of equilibrium when the total energy of the system is at a minimum. External forces are functions of single particle state and external state variables $f_i^{ext}(x_i, Q)$, where Q is a set of external state variables. The external forces can be defined as points in space attracting the surface, or boundary limit.

For vertex on mesh model T_i, inter-particle force f_i^{int} is limited acting between direct neighbor of particle i. The inter-particle elastic interaction force tend to zero at the searching radius boundary r_b. Set $r_b = \mu r_0$, by increasing scale value $\mu > 1$, $\mu \in R$, small feature is ignored in shape recovering. To meet these two conditions, we designed the following weighting function,

$$w(r) = \begin{cases} 1 & \text{if } r < r_b \\ 0 & \text{if } r > r_b \end{cases} \tag{4}$$

Then, we can rewrite inter-particle force equation (2) as

$$f_i^{int} = -w(r_{ij})\nabla_{x_i}\phi_i = -w(r_{ij})\sum_{j \in N_i}^{N} \nabla_{x_i}\phi_{ij} \tag{5}$$

where, $r_{ij}=\|r_{ij}\|=\|x_j - x_i\|$ is the distance between particles i and j. Neighboring particles N_i is defined as pairs of particles that are directly connected in mesh T_i.

The inter-particle force will be zero when length of the edge is large than r_b. When forces in a piece of the mesh facet disappear, vertexes in the mesh facet are not connected. These mesh facets will be subdivided by adding midpoint at each edge of the facet. They are set as active facets, and those newly created vertexes are set as active points.

For each new added vertexes x_i, external force $f_i^{ext}(x_i,Q)$ act on the selected free point x_j.

The external state variable Q in $f_i^{ext}(x_i,Q)$ is defined as elastic force between active points x_i and new selected free point x_j. The external force between active point and free data point is computed by equation (2), and it is tend to zero at equilibrium separation distance r_0. The internal force between active points is set as zero. The deformation potential function $\phi_{ij}(r)$ offers the mechanism to deflate the mesh. The active vertex will move along the direction of the lowest local potential energy. For each active vertex, the algorithm finds out the candidate free data point and calculates the distance related to the vertex. If the distance is smaller than the threshold r_b, the vertex will be marked as non-active and is no longer allowed to move.

4 VISIBILITY CRITERION

Spatial interaction forces may act on new added vertexes x_i. That is, the potential energy ϕ_i of particle i with respect to a system of N_i particles are given by,

$$\phi_i = \sum_{j \in N_i}^{N} \phi_{ij} \qquad (6)$$

where ϕ_{ij} is the potential energy between particles i and j. One the one hand, a complication in large-scale spatial interaction simulations is that the force calculation is $O(n^2)$ in the number of particles. If the interactions are local, efficiency may be improved through the use of spatial buckets. On the other hand, spatial buckets can filter the points which do not lie in the object boundary. For local interaction forces, each active point interact with selected free data point located in the object boundary.

Since many surfaces are not convex, mesh T_0 in general only interpolates a subset of the data points. The search direction in mesh deformation is defined to normal direction of active facet which direct to inside of the convex hull. The effect is that a model point will migrate towards the concave boundary of the object enveloped by the initial mesh. Note that interaction forces between mesh vertexes are inter-particle forces. In active facets of mesh T_i, interaction forces between vertexes disappear, replace with interaction forces between active point and free data point. For an example of 2D data points, an active edge is set and illustrated as figure 2.

(a) data points   (b) convex hull of data points
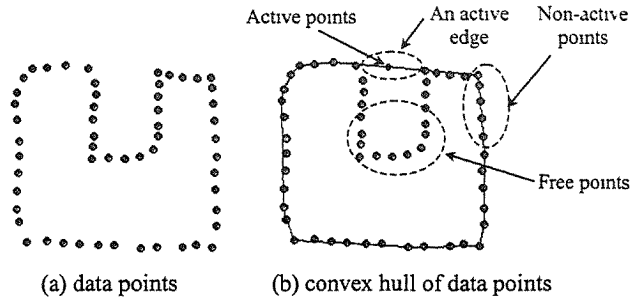
Figure 2. Categorization of data points and the selected active edge

In general, an active mesh region includes a number of triangular facets. The normal tracking method is used in computing external forces $f_i^{ext}(x_i)$. Each vertex is attracted to a point located in the vicinity of normal direction of the triangular facet, as defined by Hugues Hoppe[3]. The step length range of deformation is set as sampling rate ρ. In deformation procedure, a visibility criterion is checked in each deformation step. When an active point x_i move to a new position x'_i, a convex hull is computed which enclose point x'_i and all points in current active mesh region. If there are no free points fall into the convex hull, then original mesh facets are replaced by convex hull polyhedron. Otherwise, the step length range ρ is set as half of itself, and the convex hull is recomputed. Here a depth first search method is used. Here, the convex hull polyhedron is a visibility cone at x'_i. Point x_j and x'_i point can constitute the vector $v_j = x_j - x'_i, (j = 1,2,\cdots,m)$. The vector set $V = \{v_i | i = 1,2,\cdots,m\}$ can form a convex cone Λ,

$$\Lambda = \{\sum_{i=1}^{m} a_i v_i | a_i \in R, a_i > 0, i = 1,2,\cdots,m\} \qquad (7)$$

The detail of visibility computation can refer to our previously paper [9], and a visibility cone example in 3D shown as figure 3 (b).

(a)   (b)
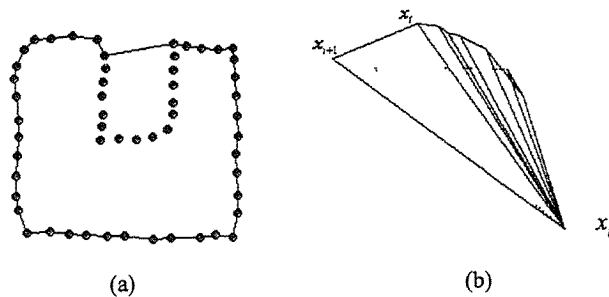
Figure 3. Computing visibility cone

5 ADAPTIVE MESH DEFORMATION

The model can be initialized with convex hull enveloping data points. After that, the mesh may then be evolved according to defined dynamic and geometrical constraints. The initialization model is closed convex polyhedron. From an unorganized set of points $X=\{x_1, ..., x_n\}$ sampled from some unknown surface U, initial mesh creates a convex hull $CH(X)$ that geometric approximates U. Figure 4 illustrate an initial convex hull of data points and a convex hull by add new point p.

We take an adaptive subdivision approach to active triangular facet that are deflating. For the initial data points P_0, P_1,..., P_n, a subdivision define a series of points P_j^k,

$$P_j^k = \sum_{i=0}^{n_k-1} \alpha_{i,j,k} P_i^{k-1} \tag{8}$$

Where, $k>0$, $\alpha_{i,j,k}$ is mask. The subdivision rule is a stationary scheme. The algorithm will introduce a new vertex at the middle position of each old edge, and connect all the three new vertices.

| (a) data points | (b)convex hull $CH(X)$ |

Figure 4 Computing convex hull of data points with increment method

The non-active particle mechanism is efficient, where lots of vertexes are initialization near their final position, and set as non-active ones. Note that only the potential $\phi_{ij}(r)$ of active vertex is computed at each iteration step. When the inter-particle forces reach a stable position, the 'energy' of the mesh is thus minimal. The equation (3) is integrated by the second order Runge-Kutta method. An adaptive step size is used when visibility criterion is not satisfied to insure mesh not getting across boundary of object. The step size can be reduced by dichotomy if the magnitude of the current step size results in an increase in the behavior function. If a step size is no longer able to reduce the cost of the vertex, then the vertex is marked as non-active and is not allowed to move any further. This heuristic rule is effective for mesh model deforming at concave area. The algorithm carrying out the evolution of the mesh can be summarized as an iteration of the pseudo-code procedure of Figure 5.

Procedure *Evolution(Mesh & T, const Points & P)*
    for each *Points P*
        compute convex hull $CH(P)$ of data points
        initial mesh $T \longleftarrow CH(P)$
    ListOfVertex L $\longleftarrow$ all vertexes of *T*
    for each *Points P* $\notin T$
        compute normal distance r_n between *P* and mesh facet,
        if $r_n > r_b$, point *P* set as a active particle
           ListOfActiveParticles $P_{act} \longleftarrow P$
    for each facet edge(*Vertex* V_i, *Vertex* V_j) **of mesh** *T*
        compute $V_{ij}.f_{int}(P)$
      if $V_{ij}.f_{int}(T)=0$ & $r > r_b$
      subdividing facet, create new vertexes V_{new}
      update T
      push V_{new} *in StackOfSubdivision SS*
    Boolean b $\longleftarrow$ **false**
    repeat
      while SS.*isNotEmpty()* **do**
          for active mesh facet
        calling dynamic particle movement sub-procedure
        and checking visibility criteria
          application of the Newtonian law of constraint on
          new added particle with the previously computed
          t, f_{int}, f_{ext}; make movement of mesh facet.
      if *equilibrium criteria is satisfied*
        seting particles in equilibrium separation distance as
        non-active particles.
      else
        calling subdivision sub-procedure
        for new added *P*
          ListOfActiveParticles $P_{act} \longleftarrow P$
        update T
      if P_{act} **is null,** **b** $\longleftarrow$ *true*
    until b
    return *T*
end

Figure 5. This procedure describes the main steps of iteration of deformation.

6 CONCLUSION

By extended physically-based mesh deformation combined with visibility check procedure,

we propose a schema of adaptive mesh for recovering shape of 3D unstructured data points. The mesh initially is constructed from convex hull of all data points. By iterative performing subdivision and deflating deformation, the mesh adapt to boundary of the object. The mesh adapt model will improve the efficiency of physically-based mesh deformation method.

REFERENCES
1 Witkin, A., Terzopoulos, D. and Kass, M., Constraints on deformable models: Recovering 3d shape and non-rigid motion. Artificial Intelligence, pages 91-123, 1988.
2 David, L. T., Dynamically coupled particle systems for geometric modeling, reconstruction, and animation, (1998) Ph.D thesis, University of Toronto.
3 Hoppe, H., Automatic reconstruction of surface models from scanned 3D points. PhD thesis, University of Washington .
4 Lachaud, Jacques-Olivier and Montanvert, A., Deformable meshes with automated topology changes for coarse-to-fine three-dimensional surface extraction, Medical Image Analysis (1998) volume 3, number 2, pp 187-207
5 Montagnat, J., Delingette, H., Ayache, N., A review of deformable surfaces: topology, geometry and deformation, Image and Vision Computing 19 (2001) 1023-1040
6 Miller, J.V., Breen, D.E., Lorensen, W.E., O'Bara, R.M., and Wozny, M.J., Geometric deformed models: a method for extracting closed geometric models from volume data. Computer Graphics (SIGGRAPH'91 Proceedings), pages 217-226, July 1991.
7 Edelsbrunner, H. and Mücke, Ernst P, Three-dimensional Alpha Shapes, ACM Transactions on Graphics, 13(1):43-72, 1994.
8 Duan, Ye and Qin, Hong, Intelligent Balloon: A Subdivision-Based Deformable Model For Surface Reconstruction Of Arbitrary Topology, Proceedings sixth ACM symposium on Solid modeling and applications, 2001, Ann Arbor, Michigan, United States, pp47-58
9 Yang, W., Ding, H. and Xiong, Y., Manufacturability Analysis for a Sculptured Surface Using Visibility Cone Computation, Int J Adv Manuf Technol (1999) 15:317–321

Development of CAD/CAM environment for one-of-a-kind production

K LAPPALAINEN
Department of Mechanical Engineering, University of Oulu, Finland

ABSTRACT

This paper deals with machining of parts in one-of-a-kind production. The manufacturing of prototypes, molds and fixtures, and scientific instruments are typical examples of one-of-a-kind production. Research is being done in development of an integrated CAD/CAM environment with databases for the one-of-a-kind production environment. Development of this kind of system depends on the manufacturing environments, product technologies and products. For example parts for space instruments are very complicated and difficult to manufacture, because each part needs its own jigs, and possibly numerous setups. Manufacturing tasks such as process planning, fixture planning, sequencing, and NC-programming have to be flexible and integrated. In CAM systems, tool paths of machining operations have to be verified before machining in order to efficiently use the machine tools.

1 INTRODUCTION

Today's competitive manufacturing environment requires a shortened product development time. One-of-a-kind or make-to-order production is getting more important, because product life cycles are shortening and prototypes must be made quicker. Also mass customization of products is increasing. Fast reconfiguration of processes and dynamic production layout are needed. For example as a result of the product process a competitive product and its production facilities are developed on a very short lead time. Although great emphasis has been placed on concurrent engineering, it is still a challenge for engineers to bridge the gap between design and manufacturing.

The one-of-a-kind production environment is composed of many simultaneous projects, and the manufacturing environment is tailored to the needs of the customer. Today, the main goal of most production systems is to fulfill the individual demands of each customer, with customers directly influencing manufacturing throughout the entire production process. It is also important that flexibility and cooperation, especially in the transfer of information, exist between design and manufacturing [14].

According to [5], characteristics of one-of-a-kind production are:
- Customer oriented production with simultaneous redesign engineering
- Incomplete relevant manufacturing information when order is first submitted
- Just-in-time (JIT) generation of product information
- Reconfigurable production equipment
- Qualification and human interaction

As in traditional batch manufacturing the goal is to increase the overall equipment effectiveness (availability * speed * quality), in one-of-a-kind manufacturing most important things are the flexibility and management of manufacturing environment. The main tasks of

the manufacturing process are process planning, fixture planning, sequencing, NC programming/simulation and tool path verification. All these must be integrated parallel into the same system. CAM systems must be very flexible in those tasks. The primary challenge is that CAM systems must be flexible to meet the demands of different projects, as well as to allow the possibility for modifications during manufacturing. These issues are very common in prototype manufacturing. The second major challenge is the integration of CAM systems into manufacturing facilities; especially the databases of CAM systems should be similar to those on the shop floor. Data generated in CAM systems must be reliable, when transferred to production. All these issues allow the "out of process" design to be of high quality. The final goals are to speed up manufacturing by raising the level of automation and efficiency of manufacturing.

2   PARTS FABRICATION IN ONE-OF-A-KIND PRODUCTION

Manufacturing systems can be classified according to production volume and variety, as seen in Fig. 1. Today's most up-to-date manufacturing systems are flexible transfer lines (FTL), flexible manufacturing systems (FMS) and flexible manufacturing cells (FMC). In one-of-a-kind (OKP) production parts are fabricated individually, or in very small batches. Thus FMSs and FMCs can also be used, as well as NC and conventional machine tools because the goal of FMSs and FMCs is to achieve the flexibility of a job shop and the efficiency of a flow line.[7]
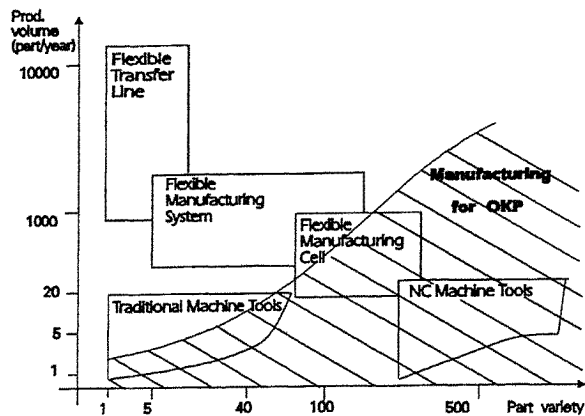
Figure 1. Manufacturing systems for one-of-a-kind production [7]

Typical to one-of-a-kind fabrication are constant setups and special fixtures. The problems are fixture planning, process planning, selection of tools and machining data. The integration of manufacturing knowledge to design process by using libraries of machine tools, cutting tools, fixtures, machining data and parts containing the manufacturing features is the target in development of machining environment. These features and accumulated know-how can be used in process planning and to reduce the redesign efforts and the lead time and to increase the quality.

3   RESEARCH ENVIRONMENT

During the last ten years the University of Oulu has taken part in several international space projects. In the Cluster/EFW-instrument (Fig. 2) project mechanical design, manufacturing and testing have been the main responsibilities [8]. The project has been carried out in cooperation with the Space Sciences Laboratory at UC Berkeley in U.S.A and Institute of Space Physics in Uppsala in Sweden [4]. In Aspera3/NPD-instrument project [9] University of Oulu was one partner in parts manufacturing. Manufacturing of space instruments is a typical example of one-of-a-kind production. The level of automation has normally been quite low. Examples of current available manufacturing equipment to be used in this kind of manufacturing are NC-machine tools (NC-level) and even flexible manufacturing unit (FMU-level) [7]. This type of flexible manufacturing equipment is also used in the prototype stage of mass production. Research has been done during the manufacturing projects. Current research is aimed at telecommunication industry. The goal is to develop operations between companies to produce prototypes and molds. The main objective is to shorten lead-time from order to delivery. Shorter manufacturing time can be reached by higher automation level.
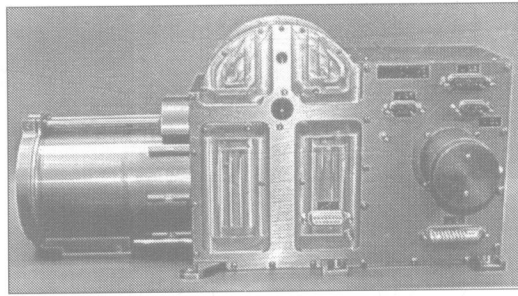
Figure 2. EFW (*Electric Fields and Waves*) instrument

Parts of EFW instruments include both prismatic and rotational parts, whose shapes and dimensions may vary widely. Examples of part shapes are plates, cubes, long parts; rotational parts may be discs, rotational shells, cups, shafts, rings, and sleeves. Examples of fixturing methods are special fixtures, technique of micro fixturing, fixtures for end-of-bar machining, and vacuum clamping. Standard fixtures such as machine vices, modular jig and fixture systems, and chucks are mainly in the primary machining operations (roughing).

4   FIXTURE CONSTRUCTION AND PROCESS PLANNING

The structure of part fixtures has a significant influence on the sequencing of operations and NC programming. There are three basic issues that must be taken into account and optimized. These are machining time, number of setups, and quality of the workpiece. These three issues must be integrated in parallel, to ensure maximum cohesion, and they must allow for smooth and flexible modifications.

The objectives of fixture design, according to [12] are:
• Determination of clamping, positioning and (if necessary) supporting features

- Selection of fixturing elements
- Calculation of the number of fixturing elements, and their positions and orientations relative to the workpiece
- Visualization of the fixturing layout
- Documentation of the resulting plan of the shop floor

An example of a fixture setup is shown in Fig. 3. Fixture setups are typically assembled from several fixture elements. Standard fixtures like vices, clamps, pins, and screws are installed on the mounting plate. Features for positioning supporting and clamping elements are designed to fit on this mounting plate where they are needed. Fixtures are joined to the surface of a pallet or table of machine tools using a standardized contact surface. This allows the location of fixture elements to be specified, which is used in simulation and production. The basic objectives of manufacturing in EFW project was to separate "out of process" time from "in process" time, then minimize "out of process" time, and eliminate the need for adjustment.[10]
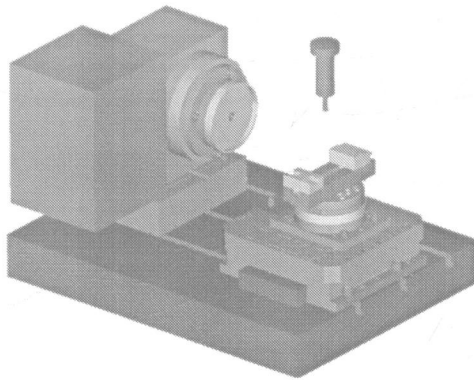
This method has many benefits. The fixture setup makes it possible to carry out reliable simulation and collision check. By using zero point clamping systems fixture accuracy is very good (typically 0.002 mm) and workpiece changes are fast and reliable. The fixture locations are always the same without any adjustments during production. The work coordinates can be calculated in CAM-system by using global origin.

There is much research being done in the development of CAPP-systems (Computer Aided Process Planning) e.g. [2], [5], [6], [11]. A CAPP system is in between CAD and CAM systems, speaking about CAD and CAM systems in a broader sense. Typical elements of CAPP are selection of machining methods, selection of cutting tools, sequencing of machining operations, path planning and generation of NC program, calculation of machine parameters, and design of jig and fixtures.

Fixture planning is a very important task, as it has a large effect on the sequencing of operations and NC programming. As well, many systems for automated or semi-automated fixture design (AFD) are developed for use in research [3], [11], [13], [15]. These systems are modules in CAPP systems or autonomous systems. Currently, systems typically are feature based. In one-of-a-kind part manufacturing, the emphasis is placed on process planning as a

continuous activity aimed at flexible use of the manufacturing equipment in both a technical and logistical sense [6]. The first requirement to reach the objectives is to make sure that the technical data is correct, and minor stoppages don't occur due to lack of manufacturing data. After this is it possible to make a reliable plan for production control. As well, an estimate of the machining time is needed from the part programming system. Quite often, problems occur in trying to optimize the individual part cost and the whole manufacturing process. However, it is extremely difficult to set up an entire CAPP system for example for space technology manufacturing for the following reasons:

- Every project has its own characteristics and demands
- Projects are usually quite short, and development during a short period is difficult
- Human resources are usually limited, and may sometimes be inadequate
- The batch size is very small and unrepeated
- Parts can be very complicated
- Parts need complicated special design fixtures
- Operation sequencing is very crucial, and rules for this are very hard to make
- All manufacturing data is tested on the computer, but not by machine tools
- Very rarely is it possible to find a "pure" part family, geometrically speaking

5   DEVELOPMENT OF CAD/CAM ENVIRONMENT

Requirements of CAD/CAM systems for one-of-a-kind manufacturing are quite common [1]:
- The system must ensure that the user always uses the latest version of the part
- The data transferred between CAD and CAM modules must be absolutely free from errors
- The system is an integrated CAD/CAM environment, i.e. no translation of data into the system is required
- If a change is made to the design model, the tool path will automatically update itself
- A part should automatically be regenerated after modifications to it or any related parts are made, following the predefined internal relationships between the interdependent parts

The utilization of CAD/CAM system has a great influence on production times and efficiency. Using high quality fixture designs and off-line programming with simulation enables:
- Higher levels of automation
- Shorter manufacturing times
- Reduction of minor stoppages
- Higher quality products
- Better utilization of machine tools
- Shorter setting times
- To change some of the in-process time to out-of-process-time

Production equipment and manufacturing knowledge are integrated into the CAD/CAM system. The CAD/CAM system has databases for workpieces, machine tools, mounting plates, standard parts, design features, tools, and method files (Fig. 4) [10].
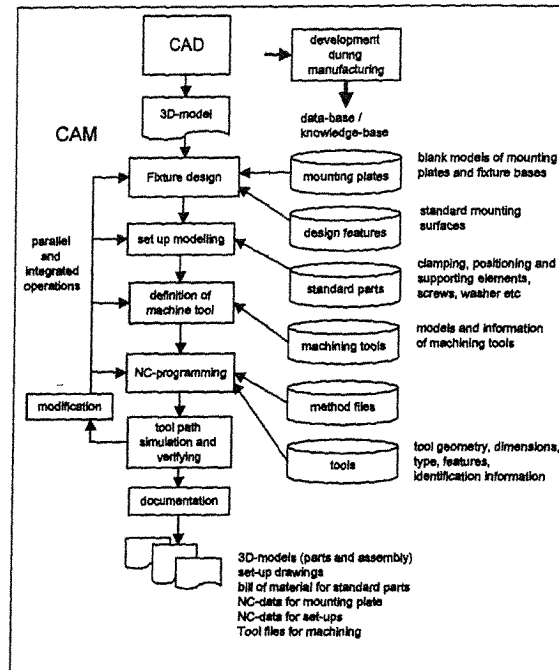
Figure 4. *The architecture of CAM-system functions and databases*

The first phase in fixture design is the selection of a mounting plate or fixture base. Blank models of mounting plates are in the library and one of them is transferred to the CAD/CAM system. The size of the mounting plate is standardized. Features, from the library containing design features, are attached to the contact surface of the mounting plate. On top of this, clamping and positioning elements are added using part geometry and features. By using a mounting plate with a standard join surface makes manufacturing easy and fast. In the design of the fixture layout, the workpiece is aligned and located on the mounting plate, and standard fixture elements are selected.[10]

Determination of the machine tools to be used involves the following functions: First, the 3D models of the machine tools and the pallet or table of machine tools is taken from the library, and the machining tool parameters are set. The mounting plate is adjusted according to the standard join surface, then the machine coordinate system and the coordinate system for NC programming is set. NC programming is done by using the entire 3D assembly model mentioned above. Programming is done directly from the user interface, or by using a method file for features or groups of features. In addition, optimal tool models are selected from the tool database. Integration of the CAD/CAM and tool system is done by the tool database, from which NC programming tools can be selected. The most essential pieces of information about tool models in NC programming are tool type, geometry, dimension, and assembly version. Tool databases also consist of further information, e.g. product number of tools, holders and chucks and information for setting up tool assembly. During NC programming, the movement of tools can be checked by using collision control. All four

44

functions (fixture design, layout modeling, machine tool specification, and NC-programming) have to be in parallel.[10]

A typical way to save manufacturing knowledge in a CAM system is to have a material library for machining parameters and macros for work cycle operations. A more sophisticated way is to use a method file, which is programmed for features or groups of features. A method file is a text-based macro file based on logical structures and variables. Method files consist of information for manufacturing features or group of features, e.g. work parameters, tool information, types of paths, exit and entry parameters, and other movement parameters. By using logical statements, variables, and mathematical operations, intelligent macros for NC programming can be created. The advantage of method files is that they are applicable to many kinds of features, of varying geometry. They may also be used for groups of features. Method files may be used for pocket and profile milling, and point group of points operations. It is also possible to bypass and modify the tool path parameters in the user interface.

During simulation, the tool path is checked, to make sure that the NC program is feasible and error-free. Typically, a single tool path is checked after every tool path generation, since incorrect tool paths can be regenerated immediately. After the individual operations belonging to specific setups are checked, their sequence is optimized. It is essential that the tool movements generated by the post processor are identical to the actual movements of the tools. Some other objectives of high-level simulations are:
- Simulation with whole fixture layout including pallet
- Modeling of tools with tool and chuck
- Collision calculations between tools and other elements
- Easily monitored working parameters and tool movements

6 CONCLUSION

One-of-a-kind production is increasing all the time when companies produce customer oriented products.The paper describes characteristics of manufacturing and its most essential tasks. It deals with the importance of fast and reliable manufacturing preparation, which is based on off-line programming and subsequent simulation. A CAD/CAM-system needs a database system, which consists manufacturing knowledge and is integrated with shop floor equipment. In one-of-a-kind manufacturing the issues mentioned above have great impact on utilization of machine tools, quality level, time from order to delivery and factory price. To survive in a competition CAD/CAM systems should be utilized effectively, machine tool and tooling as well as machining data libraries should be developed. The organization should be able to produce one-of-a-kind products almost as effectively as products in batch production. The objective will be to develop know-how so that flexible manufacturing systems could be used also for one-of-a-kind production.

REFERENCES

1   Chang J. W., Luh Y. L. & Chiou S. S. (1997) Integrated application in CAD/CAM, scheduling and control. Integrated Manufacturing System 8/6 1997. MCB University Press pp. 378-387.

2   **Deurwaader, J. M.** (1990) A new Concept for CAD-CAPP-CAM systems. 22th CIRP Int. Seminar on Manufacturing Systems. June 11-12, 1990 Enschede. Netherlands. pp. 7A.

3   **Dong, X., DeVries W. R. Wozny M. J.** (1991) Feature-Based reasoning in fixture design. Annals of the CIRP, Vol 40/1/1991 pp. 111-114.

4   **Gustafsson, G., R. Boström, B. Holback, G. Holmgren, A. Lundgren, K. Stasiewics, L. Åhlén, F. Mozer, D. Pankow, P. Harvey, P. Berg, R. Ulrich, A. Pedersen, R. Schmidt, A. Butler, A. Fransen, D. Klinge, M. Thomsen, C.-G. Fälthammar, P.-A. Lindqvist, S. Christenson, J. Holtet, B. Lybekk, T. Sten, P. Tanskanen, K. Lappalainen, and J. Wygant.** (1997) The electric field and wave experiment for the Cluster mission, Space Science Reviews, 79, pp. 137-156.

5   **Hämmerle E., Bochnick H., Hirsch B. E. & Opas J.** (1991) Knowledge based process planning for one-of-a-kind production. IFIZP Working Conference, Process Planning for Complex Machining with AI-Metdods.Germany, 27-29 th November 1991.

6   **Kals H. J. J., van Houten F. J. A. M. & Tiemersma J. J.** (1990) Cim in Small Batch Part Manufacturing, 22th CIRP Int. Seminar on Manufacturing Systems. p. 5a 1-18.

7   **Kovacs, G., Mezgar, I., Szelke, E. & Girnt, M.** (1992) One-of-a-Kind Production - A Concurrent Engineering Approach. 'One-of-a-kind' Production: New Approaches. Elsevier Science Publishers B. V. pp. 143-156.

8   **Lappalainen K.** Manufacturing of Spacecraft Instruments, Proceedings of the Int. Conf. on Manufacturing, ICM 2000 24-26 February, 2000, BUET, Dhaka, pp. 558-562.

9   **Lundin R.** (1999) Aspera 3, Analyzer of Space Plasmas and Energetic Atoms. Swedish Institute of Space Physics. Kiruna, Sweden. p. 40.

10   **Mettälä J., Lappalainen K.** (2002) Development of CAD/CAM Environment for Parts Manufacturing of Space Instruments. IDMME 2002 14-16 May, IFMA, Clermont-Ferrand, 10 p.

11   **Mäntylä M & Opas J.** (1993) Parametric Fixture Layout Models for Operative Process Planning. Helsinki University of Technology 22 p.

12   **Mäntylä M. & Shah.** (1995) Parametric and Feature-Based CAD/CAM. John Wiley & Sons, INC. New York. 620 p.

13   **Nee, A. Y. C. & Senthil Kumar, A.** (1991) A Framework for an object/rule-based automated fixture design systems. Annals of the CIRP. Vol 40, no 1/1991, pp. 147-151.

14   **Steffen, P. H.** (1992) FOCON A Communication Approach for 'One-of-a-kind' Production. 'One-of-a-kind' Production: New Approaches. Elsevier Science Publishers B. V. pp. 23-29.

15   **Wu Y., Rong Y., Ma W., Clair S.R.** (1998) Automated modular fixture planning: Accuracy, clamping, and accessibility analyses. Robotics and Computer-Integrated Manufacturing. 14(1998) pp. 17-26.

Exact G^1 continuity conditions for B-spline surfaces with applications for multiple surface fitting

W MA
Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong
N ZHAO
Currently with Trident Microsystems Inc , Sunnyvale, California, USA

SYNOPSIS

This paper presents an approach for smoothly connecting multiple B-spline surfaces with exact G^1 continuity using a finite number of check points along common boundaries. Sufficient and necessary G^1 continuity conditions for connecting two neighbouring B-spline surfaces are first studied. A sufficient condition that needs less number of check points using linear tangent blending is further developed. The selection of the linear tangent blending parameters in case of multiple surfaces is also discussed. Several examples are presented for connecting two neighbouring surfaces, three surfaces meeting at a common corner and multiple surfaces of a computer mouse with exact G^1 continuity conditions.

1. INTRODUCTION

Non-uniform rational B-splines (NURBS) are by far the most popular representation and the de-facto standard for computer aided design and manufacturing (CAD/CAM). Nevertheless, it is still awkward for practical modelling using multiple B-spline surfaces with continuity requirement. To our knowledge, most of the published work on continuity conditions focuses on Bézier surfaces, which are single patch B-splines, and little progress has been made on multiple B-spline surfaces with general topology.

Among the publications on B-spline surfaces with general topology, Milroy et al proposed a procedure for achieving approximate global G^1 continuity [4]. Apart from the weak continuity conditions obtained, this approach also needs to solve an expensive nonlinear optimization

problem. Ma et al improved this method in [3] and produced exact and high-order continuity conditions on general boundary curves using a purely linearly constrained least squares fitting approach. However, the continuity condition on boundary curve segments next to an extraordinary corner point remains approximate G^1. Another approach was developed by Eck and Hoppe for automatic construction of B-spline surfaces from unorganized points with exact G^1 or C^1 continuity [1]. However, the surface model therein is actually composed of Bézier patches rather than B-splines. Piegl and Tiller recently presented an algorithm for generating a collection of G^ε continuous NURBS surfaces that fill an arbitrary n-sided region [5]. If the tolerance ε is very small, one may however end up with a large number of surface patches due to the knots inserted.

This paper develops sufficient and necessary conditions for smoothly connecting multiple B-spline surfaces with exact G^1 continuity using a finite number of check points. A sufficient condition with linear blending is adopted for practical applications. The developed approach has been successfully applied to simultaneously fitting multiple B-spline surfaces from sample points with exact G^1 continuity. The emphasis of the present paper will be on the establishment of the continuity conditions for multiple B-spline surfaces using check points.

2. GENERAL G^1 CONTINUITY CONDITIONS

To further proceed, we define two B-spline surfaces $P(u,v) \in R^3$ and $Q(u,v) \in R^3$ as follows

$$\begin{cases} P(u,v) = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} B_{i,k_u}(u) B_{j,k_v}(v) \cdot P_{ij} \\ Q(u,v) = \sum_{i=1}^{\hat{n}_u} \sum_{j=1}^{\hat{n}_v} B_{i,\hat{k}_u}(u) B_{j,\hat{k}_v}(v) \cdot Q_{ij} \end{cases} \tag{1}$$

For $P(u,v)$, k_u and k_v are the orders of the surface along u- and v-directions, respectively, and $P_{ij} \in R^3$, for $i = 1,2,\cdots,n_u$ and $j = 1,2,\cdots,n_v$, are the $n_u \times n_v$ control points. The two sequences of knots are denoted by $U := \{u_i | = 1,2,\cdots,n_u + k_u\}$ and $V := \{v_j | j = 1,2,\cdots,n_v + k_v\}$ with $u_{k_u} = v_{k_v} = 0$, and $u_{n_u+1} = v_{n_v+1} = 1$. The parameters for $Q(u,v)$ are defined in a similar way with two sets of knots $\hat{U} := \{\hat{u}_i | = 1,2,\cdots,\hat{n}_u + \hat{k}_u\}$ and $\hat{V} := \{\hat{v}_j | j = 1,2,\cdots,\hat{n}_v + \hat{k}_v\}$

For convenience, we present the continuity conditions of two surfaces that have the same orientation and one needs to match the boundaries $P(u,0)$ with $Q(u,1)$. We assume that the two surfaces share the same knots along the common boundary, i.e., the two B-spline surfaces under discussion have the same knot vectors along u-direction with $U = \hat{U}$ and $k_u = \hat{k}_u$ (thus $n_u = \hat{n}_u$). By letting $P(u,0) = Q(u,1)$, $\forall u \in [0,1]$, and noting the fact that the two surfaces share the same basis functions on the common boundary, one obtains the sufficient and necessary G^0 continuity conditions between $P(u,v)$ and $Q(u,v)$ as follows

$$\sum_{j=1}^{k_v-1} B_{j,k_v}(0) \cdot P_{ij} = \sum_{j=\hat{n}_v-\hat{k}_v+2}^{\hat{n}_v} B_{j,\hat{k}_v}(1) \cdot Q_{ij} \qquad i = 1,2,\cdots n_u \tag{2}$$

With the pre-conditions of (2), which is assumed for all the G^1 continuity conditions discussed in the following sections, the sufficient and necessary G^1 continuity condition in a general setting can be defined as the tangent vector co-planarity in the following mixed product form

$$f(u) = \left(\mathbf{P}_v'(u,0), \ \mathbf{Q}_v'(u,1), \ \mathbf{Q}_u'(u,1) \right)$$
$$= \left\langle \mathbf{P}_v'(u,0) \cdot \left(\mathbf{Q}_v'(u,1) \times \mathbf{Q}_u'(u,1) \right) \right\rangle \equiv 0 \qquad \forall u \in [0,1] \tag{3}$$

We first consider a single boundary segment, i.e. the l-th knot interval, whose tangent vectors are denoted by $\mathbf{P}_{v,l}'(u,0)$, $\mathbf{Q}_{v,l}'(u,1)$ and $\mathbf{Q}_{u,l}'(u,1)$, $\forall u \in \left[u_{k_u+l}, u_{k_u+l+1} \right]$ and with $0 \le l \le n_u - k_u$. It is noticed that $\mathbf{P}_{v,l}'(u,0)$ and $\mathbf{Q}_{v,l}'(u,1)$ are vector-valued polynomial functions of degree $k_u - 1$, while $\mathbf{Q}_{u,l}'(u,1)$ is a vector-valued polynomial function of degree $k_u - 2$. The mixed product function $f(u)$, $\forall u \in \left[u_{k_u+l}, u_{k_u+l+1} \right]$ of (3) can thus be expressed in algebraic polynomial form as the following function of degree $3k_u - 4$

$$f_l(u) = \left(\mathbf{P}_{v,l}'(u,0), \ \mathbf{Q}_{v,l}'(u,1), \ \mathbf{Q}_{u,l}'(u,1) \right)$$
$$= \sum_{i=0}^{3k_u-4} a_{i,l} \cdot u^i \equiv 0 \qquad \forall u \in \left[u_{k_u+l}, u_{k_u+l+1} \right] \tag{4}$$

Further, let $\bar{u}_{t,l} \in \left[u_{k_u+l}, u_{k_u+l+1} \right]$ for $t = 1,2,\cdots,3k_u - 3$, be a set of $3k_u - 3$ distinct sample parameters in the l-th knot interval. As explained in the following paragraph, one could also obtain another equivalent set of conditions for equation (3)

$$f_l(\bar{u}_{t,l}) = \left(\mathbf{P}_{v,l}'(\bar{u}_{t,l},0), \ \mathbf{Q}_{v,l}'(\bar{u}_{t,l},1), \ \mathbf{Q}_{u,l}'(\bar{u}_{t,l},1) \right)$$
$$= \sum_{i=0}^{3k_u-4} a_{i,l} \cdot \bar{u}_{t,l}^i \equiv 0 \qquad \text{for } t = 1,2,\cdots,3k_u - 3. \tag{5}$$

In other words, the two surfaces are G^1 continuous $\forall u \in \left[u_{k_u+l}, u_{k_u+l+1} \right]$ if and only if the tangent vector coplanar conditions are satisfied at the above mentioned $3(k_u - 1)$ sample parameters, called check points. This leads to the basic idea of using check points for setting up G^1 continuity conditions. By checking the continuity constraints at a finite number of check points within each of the boundary curve segments, G^1 continuity condition can be achieved along the entire boundary curve. We shall discuss the actual number of check points needed for maintaining G^1 continuity in the next section.

Equation (5) can be explained as follows. Following the basic properties of polynomials, equation (4) holds true if and only if all of the coefficients of the polynomial vanishes, i.e. $a_{i,l} = 0$ for $i = 0,1,2,\cdots,3k_u - 4$. By introducing $a_{i,l} = 0$ for all i's into (5), one is assured that equation (5) holds true. On the other hand, equation (5) forms a set of homogeneous linear equations. For distinct sample parameters, the system matrix is of full rank and equation (5) has only one set of zero solutions, i.e. $a_{i,l} = 0$ for all i's, and hence equation (4) holds true.

3. G^1 CONTINUITY CONDITIONS WITH LINEAR TANGENT BLENDING

It should be noticed that the number of check points, or constraints, defined in (5) using check points is $3(k_u - 1)$, which is a bit too many. In addition, the constraints are nonlinear and are mixed with regard to x-, y- and z-coordinates of the control points. It is therefore difficult to use in practice. In this section, we adopt the following sufficient G^1 continuity condition

$$\mathbf{P}_v'(u,0) = \Phi(u) \cdot \mathbf{Q}_v'(u,1) + \Psi(u) \cdot \mathbf{Q}_u'(u,1) \qquad \forall u \in [0,1] \qquad (6)$$

with simple blending functions $\Phi(u)$ and $\Psi(u)$ for the tangent vectors. Equation (6), combined with the pre-conditions of (2), also guarantees that the three tangent vectors be on the same tangent plane, and hence achieve G^1 continuity across the boundary curve.

We again consider the l-th knot interval at first and use simple blending functions $\Phi_l(u) = \alpha_l$ and $\Psi_l(u) = \beta_l \cdot u + \gamma_l$ with $\alpha_l > 0$, $\beta_l \in R$ and $\gamma_l \in R$, i.e.

$$\mathbf{P}_{v,l}'(u,0) = \alpha_l \cdot \mathbf{Q}_{v,l}'(u,1) + (\beta_l \cdot u + \gamma_l) \cdot \mathbf{Q}_{u,l}'(u,1) \qquad (7)$$
$$\forall u \in \left[u_{k_u+l}, u_{k_u+l+1}\right]$$

Following a similar reasoning of the last section, one could also obtain the following set of equivalent G^1 continuity conditions using check points

$$\mathbf{P}_{v,l}'(\bar{u}_{t,l},0) = \alpha_l \cdot \mathbf{Q}_{v,l}'(\bar{u}_{t,l},1) + (\beta_l \cdot \bar{u}_{t,l} + \gamma_l) \cdot \mathbf{Q}_{u,l}'(\bar{u}_{t,l},1) \qquad (8)$$
$$\forall \bar{u}_{t,l} \in \left[u_{k_u+l}, u_{k_u+l+1}\right], \qquad t = 1,2,\cdots,k_u$$

Compared with equation (5), the total number of check points needed for maintaining G^1 continuity for an independent knot interval is dramatically reduced to k_u. In addition, the individual x-, y- and z-coordinates are independent and the G^1 constraints are purely linear with regard to the control points.

When the constraints of (8) are extended to the entire common boundary curve, we need to further address two issues. One of the issues is whether there exists any restriction over the coefficients α_l, β_l and γ_l for $0 \le l \le n_u - k_u$ in (7) between two neighbouring knot intervals. The other issue is the determination of the number of check points along the entire boundary curve.

In the following discussions, we assume that there are no multiple interior knots. For the first issue, one must use the same blending functions $\Phi(u) = \alpha$ and $\Psi(u) = \beta \cdot u + \gamma$ along the entire common boundary curve due to the continuity constraints of the three tangent vectors across an interior knot. To clarify this issue, we consider the continuity conditions of the first two knot intervals for convenience, i.e. with $l = 0$ and $l = 1$. Following (7), the corresponding G^1 continuity conditions can be written as

$$\begin{cases} \mathbf{P}_{v,0}'(u,0) = \alpha_0 \cdot \mathbf{Q}_{v,0}'(u,1) + (\beta_0 \cdot u + \gamma_0) \cdot \mathbf{Q}_{u,0}'(u,1) & \forall u \in \left[u_{k_u}, u_{k_u+1}\right] \\ \mathbf{P}_{v,1}'(u,0) = \alpha_1 \cdot \mathbf{Q}_{v,1}'(u,1) + (\beta_1 \cdot u + \gamma_1) \cdot \mathbf{Q}_{u,1}'(u,1) & \forall u \in \left[u_{k_u+1}, u_{k_u+2}\right] \end{cases} \qquad (9)$$

Following the properties of B-spline surfaces, the continuity of the three tangent vectors at knot $u = u_{k_u+1}$ are as follows:

- $\mathbf{P}'_v(u,0)$ is continuous at $u = u_{k_u+1}$ up to C^{k_u-2},
- $\mathbf{Q}'_v(u,1)$ is continuous at $u = u_{k_u+1}$ up to C^{k_u-2}, and
- $\mathbf{Q}'_u(u,1)$ is continuous at $u = u_{k_u+1}$ up to C^{k_u-3}.

As a result of the C^0 continuity of the three tangent vectors at $u = u_{k_u+1}$ and noting the fact that $\mathbf{Q}'_v(u,1)$ and $\mathbf{Q}'_u(u,1)$ are linearly independent, we obtain the following equalities

$$\begin{cases} \alpha_0 = \alpha_1 = \alpha \\ \beta_0 \cdot u_{k_u+1} + \gamma_0 = \beta_1 \cdot u_{k_u+1} + \gamma_1 \end{cases} \tag{10}$$

We can further evaluate the derivatives of $\mathbf{P}'_{v,0}(u,0)$ and $\mathbf{P}'_{v,1}(u,0)$ versus u from (9), which yields

$$\begin{cases} \left(\mathbf{P}'_{v,0}\right)_u(u,0) = \alpha_0\left(\mathbf{Q}'_{v,0}\right)_u(u,1) + \beta_0\,\mathbf{Q}'_{u,0}(u,1) + (\beta_0 \cdot u + \gamma_0)\left(\mathbf{Q}'_{u,0}\right)_u(u,1) \\ \qquad\qquad\qquad\qquad \forall u \in \left[u_{k_u}, u_{k_u+1}\right] \\ \left(\mathbf{P}'_{v,1}\right)_u(u,0) = \alpha_1\left(\mathbf{Q}'_{v,1}\right)_u(u,1) + \beta_1\,\mathbf{Q}'_{u,1}(u,1) + (\beta_1 \cdot u + \gamma_1)\left(\mathbf{Q}'_{u,1}\right)_u(u,1) \\ \qquad\qquad\qquad\qquad \forall u \in \left[u_{k_u+1}, u_{k_u+2}\right] \end{cases} \tag{11}$$

As a result of the C^1 continuity of the three tangent vectors at $u = u_{k_u+1}$ and the C^0 continuity of $\mathbf{Q}'_u(u,1)$, and observing (10), we obtain

$$(\beta_0 - \beta_1) \cdot \mathbf{Q}'_u(u_{k_u+1},1) = 0 \tag{12}$$

To avoid degeneracy, we need $\mathbf{Q}'_u(u_{k_u+1},1) \neq 0$. This leads to the conclusion that

$$\beta_0 = \beta_1 = \beta \tag{13}$$

Observing both equations (10b) and (13), we finally have

$$\gamma_0 = \gamma_1 = \gamma \tag{14}$$

In a similar way, one could conclude that the blending functions should also be the same across all other interior knots. This clarifies the first issue and we obtain the continuity conditions using check points with linear blending along the whole common boundary as follows

$$\mathbf{P}'_{v,l}\left(\overline{u}_{t,l},0\right) = \alpha \cdot \mathbf{Q}'_{v,l}\left(\overline{u}_{t,l},1\right) + \left(\beta \cdot \overline{u}_{t,l} + \gamma\right) \cdot \mathbf{Q}'_{u,l}\left(\overline{u}_{t,l},1\right) \tag{15}$$
$$\forall \overline{u}_{t,l} \in \left[u_{k_u+l}, u_{k_u+l+1}\right], \quad t = 1,2,\cdots,k_u, \text{ and } \quad l = 0,1,\cdots,n_u - k_u$$

For the second issue, we first rewrite equation (6) in the follow form, where $g(u)$ is a piecewise vector-valued polynomial function.

$$\mathbf{g}(u) = \mathbf{P}'_v(u,0) - \left[\alpha \cdot \mathbf{Q}'_v(u,1) + \left(\beta \cdot u + \gamma\right) \cdot \mathbf{Q}'_u(u,1)\right] \qquad \forall u \in [0,1] \tag{16}$$

We further study the continuity conditions of $\mathbf{g}(u)$ in the first two knot intervals, and rewrite the above equation for the two intervals separately as

$$\begin{cases} \mathbf{g}_0(u) = \sum_{i=0}^{k_u-1} \mathbf{b}_{i,0} \cdot u^i = \sum_{i=0}^{k_u-1} \overline{\mathbf{b}}_{i,0} \cdot \left(u - u_{k_u+1}\right) & \forall u \in \left[u_{k_u}, u_{k_u+1}\right] \\ \mathbf{g}_1(u) = \sum_{i=0}^{k_u-1} \mathbf{b}_{i,1} \cdot u^i = \sum_{i=0}^{k_u-1} \overline{\mathbf{b}}_{i,1} \cdot \left(u - u_{k_u+1}\right) & \forall u \in \left[u_{k_u+1}, u_{k_u+2}\right] \end{cases} \tag{17}$$

As a result of the continuity constraints of the three tangent vectors, we can conclude that the two vector-valued algebraic polynomials $\mathbf{g}_0(u)$ and $\mathbf{g}_1(u)$ are C^{k_u-3} continuous across $u = u_{k_u+1}$, i.e., $\mathbf{g}_0^{(d)}(u)\big|_{u=u_{k_u+1}} = \mathbf{g}_1^{(d)}(u)\big|_{u=u_{k_u+1}}$ for $d = 0,1,2,\cdots,k_u - 3$. This leads to the following connections between the continuity conditions of two neighbouring patches

$$\overline{\mathbf{b}}_{i,0} = \overline{\mathbf{b}}_{i,1} = 0, \qquad i = 0,1,\cdots,k_u - 3 \tag{18}$$

In other words, if the G^1 continuity conditions are satisfied on one of the patches, the number of constraints or check points needed for its neighbouring patches will be $k_u - 2$ less, i.e. only 2 check points are needed by observing equation (8) and (18). The analysis for other successive knot intervals is similar and it can be concluded that, in case of no multiple interior knots, we need k_u check points in only one of the interior patches and 2 check points for all remaining patches. The total number of check points for the entire boundary curve, i.e., $n_u - k_u + 1$ patches, will therefore be

$$n_c = k_u + 2 \cdot \left(n_u - k_u\right) = 2 \cdot n_u - k_u \tag{19}$$

Let us assume that there are no multiple interior knots. In this case, one can select all the interior and the two end knots as part of the check points. One could further pick up one additional check point in each of the interior knot intervals. The remaining $k_u - 2$ check points can finally be picked up at any position on the boundary curve provided that all selected check points are distinct. Following a similar reasoning, one could also prove that the total number of check points for the sufficient and necessary G^1 continuity condition of equation (3) or (5) for the entire boundary curve is

$$n_c = \left(n_u - k_u\right) \cdot \left(2d_u + 1\right) + 3 \cdot d_u \tag{20}$$

where $d_u = k_u - 1$ is the u-degree of the B-spline surfaces.

4. G^1 CONTINUITY CONDITIONS FOR MULTIPLE SURFACES

When constructing the topological structure for smooth multiple surface modelling, we adopt a boundary representation (B-rep) with quadrilateral domains. Each of the domains is filled with a B-spline surface defined by equation (1). Each surface boundary or domain boundary may be shared by at most two adjacent surfaces. Any number of boundaries may however meet at a common corner. To agree with the check point technique, we assume that each pair of the two adjacent surfaces share the same order k, number of control points n and the knot vector along the common boundary. We could then apply G^1 continuity constraints defined by equation (6) at $n_c = 2 \cdot n - k$ check points and G^0 continuity constraints defined by equation (2) along each of the common boundary curves.

To determine the blending functions, we proceed corner by corner similar to Bézier surfaces widely accepted in literature. Let $\{\mathbf{T}_i\}_{i=1}^N$ be a set of outwards oriented and anti-clockwise unit tangent vectors at a common corner $\mathbf{P}$. A set of parameters $\{\alpha_i > 0, \beta_i\}_{i=1}^N$ should then be assigned to each of the common corners such that

$$\mathbf{T}_{i+1} = -\alpha_i \cdot \mathbf{T}_{i-1} + \beta_i \cdot \mathbf{T}_i, \qquad i = 1, 2, \dots, N \quad \text{mod with } N \tag{21}$$

For the i-th common boundary, the γ_i parameter at one end is actually the β_i parameter at the other end and, therefore, needs not be addressed separately. Similar to the construction of multiple Bézier patches, we should carefully define the parameters $\{\alpha_i > 0, \beta_i\}_{i=1}^N$ for all common corners without conflict and degeneracy at each corner and with twist compatibility for all corners. In this paper, we adopt a symmetrical solution reported in [6] at internal extraordinary corners and an unsymmetrical solution reported in [2] at internal regular corners as follows

$$\begin{cases} \alpha_i = 1 \\ \beta_i = \lambda = 2\cos(2\pi / N) & i = 1, 2, \cdots, N \quad \text{when } N \neq 4 \\ \beta_1 = \beta_3 = 0, \beta_2 = -\beta_4 = \lambda & \lambda \neq 0 \quad \text{when } N = 4 \end{cases} \tag{22}$$

5. EXAMPLES AND DISCUSSIONS

This section presents three examples, all fitted from sample points. The fitting procedure is similar to that published in [3], however, the new check point method is applied for maintaining exact G^1 continuity. All B-spline surfaces shown in Figs. 1 and 2 are defined with bicubic splines and uniform knots with $k_u = k_v = 4$, and the number of control points for each surface is $n = n_u \times n_v = 6 \times 6 = 36$. Each surface has $3 \times 3 = 9$ patches with three knot intervals along a common shared boundary. In Fig. 1, we adopt the check point method with continuity constraints given by (2) and (6) and we set $\alpha = 1$ and $\beta = 0.5$ at the upper side corner and

$\alpha = 1$ and $\beta = 0$ at the lower side corner. There are in total 8 check points. In Fig. 2, we show an example with three B-spline surfaces meeting at a common corner. The orders and the number of control points used are the same as surfaces of Fig. 1. We set $\{\alpha_i = \beta_i = 1.0\}_{i=1}^{3}$ for the central common corner, and $\{\alpha_i = 1.0, \beta_i = 0.0\}$ for other external corners. Fig. 1b and Fig. 2b show the simultaneously fitted multiple surfaces with exact G^1 continuity.

(a)   (b)
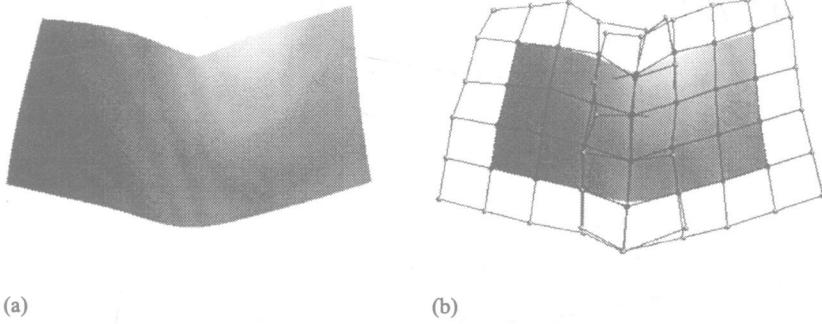
Figure 1. Two fitted B-spline surfaces with exact G^1 continuity using check points: (a) surfaces in shaded image; (b) surface shading with control points.

(2.a)   (2.b)
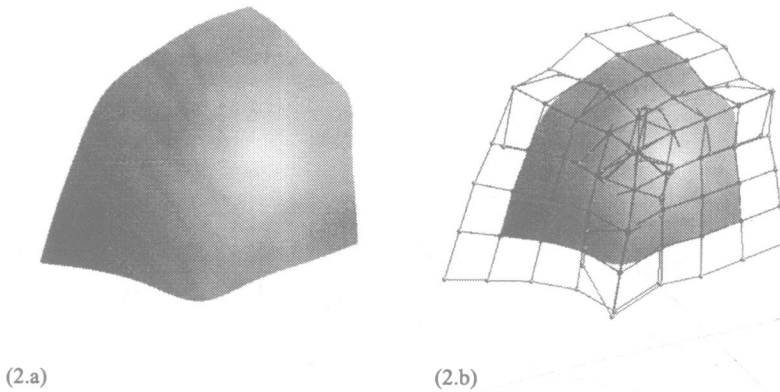
Figure 2. Three fitted B-spline surfaces with exact $G1$ continuity using check points: (a) surfaces in shaded image; (b) surface shading with control points.

The third example is a computer mouse and is shown in Fig. 3. The original measured data are acquired from a plastic body of a computer mouse through a medical CT-scanner. The dimension of the physical object is about 115mm × 70mm × 35mm. The total number of measured points for surface fitting is 7170 as shown in Fig. 3.a with the topological structure of the mouse. It contains eleven inter-connected and two independent quadrilateral domains.

For this example, we set $\alpha = 1$ for all the boundaries/corners and $\beta = 1$ (or $\gamma = 1$) for all extraordinary corners on the upper side of the mouse shown in Fig. 4. The β (or γ) parameter for the latitudinal/vertical boundaries is gradually reduced to 0 when reaching the other side of the mouse. For all other longitudinal/horizontal boundaries, we set $\beta = 0$ (and/or $\gamma = 0$). Fig. 4 illustrates some further details for parameter setting. The final fitted mouse model is shown with shaded image in Fig. 3b. The maximum, average, minimum and standard deviations of the fitting errors are 0.9837mm, 0.1304mm, 0.0013mm and 0.0011mm, respectively.

(3.a)   (3.b)
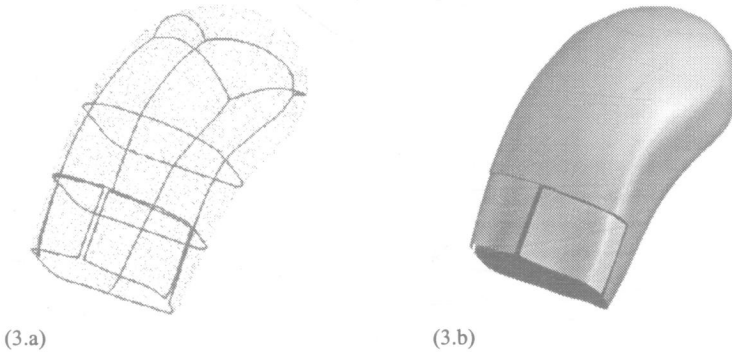
Figure 3. Simultaneous multiple B-spline surfaces fitting with exact G^1 continuity conditions using check points: (a) measured points and the quadrilateral topological domains; (b) shaded image of the final fitted surfaces.
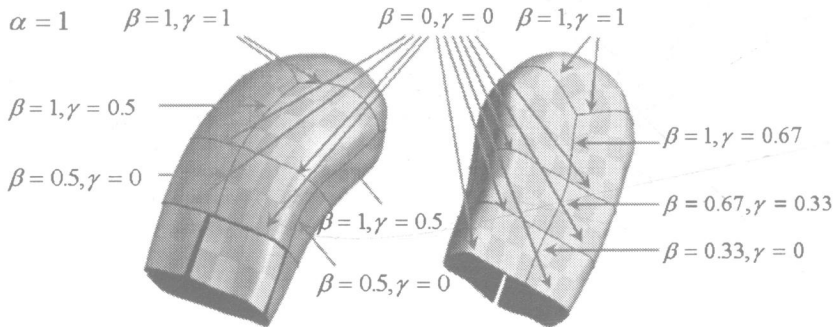
Figure 4. Illustration of the tangent blending parameters for fitting the computer mouse surfaces.

6. CONCLUSIONS

This paper presents an original approach for smoothly connecting multiple B-spline surfaces with exact G^1 continuity using check points. The sufficient and necessary G^1 continuity conditions using check points for connecting two neighbouring B-spline surfaces are first studied. A sufficient condition with linear tangent blending is developed for connecting two neighbouring surfaces and multiple surfaces meeting at a common corner. The proposed method can be used to simultaneously fit smoothly connected multiple B-spline surfaces over irregular topological domains with extraordinary corners. The proposed method can be extended to applications such as surface blending and multi-sided region filling. It can also be extended to support NURBS representation.

ACKNOWLEDGEMENTS

REFERENCES

1   **Eck, M., and H. Hoppe** (1996), Automatic reconstruction of B-spline surfaces of arbitrary topology type, *Computer Graphics*, Vol. 30, pp. 325-334.

2   **Ma, L., and Q. Peng** (1995), Smoothing of free-form surfaces with Bézier patches, *Computer Aided Geometric Design*, Vol. 12, pp. 231-249.

3   **Ma, W., P. C. Leung, E. H. M. Cheung, S. Y. T. Lang, and H. Cui** (1999), Smooth multiple surface fitting for reverse engineering, in *Proceedings of the 32nd CIRP International Seminar on Manufacturing Systems*, Katholieke Universteit Leuven, Belgium, pp. 53-62.

4   **Milroy, M. J., C. Bradley, G. W. Vickers, and D. J. Weir** (1995), *G*1 continuity of B-spline surface patches in reverse engineering, *Computer Aided Design*, Vol. 27, pp. 471-478.

5   **Piegl, L. A., and W. Tiller** (1999), Filling n-sided regions with NURBS patches, *Visual Computer*, Vol. 15, pp. 77-89.

6   **Wijk, J. J.** (1986), Bicubic patches for approximating non-rectangular control-point meshes, *Computer Aided Geometric Design*, Vol. 3, pp. 1-13.

Heterogeneous materials and their applications in high-tech product design

X-J ZHANG and **K-Z CHEN**
Department of Mechanical Engineering, The University of Hong Kong, Hong Kong

ABSTRACT

With rapid developments of high technology in various fields, there appear more critical requirements for special functions of components/ products. Since conventional homogeneous materials cannot easily satisfy these requirements, heterogeneous materials become more and more important. Heterogeneous materials include composites, functionally gradient materials, smart materials, and heterogeneous material with a periodic microstructure. Heterogeneous materials have many special properties, which will give designer more freedom and convenience, and make high-tech product design much more handy than before. This paper presents current progress of the research on heterogeneous materials, stressing on their definitions, properties, and applications in product design.

1 INTRODUCTION

With rapid developments of high technology in various fields, there appear more critical requirements for special functions of components or products now. For example, the thermal deformation of satellite's paraboloid antenna (10 meters in diameter) should be controlled within 0.2 mm in order to work well under the environment with large variation in temperature (-180°C ~120°C). To fulfill it, its thermal expansion coefficient should be close to zero. Since conventional homogeneous materials cannot satisfy these requirements, attention has been focused on heterogeneous materials. Heterogeneous materials include composites, functionally gradient materials, smart materials, and heterogeneous material with a periodic microstructure. In particular, heterogeneous materials with a periodic microstructure have many special properties such as negative Poisson's ratio, zero or negative thermal expansion coefficient, and extreme bulk modulus. Materials are the bases of high technology developments. The application of each new material will surely improve our design abilities in high-tech product design. In order to meet the needs of high-tech product design, a literature survey on current progress of the research on heterogeneous materials, including composites, functionally gradient materials, smart materials, and heterogeneous materials with a periodic microstructure, has been done seriously. This paper introduces it concisely, and focuses on heterogeneous materials' definitions, properties, mechanisms, and current applications in product design.

2 COMPOSITES

Composites are the materials that are made of two or more different constituents, one is matrix and others are reinforcements, by combining together to offer superior properties for a specific application, such as higher strength, higher stiffness, higher fatigue life, and/or better wear resistance. According to the types of matrices, composites can be divided into four subclasses: metal matrix composites, polymer matrix composites, ceramic matrix composites, and carbon matrix composites (1).

2.1 Metal matrix composites
Metal Matrix Composites (MMCs) are the materials in which rigid reinforcement fibers are embedded in a ductile metal or alloy matrix (2). The reinforcement fibers can be short fibers, whiskers, filaments, yarns, or wires. Numerous metals have been used as matrices, such as Aluminum, Magnesium, Titanium and Copper. Because MMCs combine metallic properties (ductility and toughness) with reinforcement fibers' characteristics (high strength and high modulus), they have greater strength in shear and compression (high specific modulus and high strength), high thermal stability, and better wear resistance (3). MMCs are widely used in the aerospace and automotive industries, and other structural applications, such as landing gear drag link of the Space Shuttle Orbiter and high-gain antenna boom for the Hubble Space Telescope (4).

2.2 Polymer matrix composites
Polymer Matrix Composites (PMCs) are the materials combining a polymer-based resin with reinforcing fibers, such as glass fibers and carbon fibers, to obtain exceptional properties. The properties of PMCs are determined by many factors, such as the properties of the fiber, the properties of the resin, fiber volume fraction, and the geometry and orientation of the fibers in the composites. Because the resin matrix spreads the load to fibers and also protects the fibers from damage caused by abrasion and impact, PMCs have high strength, stiffness, and environmental resistance; and are easy to mold complex shapes (5). The applications of PMCs increase very quickly. For example, glass-fiber-reinforced polyester is a very good material for tanks, scrubbers, reactors, mixers, pipe systems, etc. (6,7).

2.3 Ceramic matrix composites
Ceramic Matrix Composites (CEMCs) use ceramic as the matrix and short fibers or whiskers as the reinforcements (4). The advantages of CEMCs are light–weight, which can be used for automobiles to reduce fuel consumption, and high-temperature resistance, which can be used for heat exchangers of gas turbines and fluidized bed combustion units.

2.4 Carbon matrix composites
Carbon Matrix Composites (CAMCs) are the materials consisting of fibers embedded in the carbonaceous matrix (5). They have low densities, high thermal conductivities, and excellent mechanical properties at high temperatures. One kind of CAMCs is carbon fiber reinforced CAMCs (carbon-carbon composites), which is widely used in product design due to their special properties. These materials can withstand temperatures in excess of 2000°C without major deformations. They have very high thermal and electrical conductivities which strongly depend on the direction concerned. In the fiber longitudinal direction, their values are 12-15 times higher than that in the direction perpendicular to the fibers. Their tensile and flexural

properties are fiber-dominated, and their compression behaviors are mainly affected by fiber densities and matrix morphologies. However, their susceptibility to oxidation is very severe over 500°C; and at about 800°C, the rate of oxidation is limited only by the diffusion rate of oxygen through the surrounding gas to the carbon surface. There are many factors influencing their properties, including the choice of pitch and fiber, the use of resin, the weave pattern of carbon fabric, fiber matrix bond strength, carbonization method, and surface treatment of carbon fibers (8). Carbon-carbon composites have very widely applications and have been mostly used in aerospace applications, mainly for aircraft disc brakes, rocket re-entry nose tips, and the parts of rocket nozzles.

3 FUNCTIONALLY GRADIENT MATERIALS (FGMs)

The functionally gradient materials appeared in the 1980's. The initial idea of FGMs was to combine the incompatible properties of heat resistance and toughness with low internal thermal stress, by producing a compositionally graded structure of distinct ceramic and metal phases. Later on, this concept was broadened to include a combination of dissimilar materials without having explicit boundaries for creating materials with new functions. The FGMs with proper spatial gradient have superior properties compared with conventional isotropic composite. The effective thermal expansion coefficient of FGMs is normally between zero and one, and is proportional to volume fraction (volume ratio of metal phase to ceramic phase) (9). The plasticity of FGMs is also dependent on the volume fraction of their constituents. With low volume fraction of ceramic, the FGMs display typical elastic-plastic deformation similar to that of pure metal; but with the increment in volume fraction of ceramic, the response behavior is more complicated due to a microstructure transition at this composition. According to their properties, FGMs can be used into many fields (10): (a) thermal barrier/anti-oxidation coatings; (b) cutting tools: cemented carbides or diamond/SiC; (c) thermoelectric materials; (d) optical film: bandpass filter; (e) piezoelectric actuator.

4 SMART MATERIALS

Smart materials are the materials that have the ability to perform sensing and actuating functions and thus are capable of imitating living systems. Four of the most widely used classes of smart materials are piezoelectrics, magnetostrictives, electrorheological fluids, and shape memory alloys (11).

4.1 Piezoelectrics (PZT)
Piezoelectrics are the materials that can produce an electrical field when they are subjected to a mechanical strain and a deformation when an electrical field is applied to them. In the former case, the constraint for deformation produces mechanical strain and then an electrical field that can be used for stress or strain sensing. In the latter case, piezoelectrics exhibit actuation behaviors that can be used like an actuator. Piezoelectrics can be used in many fields. For example, piezoelectric actuators have been used for active shape, vibration, and acoustic control of structures because of their adaptability and light-weight.

4.2 Magnetostrictives
Magnetostrictiives are the materials in which magnetostriction can be observed when they are put in the magnetic field and a magnetic field is generated when they are stressed. Magnetism

strength is proportional to the materials' strain rate. When distributed as micro-scale devices in a host, magnetostrictives can act as distributed sensors in multifunctional composites. Magnetostrictives can distribute actuation for composites, which can be used for vibration suppression, micropositioning, damage mitigation, and shape control.

4.3 Electrorheological fluids (ER)

ER materials are suspensions which experience reversible changes in rheological properties such as viscosity, plasticity, and elasticity when subjected to electric fields. ER materials can be used in controllable devices and adaptive structures.

4.4 Shape memory alloys (SMAS)

Shape memory alloys are the alloys that can remember and return to a specific shape after considerable deformation. This function results from the thermoelastic martensitic transformation (12). They have many special properties: (a) shape memory effect which refers to SMAs' ability of remembering and returning to a specific shape after considerable deformation; (b) point defect which refers to SMAs' vacancy and anti-site defects because most of SMAs are in fact intermetallic and ordered compounds, even near to the melting temperature, many of them remain ordered; (c) martensite aging effect stabilization which refers to the phenomenon that martensite appears to be more stable with respect to the parent phase during aging so that the reverse transformation temperature is increased; (d) rubber-like behavior which refers to the phenomenon that martensite exhibits a recoverable or pseudo-elastic deformation behavior after aging together with an increase in critical stress; (e) super-elasticity. Due to these special properties, they have extensive application, such as couplings, actuators, orthodontic arch wire, brassieres for women, antennas for cellular phones, and guide wires for catheters in medical use (13,14).

5 HETEROGENEOUS MATERIALS WITH A PERIODIC MICROSTRUCTURE

Heterogeneous materials with a periodic microstructure are the materials that are constituted by the base cell, which is the smallest periodically repetitive unit of material and comprises of a material phase and a void phase at a microscopic scale, as shown in fig. 1. The effective properties of these materials are determined by the topology of base cell and the properties of its constituents. In other words, the effective properties of these materials or their property characteristics can be changed by designing various topologies of their base cells. According to the base cell, heterogeneous materials with a periodic microstructure can be divided into many subclasses, such as periodic truss microstructure, periodic cellular or sandwich microstructure, and periodic fibrous microstructure. With the help of topology optimization
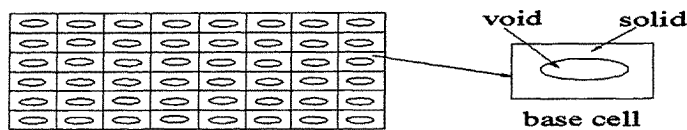

Figure 1. Heterogeneous materials with a periodic microstructure

and homogenization approach, they can be designed to possess very special properties, such as negative Poisson's ratio, zero or negative thermal expansion coefficient, and extreme bulk modulus.

5.1 Periodic truss microstructure

Its base cell is characterized by the triangulation of its members, which induce axial forces in the individual members, giving it a high specific stiffness and strength in all three orthogonal directions. The materials composed of these base cells can be exploited in multi-functional applications. For instance, they could be used in the heat transfer device (15).

5.2 Periodic cellular or sandwich microstructure

This kind of materials (16) consists of three layers: two high-density face sheets are bonded to a low-density core. The core is used to keep the face sheets at the desired distance and carry the transverse shear loads. By choosing appropriate core materials, this kind of materials is able to achieve not only high stiffness and strength but also great savings in weight. Furthermore, different microstructures of cores can lead to very different mechanical properties. For example, the materials with a periodic triangular structure are stiff in both compression and shear, and have the moduli scaled linearly with their densities; but the materials with a periodic hexagonal microstructure are stiff in compression and soft in shear, and have the compressional moduli scaled linearly with their densities. The creep properties of the materials also depend on the architectures of base cells. The materials with a periodic Voronoi structure (17) are stiffer than the materials with a periodic hexagonal honeycomb microstructure, but have greater creep rates due to its random structures. This kind of materials can suppress the dissemination of sound and vibration, and absorb the high crash energy. Therefore, this kind of materials can be used as vibration-proof materials, anti-noise materials, and heat insulation materials, which can be used in many fields such as the wing of airplane, vehicle construction, and transport systems.

5.3 Periodic fibrous microstructure

This kind of materials is composed of periodically arranged fibers. Its elastic properties strongly depend on the direction concerned. Its strength and stiffness are considerably greater in the fiber longitudinal direction than that in the fiber transversal direction, and are strongly dependent on the fiber density and will rise with the increment of the fiber density (18). As to plastic properties, these materials reveal different behaviors, which strongly depend on the deformation. In the case of tension in the direction perpendicular to fibers, its non-linear hardening is observed for a certain range of strain; in the case of tension in the direction parallel to fibers, its linear hardening is observed, due to the fact that the distribution of stress is nearly uniform along fibers. This kind of materials can be used in civil engineering, aircraft, and motor industries (19).

5.4 Other periodic microstructures

With the help of homogenization method and topology optimization, designing various topologies of base cells for the materials can obtain some special properties. The topology of base cell for a material with negative Poisson's ratio is shown in fig. 2 (20,21). In the picture, different color regions represent different constituents. Black regions are filled by solid material, and grey regions correspond to void. Materials with negative Poisson's ratio can be used in fasteners and panels, because, when a plate or bar is bent, it assumes a saddle shape if Poisson's ratio is positive and a convex shape if Poisson's ratio is negative. The convex shapes are more appropriate than saddle shapes for sandwich panels in aircraft or automobiles. Besides, if Poisson's ratio is as small as possible, stress distribution in a flexible pad such as a wrestling mat is more favorable for reducing impact forces upon both small objects such as an elbow and large objects such as a leg or back.

Figure 2. Topology for a material with negative Poisson's ratio

Fig.3 shows the topology of base cell for a material with negative thermal expansion coefficient (22). In the picture, different colors represent different constituents. White regions correspond to void (no material), black regions are filled by Invar alloy (Fe-36% Ni), and red regions are occupied by Nickel metal. Because materials with zero or negative thermal expansion coefficients are able to eliminate thermal mismatch between parts in structures subjected to temperature changes, they are needed for use in structures subjected to temperature changes, such as civil engineering, space structures, and piping systems; and also can be used to overcome positive thermal expansion of other materials. For example, they can be used as thermally operated fasteners.

Void (white): no material
Invar (black): E=150Gpa, v=0.31, α=0.8μm/mk
Nickel (red): E=200Gpa, v=0.31, α=13.4μm/mk
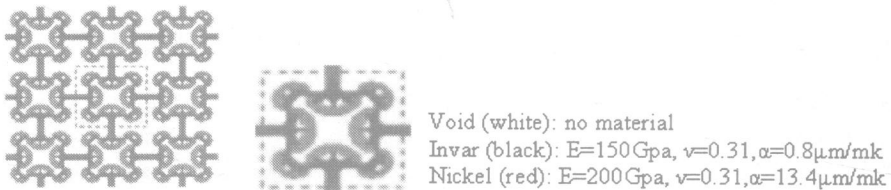
Figure 3. Topology for a material with negative thermal expansion coefficient. E is Young's module, v is Poisson's ratio, α is thermal expansion coefficient.

Fig.4 shows the topology of base cell for a material with extreme bulk modulus. In the

Phase 1(white): E=0.0, v=0.3, f=0.2
Phase 2(grey): E=0.2, v=0.3, f=0.4
Phase 3(black): E=1, v=0.3, f=0.4
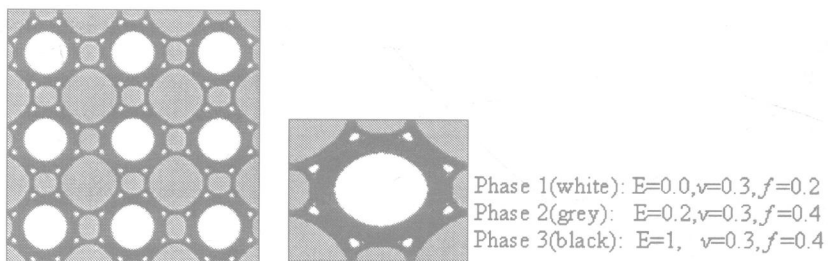
Figure 4. Topology for a material with extreme bulk modulus 0.222. E is Young's module, v is Poisson's ratio, f is volume fraction.

picture, the white regions are occupied by the weakest phase 1, the grey regions are filled by the intermediate phase 2, and the black regions correspond to the strongest phase 3. The materials with extreme bulk properties are of importance in material science and structural optimization. They can be used to improve or optimize man-made materials with extreme

rigidity, and have been widely used in the microelectronic industry, such as a charge storage medium in memories and a gate dielectric in thin films (23).

6 CONCLUSIONS

Materials are the bases of high technology developments. The application of each new material will surely improve our design abilities in high-tech product design. In order to meet the needs of high-tech product design, a literature survey on current progress of the research on heterogeneous materials, including composites, functionally gradient materials, smart materials, and heterogeneous materials with a periodic microstructure, has been done seriously. This paper introduces it concisely, and focuses on heterogeneous materials' definitions, properties, mechanisms, and current applications in product design. Based on this survey, it can be concluded that heterogeneous materials are much superior to conventional homogeneous materials due to their special properties like reducing thermal stress, preventing peeling off of coated layer, preventing micro-crack propagation, and providing high-temperature and impact resistant capability. Therefore, they can surely satisfy the special requirements from high-tech product design and should be widely used in product design further.

ACKNOWLEDGEMENTS

REFERENCES

1. http://me.lsu.edu/~woldesen/intro.html.

2. **Newaz GM, Neber-Aeschbacher H, Wöhlbier FH.** (1995) Metal matrix composites. Translate technical publications. Aedermannsdorf, Switzerland.

3. **Kosoffski R.** (1997) Review of metal matrix materials. In *Journal of Aeronautical Materials*. Vol.17, Iss.3, pp. 2-9.

4. http://www.geocities.com/materialsworldweb/Composites.html

5. http://www.netcomposites.com/education.asp

6. **Nielsen LE, Landel RF.** (1994) Mechanical properties of polymers and composites. M. Dekker, New York.

7. **Wright PK.** (2001) 21st Century Manufacturing. Prentice Hall, Upper Saddle River, N. J.

8. **Winddorse T, Blount G.** (1997) Carbon-carbon composites: a summary of recent development and applications. In *Materials and Design*. Vol.18, Iss.1, pp. 11-15.

9. **Marsan A, Dutta D.** (1998) On the application of tensor product solids in heterogeneous solid modeling. Proceedings of DETC'98, ASME Design Engineering Technical Conferences, DETC98/DAC-5622, pp.1-9.

10. **Hohe J, Becker W.** (2000) A mechanical model for two-dimensional cellular sandwich cores with general geometry. In *Computational Materials Science*. Vol.19, pp. 108-115.

11. **Shakeri C, Noori MN, Hou Z.** (1996) Smart materials and structures. Proceedings of the fourth materials engineering conference, Washington, D.C. Vol. 2, pp. 863 -876.

12. **Yang D.** (2000) Shape memory alloy and smart hybrid composites-advanced materials for the 21st century. In *Materials and Design*. Vol. 21, pp. 503-505.

13. **Otsuka K, Ren X.** (1999) Recent developments in the research of shape memory alloys. In *Intermetallics*. Vol. 7, pp. 511-528.

14. **Gandhi MV, Thompson BS.** (1992) Smart material and structure. Chapman & Hall, New York.

15. **Wallach JC, Gibson LJ.** (2001) Defect sensitivity of a 3D truss material. In *Scripa Materiala*. Vol. 45, pp. 639-644.

16. **Kim B, Christensen RM.** (2000) Basic two-dimensional core types for sandwich structures. In *International Journal of Mechanical Sciences*. Vol.42, pp. 657-676.

17. **Andrews EW, Gibson LJ.** (2001) The role of cellular structure in creep of two-dimensional cellular solids. In *Materials Science and Engineering*. A303, pp. 120–126.

18. **Holmberga S, Perssona K, Peterssonb H.** (1999) Nonlinear mechanical behavior and analysis of wood and fibre materials. In *Computers and Structures*. Vol. 72, pp. 459-480.

19. **Wiceckowski Z.** (2000) Dual finite element methods in homogenization for elastic-plastic fibrous composite material. In *International Journal of Plasticity*. Vol.16, pp. 199-221.

20. **Larsen VD, Sigmund O, Bouwstra S.** (1997) Design and fabrication of compliant micromechanisms and structures with negative Poisson's ratio. In *Journal of Microelectromechanical Systems*. Vol.6, Iss. 2, pp. 99-106.

21. **Zohdi TI, Wriggers P.** (2001) Aspects of the computational testing of the mechanical properties of microheterogeneous material samples. In *International Journal for Numerical Methods in Engineering*. Vol. 50, pp. 2573-2599.

22. **Sigmund O, Torquato S.** (1997) Design of materials with extreme thermal expansion using a three-phase topology optimization method. In *Journal of the Mechanics and Physics of Solids*. Vol. 45, Iss. 6, pp. 1037-1067.

23. **Leonid V. Gibiansky, Ole Sigmund.** (2000) Multiphase composites with extremal bulk modulus. In *Journal of the Mechanics and Physics of Solids*, Vol. 48, pp. 461-498.

A modelling method of heterogeneous components

K-Z CHEN
Department of Mechanical Engineering, The University of Hong Kong, Hong Kong
X-A FENG
Faculty of Mechanical Engineering, Dalian University of Technology, China

ABSTRACT

In order to represent, analyze, optimize, and manufacture a component made of heterogeneous materials, a computer model of the heterogeneous component needs to be built first. Current modeling techniques focus only on capturing the geometric information and cannot satisfy the requirements from modeling the heterogeneous components. This paper develops a modeling method, which can be implemented by employing the functions of current CAD graphic software and can obtain the model including both the material information (about its microstructures and constituent composition) and the geometry information without the problems arising from too much data.

1. INTRODUCTION

With rapid developments of high technology in various fields, there appear more critical requirements for special functions of components/products, such as negative Poisson's ratio, zero thermal expansion coefficient, and extreme bulk modulus. These special requirements cannot be satisfied by using conventional homogeneous materials. The attention has now focused on heterogeneous materials (HM), including composite materials, functionally graded materials, and heterogeneous materials with a periodic microstructure. A heterogeneous component (HC) can be made of single material in a heterogeneous form or multiple heterogeneous materials, according to the requirements for special functions of the components/products. In order to design and manufacture the heterogeneous components, the computer models for representing the heterogeneous components, especially those made of multiple heterogeneous materials, need first to be built, so that further analysis, optimization and layered manufacturing can be implemented based on the models.

Current modeling techniques focus only on capturing the geometric information [1]. Some researchers are now focusing on modeling heterogeneous objects by including both the variation in constituent composition and the geometry in the solid model [2-4]. But modeling or representing the microstructure of heterogeneous components is beyond their scope [2]. This paper develops a modeling method, which can be implemented by employing the functions of current CAD graphic software and can obtain the model including both the material information about its microstructures and constituent compositions and the geometry information without the problems arising from too much data.

2. REQUIREMENTS OF MODELING HETEROGENEOUS COMPONENTS

Before developing a modeling method for heterogeneous components, the requirements from representing a component made of heterogeneous materials should be made clear first. Since heterogeneous materials cover composite materials, functionally graded materials, and heterogeneous materials with a periodic microstructure, the requirement for each one of them is analyzed, respectively, as follows:

A composite material [5-7] consists of one or more discontinuous phases distributed in one continuous phase. The continuous phase is called the matrix and may be resin, ceramic, or metal. The discontinuous phase is called reinforcement and may be fibers, particles, or voids. Reinforcement is used to improve certain properties of matrices, such as stiffness, behavior with temperature, and resistance to abrasion. For instance, lead inclusions in copper alloys make them easier to be machined. The properties of composite materials result mainly from the properties of both their matrix and reinforcements and the geometry and distribution of their reinforcements. Thus, to describe a component made of a composite material, its model will have to specify the geometry, material, and distribution of reinforcement and the matrix material as well as the geometry of the component.

Functionally graded materials are used to join two different materials without stress concentration at their interface. Gradation in properties from one portion to another portion can be determined by material constituent composition. The volume fraction of one material constituent should be changed from 100% on one side to zero on another side, and that of another material constituent should be changed the other way round. In fact, there are many material composition functions [8]. The designers can choose certain composition functions from them for their applications. For example, the following parabolic function is selected for material composite function of the metal/ceramic functionally graded material in the cylinder of vehicular engines or pressure vessels:

$$V_A = a_0 + a_1 x + a_2 x^2 \tag{1}$$

where V_A is the volume fraction of metal and x is the distance from one side. The coefficients of the parabolic function are optimized subject to criteria that the thermal flux across the material is minimized, and the thermal stresses are minimized and restricted below the yield stress of the composite material. After the determination of composition function, physical properties can also be estimated based on property estimation models [8]. Thus, in order to describe a component made of a functionally graded material, its model will have to specify its material constituents and their constituent composition as well as the geometry of the component.

Heterogeneous material with a periodic microstructure is indicated in Figure 1. Such materials are described by the base cell, which is the smallest repetitive unit of material and comprises of a material phase and a void phase. It should be emphasized that, in comparison with the dimensions of the component, the size of these non-homogeneities should be very small. The effective properties of the heterogeneous material are determined by the topology of its base cell and the properties of its constituents, and can be predicted by the mathematical theory of homogenization [9-11]. In other words, the effective properties of the heterogeneous material can be changed by designing various topologies of its base cells. With the homogenization method, the topology of base cell can be designed, according to certain requirements, using topology optimization [12,13]. Thus, in order to describe a component made of a heterogeneous material with a periodic microstructure, its model will have to

specify its material and its microstructure (or base cell) as well as the geometry of the component.

| (a) | (b) |
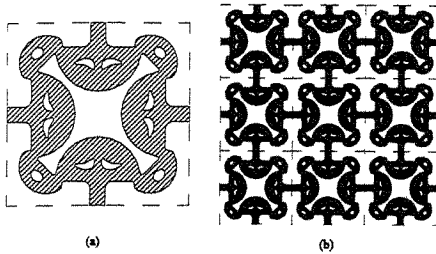
Fig.1 A HM with a periodic microstructure   **Fig.2 An example of HC**
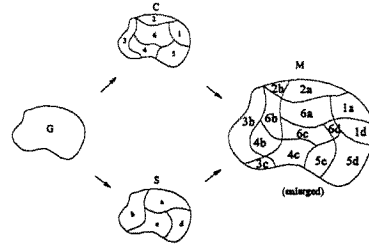
According to the above analysis, it is obvious that the functional requirement (FR) of its model can be decomposed into three sub-FRs: representing geometry, material constituent composition, and material microstructures (including the geometry and distribution of reinforcement for composite materials) of the component, which are noun phrases corresponding to "what we want to achieve" and can be written as:

FR_1 = Representing geometry of the component
FR_2 = Representing material constituent compositions of the component
FR_3 = Representing material microstructures of the component

In other words, in order to represent a component made of composite, functionally graded materials, and/or heterogeneous materials with a periodic microstructure, its model have to be able to satisfy the above three functional requirements. According to Axiomatic Design [14], the model should be decomposed into three sub-models as the design solution (DS) to satisfy the three FRs, respectively. These DSs are stated starting with a verb corresponding to "how we achieve it" and can be written as:

DS_1 = Build its geometric model
DS_2 = Build its material constituent composition model
DS_3 = Build its material microstructure model

Thus, according to Axiomatic Design, the design equation for it can be obtained as follows:

$$\begin{Bmatrix} FR_1 \\ FR_2 \\ FR_3 \end{Bmatrix} = \begin{bmatrix} X & 0 & 0 \\ X & X & 0 \\ X & X & X \end{bmatrix} \begin{Bmatrix} DS_1 \\ DS_2 \\ DS_3 \end{Bmatrix} \tag{2}$$

From the equation obtained, it can be seen that the design matrix is a triangular matrix, which indicates that the design is a decoupled design and satisfies Independence Axiom [14]. Therefore, these DSs can be adopted, and it is correct to build the three sub-models to represent the component made of heterogeneous materials.

3. MODELING OF HETEROGENEOUS COMPONENTS

3. MODELING OF HETEROGENEOUS COMPONENTS

A 3D model representing the geometry of a component can be made by using current CAD graphic software[1], and is indicated by G. This model can be divided into n regions or parts based on their different types of material constituent compositions. Thus, the material constituent composition set can be indicated by:

$$C = \{C_i, i = 1, 2, ..., n\} \tag{3}$$

According to its material microstructures, the geometry model can also be divided into m regions if there are m different types of microstructures. The material microstructure set can be written as:

$$S = \{S_j, j = 1, 2, ..., m\} \tag{4}$$

Thus, the material set (M) of the component can be obtained by solving Cartesian product of C and S as:

$$M = C \times S = \{M_{ij} \mid i \in (1, 2, 3..., n), j \in (1, 2, ..., m\} \tag{5}$$

For example, there are six different material constituent composition regions ($n = 6$) and four different material microstructure regions ($m = 4$) in the component G as shown in Figure 2. Solving its Cartesian product of C and S can obtain fourteen material regions, each of which contains the information of both material constituent composition and microstructure. The first Arabic figure of the symbol of each region is the code name of material constituent composition region, and the second English letter is the code name of material microstructure region. Region $6a$, for instance, indicates that its material constituent composition in this region is determined by that in Region 6 of material constituent composition set and its material microstructure is specified by that in Region a of material microstructure set.

3.1 Material constituent composition models

Each region in material constituent composition set has a specified material constituent composition. The volume fraction of the k-th material constituent at the position (x, y, z) can be represented as:

$$V_k = f_k(x, y, z) \tag{6}$$

This material composition function along with primary material combinations and intended applications can be obtained from many literatures [8], and organized into a database for applications. Designers may select suitable material composition functions from it for their applications according to the functional requirements of component.

Based on schema theory [15], frame and slots can be used to organize the knowledge for modeling. The model for the i-th material constituent composition region is then designed using Axiomatic Design as the following schema form:

$C_i = \{$ Number of material types: q_i
       Material type: $m_1, m_2, ..., m_{qi}$
       Coordinate system type: (Cartesian, cylindrical, or spherical
                  coordinate system)
       Origin of coordinate system: X_0, Y_0, Z_0
       Orientation of coordinate system: α, β, γ
       Material constituent composition function: $[V_k = f_k(x, y, z), k = 1,$

$$2,\ldots, q_i \mid \sum_{k=1}^{q_i} V_k = 1, \ (x, y, z) \in C_i \ ]$$

}   (7)

3.2 Material microstructure models

The material microstructure model introduced in this paper covers both the microstructure of composite materials and the periodic microstructure of heterogeneous materials.

3.2.1 Material microstructure models for composite materials

The composite material consists of matrix and reinforcements (or inclusions). The later may be particles or voids, which may have various shapes, sizes, and distributions. Since the components are considered to be made by layered manufacturing technology in this paper, the modeling is for the composite with the inclusions having similar shapes and the distributions which are according to certain pattern, which can have more stable properties than those with the inclusions having randomly changed shapes, sizes, and distributions.

Using the same method, the model for the j-th material microstructure region is designed as the following form:

S_j = { Insertion 1: "code name of inclusion parametric model"
　　　　　Inclusion material: "code name of material"
　　　　　Coordinate system type: (Cartesian, cylindrical, or coordinate
　　　　　　　　　　　　　　system)
　　　　　Origin of coordinate system: X_1, Y_1, Z_1
　　　　　Orientation of coordinate system: α_1, β_1, γ_1
　　　　　(x, y, z): $[(x_1(t_1), y_1(t_2), z_1(t_3)) \mid (t_1, t_2, t_3) \in$ "Integer", $(x, y, z) \in S_j]$
　　　　　D: $F_{d1}(x, y, z)$
　　　　　θ : $f_{\theta 1}(x, y, z)$
　　　　　Type of RBO: (matrix dominant subtraction, inclusion dominant
　　　　　　　　　　　　complex union, or matrix dominant complex union)
　　　　Insertion 2: ......
　　　　......
}   (8)

This model consists of several sub-frames based on the number of inclusion types, each of which describes one type of inserting operation. In each sub-frame, each slot is defined as follows:

The first slot is the shape of inclusion (particle or void), which is represented by the code name of its 3D parametric model and can be retrieved from a parametric graphics library according to its code name. The second slot is the material type of inclusion, which is represented by the code name of material type and can be retrieved from a material database according to its code name. When the inclusion is a void, not a particle, "nil" will be filled in. The third slot is the local coordinate system type of material structure region, which may be Cartesian, cylindrical, or spherical coordinate system. The fourth and fifth slots are the origin and the orientation of the local coordinate system, respectively, which are based on global coordinate system. The sixth slot is the position of inclusion, which should be at a point in a matrix of local coordinate system and within its material microstructure region. For example, if Cartesian coordinate system is applied, its matrix can be determined, as shown in Figure 3, by:

$$x = \{x(t_1), \ t_1 = 1, 2, \ldots\} = \{x_1, x_2, \ldots\}$$
$$y = \{y(t_2), \ t_2 = 1, 2, \ldots\} = \{y_1, y_2, \ldots\} \qquad (9)$$
$$z = \{z(t_3), \ t_3 = 1, 2, \ldots\} = \{z_1, z_2, \ldots\}$$

The seventh slot is the dimension of inclusion, which consists of all the parameters of its 3D parametric model and is determined by a special function set, $D = F_d (x, y, z)$, based on the position of inclusions. When all the inclusions have same dimension, its functions are constant. The eighth slot is the orientation of inclusions, which is also determined by a special function set,

$\theta = f_\theta (x, y, z)$, based on the position of inclusions. The ninth slot is the type of Reasoning Boolean Operation (RBO) [16]. Unlike conventional Boolean operations, the RBO needs to be executed according to the dominant material information, which is defined either matrix dominant or inclusion dominant union, subtract, and intersect according to the design intent. Here, three types of RBO, matrix dominant subtraction, inclusion dominant complex union, or matrix dominant complex union, will be used.
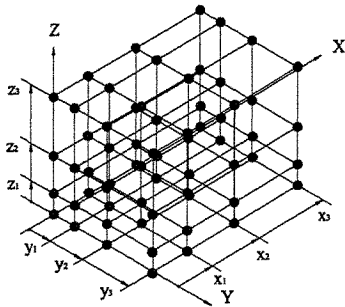
Fig.3 Inclusion positions   Fig.4 Base cell with three types of materials
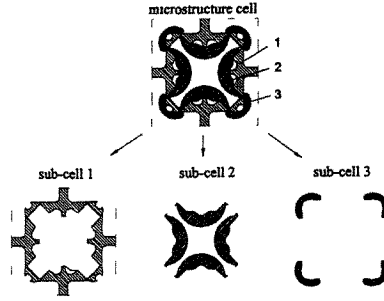
3.2.2 Material microstructure models for heterogeneous materials with a periodic microstructure

A heterogeneous material with a periodic microstructure is described by its base cell, which is the smallest repetitive unit of material and comprises of a material phase and a void phase. The method for building its model is to implement matrix dominant complex union. Taking the base cell shown in Figure 1(a) as an example, its operation result can be shown in Figure 1(b). Its model can be expressed as follows:

$S_j = \{$ Insertion 1: "code name of basic-cell"

   Inclusion material: "nil"

   Coordinate system type: Cartesian coordinate system

   Origin of coordinate system: X_0, Y_0, Z_0

   Orientation of coordinate system: α, β, γ

   $(x, y, z) : [(k_x t_1 + c_x, k_y t_2 + c_y, k_z t_3 + c_z) \mid (t_1, t_2, t_3) \in$ "Integer",

     $(x, y, z) \in S_j]$

   D: "constant"

   θ : "constant"

   Type of RBO: "matrix dominant complex_union"

   $\}$ (10)

In this model, the first slot is the pattern of basic cell, which can be retrieved from parametric microstructure graphics library according to the code name of its pattern. The second slot is the type of inclusion material. When the heterogeneous material with a periodic

microstructure consists of only one type of material, its material is matrix material, which has been determined by material constituent composition model. Therefore, the "Inclusion material" can be filled by "nil". The third slot is coordinate system, which may also be Cartesian, cylindrical, or spherical coordination system. In Cartesian coordination system, all the dimensions for basic cell are constant. But, if cylindrical or spherical coordination system is used, the dimensions of basic cells in the same circle are constant and those in different circles are the linear function of their radial position coordinates.

When the heterogeneous material with a periodic microstructure consists of several types of materials, such as three types of materials as shown in Figure 4, its basic cell can be decomposed into three sub-cells. The material of sub-cell with the largest volume or functionally graded material in a cell is defined as matrix material. For the basic cell in Figure 4, sub-cell 1 has the largest volume, and its material is taken as matrix material and is determined by its material constituent composition model. The other two sub-cells can be inserted into it using RBO: inclusion dominant complex_union. Its model can be written as:

Sj = { Insertion 1: "sub-cell 1"
  Inclusion material: "nil"
  Coordinate system type: Cartesian coordinate system
  Origin of coordinate system: X_0, Y_0, Z_0
  Orientation of coordinate system: α, β, γ
  (x, y, z): $[(k_x t_1 + c_x, k_y t_2 + c_y, k_z t_3 + c_z) \mid (t_1, t_2, t_3) \in$ "Integer",
          $(x, y, z) \in S_j]$
  D: "constant"
  θ : "constant"
  Type of RBO: "matrix dominant complex_union"
Insertion 2: "sub-cell 2"
  Inclusion material: "material of sub-cell 2"
  Coordinate system type: Cartesian coordinate system
  Origin of coordinate system: X_0, Y_0, Z_0
  Orientation of coordinate system: α, β, γ
  (x, y, z): $[(k_x t_1 + c_x, k_y t_2 + c_y, k_z t_3 + c_z) \mid (t_1, t_2, t_3) \in$ "Integer",
          $(x, y, z) \in S_j]$
  D: "constant"
  θ : "constant"
  Type of RBO: "inclusion dominant complex_union"
Insertion 3: "sub-cell 3"
  Inclusion material: "material of sub-cell 3"
  Coordinate system type: Cartesian coordinate system
  Origin of coordinate system: X_0, Y_0, Z_0
  Orientation of coordinate system: α, β, γ
  (x, y, z) : $[(k_x t_1 + c_x, k_y t_2 + c_y, k_z t_3 + c_z) \mid (t_1, t_2, t_3) \in$ "Integer",
          $(x, y, z) \in S_j]$
  D : "constant"
  θ : "constant"
  Type of RBO : "inclusion dominant complex_union"
}   (11)

4. CONCLUSIONS

This paper develops a modeling method for heterogeneous components using Axiomatic

Design. The method divides a component into two region sets (C and S) based on its material constituent compositions and material microstructures. The models for the two region sets are built, respectively, using schema theory. Thus a component can be decomposed, by Cartesian product of C and S, into material regions (M_y), each of which has the same material constituent composition and microstructure. Integrated with conventional 3D computer model of a component, this method can employ the functions of current CAD graphic software and obtain the model including the material information (about microstructures and constituent compositions) along with geometry information in current 3D solid modeling without problems arising from too much data.

ACKNOWLEDGEMENTS

REFERENCES

1 Lee, K. (1999) Principle of CAD/CAM/CAE System. Addison-Wesley Longman, Inc.,Reading.

2 Kumar, V. and Dutta, D. (1998) An approach to modeling & representation of heterogeneous objects. Journal of Mechanical Design, Vol. 120, pp. 659-667.

3 Kumar, V., Burns, D., Dutta, D. and Hoffmann, C. (1999) A framework for object modeling. Computer-Aided Design, Vol. 31, pp. 541-556.

4 Chiu, W.K. and Tan, S.T. (2000) Multiple material objects: from CAD representation to data format for rapid prototyping. Computer-Aided Design, Vol. 32, pp. 707-717.

5 Berthelot, J.M. (1999) Composite materials: mechanical behavior and structural analysis. Springer-Verlag, New York.

6 Chawla, K.K. (1998) Composite materials: science and engineering. Springer-Verlag, Inc., New York.

7 Barbero, E.J. (1998) Introduction to composite materials design. Taylor & Fancis, Ann Arbor, MI.

8 Bhashyam, S., Shin, K.H. and Dutta, D. (2000) An integrated CAD system for design of hetergeneous objects. Rapid Prototyping Journal, Vol. 6, pp. 119-135.

9 Bendsoe, M.P. (1995) Optimization of structure topology, shape, and material. Springer Verlag, Berlin.

10 Hassani, B. and Hinton, E. (1998) A review of homogenization and topology optimization I --- homogenization theory for media with periodic structure. Computers and Structures, Vol. 69, pp. 707-717.

11 Hassani, B. and Hinton, E. (1998) A review of homogenization and topology optimization II --- analytical and numerical solution of homogenization equations. Computers and Structures, Vol. 69, pp. 719-738.

12 Silva, E.C.N., Fonseca, J.S.O. and Kikuchi, N. (1997) Optimal design of piezoelectricmicrostructure. Computational Mechanics, Vol. 19, pp. 397-410.

13 Larson, U.D., Sigmund, O. and Bouwstra, S. (1997) Design and fabrication of compliantmicromechanisms and structures with negative Poisson's ratio, Journal of Microelectromechanical Systems, Vol. 6, pp. 99-106.

14 Suh, N.P. (1990) The Principle of Design. Oxford University Press, Inc., New York.

15 Jonassen, D.H., Beissner, K. and Yacci, M. (1993) Structural knowledge: techniques for representing, conveying, and acquiring structural knowledge. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

16 Sun, W., Lin, F. and Hu, X. (2001) Computer-aided design and modeling of composite unit cells. Composite Science and Technology, Vol. 61, pp. 289-299.

Enhancing conceptual CAD interface by haptic feedback and two-handed input

G ZHAN and **I GIBSON**
Department of Mechanical Engineering, University of Hong Kong, Hong Kong

SYNOPSIS

This paper proposed an approach to facilitate the conceptual geometric design process with haptic feedback and two-handed technologies. In the conceptual stage of design, the shape and dimension are not rigidly defined and therefore designers have great freedom and control over the model. However, conventional CAD systems do not satisfy the requirements of conceptual design due to lack of intuitiveness and similarity to traditional modelling methods. Haptic feedback can increase information flow between users and computer and improve reality and intuitiveness of simulation. Two-handed input provides user more degrees of freedom and the potential to incorporate users' learnt skill of traditional modelling into conceptual CAD system. The architecture of this CAD system, physical metaphor of traditional tools and two-handed task integration are discussed.

1 INTRODUCTION

In the conceptual phase of the design process, designers are concerned with the generation of ideas and concepts for the shapes of products. Although the conceptual phase is only a short period in the product development process, it has many important implications for the later phases.

It is necessary to support conceptual modelling with computers. Computer based systems do have some benefits over the traditional tools, including the easier integration of conceptual model and the later phases of the computer supported activities such as finite element analyses and CAM, preserving the development history of a design and the convenience to store and share a model by electronic means. [14]

There is abundant reason to use computers in conceptual design. However, the conventional CAD systems do not match the requirements of conceptual modelling. In conceptual design, the product details such as shape features and dimensions are not rigidly or precisely defined. The shapes and dimensions of products are more qualitative than quantitative. However, the conventional CAD requires designers to specify shape and dimension to create model even if

this information perhaps is not necessary at this stage. These drawbacks of conventional CAD used for conceptual design become requirements for new conceptual CAD tools. A CAD system is useful only if it meets the requirements of the designer. Some works have been done to research conceptual CAD. [6] In this paper, we refer to C-CAD as conceptual CAD to distinguish it from conventional CAD. C-CAD should support user's skills of traditional modelling, which allow designers to fluently convey their idea. Also, C-CAD systems should work in the way designers think. The thinking style of the designer is not mathematical, especially in the conceptual phase. Therefore, C-CAD applications should hide their underlying mathematical foundation from the designer.

In this paper, we introduce haptic feedback and two-handed technologies with potential to fulfil the requirements of a C-CAD system. Haptic feedback is the state of the art technology to increase the information flow between the user and computer system. Hundreds of papers published each year in the haptics area indicate that haptic feedback technology would be a promising technology to use. Two-handed interfaces are also being intensively researched. An interface featuring these two technologies is proposed and is partially implemented for C-CAD.

2 HAPTIC FEEDBACK

Haptic perception has two components: 1) kinesthesis, or the perception of forces and joint angles, and 2) tactile or cutaneous perception, or the sensation of surface feature, texture, slip, and other aspects of local contact on the skin.

A haptic feedback device can provide force and/or tactile sensations to the user. The value of haptic feedback in improving performance in teleoperation, virtual surgery and manipulation tasks in virtual environments is well documented. [3][20]

2.1 A brief history of haptics research
One consistent trend in the development of computing is the continuing expansion of information flow between human and computer. As the technology developed, input came from punch cards, magnetic tape to keyboards and mouse, while output became words on a display screen, images and sounds. Haptic devices are a further way to enhance this information flow.

The earliest haptic device was Braille reader, through which a blind user can read computer output text. 2D mouse and joystick with force feedback are quite common nowadays. These use vibrating motors to simulate different force related events and different material or texture. However, these devices have little to do with 3D force feedback.

More sophisticated devices include PHANToM from SensAble Inc and CyberGrasp and CyberForce from Immersion Inc. [18] Among them, perhaps PHANToM is the best known. PHANToM has the appearance of a small robotic arm from which a stylus suspended. As the user moves the stylus, the robotic arm detects the motion and a cursor moves around in 3D virtual space. Using the device, if one encounters a "solid" object in the computer simulated 3D space, the stylus is stopped by a motor on the robotic arm. Moving along a surface can provide tactile information about the surface's texture, and the user can readily and intuitively sense curves and corners of an object.

2.2 Haptic display

In 1965 Ivan Sutherland set forth a vision of "The Ultimate Display," a view of a display as a window into a virtual world. This display included seeing, hearing, and feeling in the virtual world. Calculation variables such as force, acceleration and vibration, etc., which are necessary for haptic interface control, is referred to as haptic display or haptic rendering. A haptic device is able to faithfully present forces due to tool-model interaction only if the desired forces are computed accurately and updated in time. Haptic display has the challenging tasks of determining static and dynamic interactive behavior of objects in the virtual environment according to Newtonian physical law.

The first step for a haptic display is collision detection, computing whether and where objects are contact. Collision detection assures objects do not traverse into each other. This computation is global in the whole virtual scene because any objects could touch anything else, therefore, it can be arbitrarily complex according to the geometry of environment. Bounding box techniques were developed to enhance the computation efficiency. [4]

When a virtual tool contacts with a model, the surface of the model deforms according to the shape feature of the tool and the user-applied force. If the object regains its original shape once the applied force released, the deformation is elastic, otherwise the object remains deformed and the deformation is plastic. Great deals of papers have documented the geometric deformation of a curve, a surface or a volume. A force is computed according to a certain physical deformation model. For elastic deformation, Hooke's law is used in most circumstances where a force is determined by the stiffness constant of a virtual object and the surface deformation distance along a specified direction. Walls or some smooth rigid surface in a C-CAD environment can often be simplified and treated as springs with high stiffness constant so that the contact force can be calculated according to Hooke's law. A more complex elastic deformation model based on finite element method was developed by Frank A. et al. for real time haptic feedback of soft tissue. [8]

Unlike, for example, the virtual simulation of surgery, where the simulated tissue may be more elastic and softer, the material of C-CAD models could be more plastic or solid, so that the haptic display model can be simpler. One example is Burdea et al.'s illustration of contact forces when grasping and deforming a virtual soda can. [2] The surface-based can model first has an elastic deformation when force is below a certain value. When increasing force exceeds the value, the deformation becomes plastic. Another example of plastic object deformation is the carving model developed by Yamamoto et al. Their approach is based on voxel deformation. Voxels eliminated by a carving tool and the translational vector of the tool tip are coupled to determine the reaction force. [21]

Other aspects of haptic display include dynamic force interaction, physical constraint and friction force. All these aspects help to enhance the realistic and intuitiveness of haptic feedback. [16]

2.3 Related works – haptic feedback in CAD applications

Haptic interface can be employed in CAD applications. The FreeForm modelling system was the first commercial CAD application to take advantage of haptic feedback. Integrating haptic technology with 3D modelling technology, Freeform provides designers with familiar physical metaphors such as sculpting, wire cutting, clay shaping and deforming, which designers have

liked to use for years. Moreover, FreeForm's sophisticated modelling techniques help designers faithfully express their design ideas.

McDonnell et al developed a real-time sculpting system, namely Virtual Clay, which provides the user with a natural interface for direct, force-based deformation. The Virtual Clay system is featured with its natural, haptic-based interaction to provide the user with a realistic sculpting experience. [16]

Gregory et al presented a 3D interface for interactively editing and painting a polygonal mesh using a PHANToM device. A designer uses the system to create and refine a 3D multi-resolution polygonal mesh. Its appearance can be further enhanced by directly painting onto its surface. [9]

Dachiile et al presented a haptic approach for the direct manipulation of physics-based B-spline surfaces. His method permits users to interactively sculpt virtual material with a haptic device, and feel the physically realistic presence of virtual B-spline objects with force feedback. [5]

3 TWO-HANDED INTERACTION

3.1 Guiard's Three Principles
Guiard's three principles explain how our two hands work in harmony. Understanding the principles helps us understand two-handed interaction with computers. Guiard's analysis of human skilled bimanual action provides an insightful theoretical framework for classifying and understanding the roles of the hands. [10]
1. Right-to-left reference: Motion of the right hand typically finds spatial reference in the results of motion of the left hand (for right handed people). For example, when writing, the nonpreferred hand controls the position and orientation of the page, while the preferred hand performs the actual writing by moving the pen relative to the nonpreferred hand.
2. Asymmetric scales of motion: The right and left hands are involved in asymmetric scales of motion. The right hand specializes in rapid, small-scale movements; the left, in slower, larger-scale movements. During handwriting, for example, the movements of the left hand adjusting the page are infrequent and coarse in comparison to the high-frequency, detailed work done by the right hand.
3. Left-hand precedence: The left hand precedes the right hand: the left hand first positions the paper, and then the right hand begins to write.

3.2 Related works
Several research systems have attempted to incorporate two-handed interaction into 2D desktop applications. Buxton et al developed a two-handed tool-glass and magic-lens system for use in a 2D drawing program. [1] In their system, the user controlled two cursors. The nonpreferred hand controlled the coarse placement of the tool, while the preferred hand performed more precise selection and drawing operations. Two-handed interface with 2D input devices can not only be applied in 2D graphical applications but also in 3D systems. An example of such a system is the SKETCH system that is operated with 2D input devices with 2 degrees of freedom (DOFs). [22] The system's two devices altogether provide 4-DOF and allow users to perform a number of CAD operations with two hands. Objects can be moved,

rotated and scaled, the viewpoint and other camera parameters can be manipulated and several other editing operations are supported.

Two-handed interaction has been incorporated into virtual reality and 3D space systems. Hinckley has done some notable work in two-handed interaction in a desktop environment. [11] In his neurosurgical operation planning system, he embedded a six DOFs tracker in a prop (a doll's head) that was controlled by the nonpreferred hand and then manipulated another 6 DOFs tracker with the preferred hand in order to define slicing planes. Shaw et al developed a CAD system for creating hierarchical quadrilateral polygon based surfaces. The interface of the system uses two hands to interact with the surface, with the left hand setting geometric and other contexts and the right hand manipulating the surface geometry. [19] Two handed interaction is increasingly used in immersive virtual environments. Research in two-handed interaction for virtual reality is ongoing in many places and is increasingly based on simultaneous manipulation of 3D objects with both hands.

3.3 Benefits of Two-handed Operation

After summarizing the research works above, some potential benefits are: [14]

1. The nonpreferred hand as a dynamic frame of reference – improve efficiency and intuitiveness. An important design principle is for the interface to preserve the mobile, dynamic role of the nonpreferred hand as a base frame of reference. The nonpreferred hand adjusts to and cooperates with the action of the preferred hand, allowing users to restrict necessary hand motion to a small working volume. Unlike a clamp (whether physical or virtual), the nonpreferred hand provides mobile stabilization.
2. Help incorporate the learnt two-handed skills and natural skills of designers. Most everyday tasks or working skills are two-handed. Two-handed interaction mimics the way people use their hands in everyday tasks. For 3D conceptual modelling, this could mean that working with the computer resembles working with traditional tools. Therefore, two-handed computer interfaces are potentially easier to operate than one-handed interfaces.
3. Reducing task completion time and the workload per hand.
4. Employing the kinesthetic sense. When working with two hands, people can sense where one hand is relative to the other (kinesthetic sense).

It seems that two-handed interface will naturally result in better performance than one-handed interface because two hands' input simply increases the DOFs of input. However, this is not always true. Kabbash et al have shown that two hands are not always better than one. [13] When a user is working on a complicated task, two-handed interface sometimes further complicates the situation by forcing the user to coordinate the actions of their hands. A problem of divided attention may occur here. Two-handed input is most likely to be effective when task integration of two hands is supported and divided attention is minimized. [17] There are two types of task integration – visual and conceptual.

In Fitzmaurice's bricks system, two bricks can be used to move and scale a rectangle with one brick as an anchor point and the other brick as a stretching point. In this case, the operations of moving and scaling are visually integrated and therefore divided attention is minimized. [7] Jacob provided evidence of conceptual integration effecting computer interaction. [12] Conceptual task integration is defined as when the user prefers to think of the operations of the two hands not as separate activities but as a single activity. For example, holding a chisel and hitting it with a hammer can be conceptually integrated into the single activity of "chiselling".

4 PROPOSED INTERFACE FOR CONCEPTUAL CAD

Some work has been done to demonstrate that the two-handed interface is beneficial to tasks such as assembly, operation planning and so on. However, few works of two-handed interface for geometric modelling can be found. While haptic feedback CAD exists, Freeform is already commercially viable, no commercial CAD system support two-handed interface. A two-handed haptic feedback interface is proposed to facilitate conceptual geometric modelling.

4.1 Architecture of the interface

The interface consists of one PHANToM haptic device and a Polhemus tracker. PHANToM device provides 6 DOFs of input and 3 DOFs of force output for the preferred hand. A Polhemus stylus is held in the user's nonpreferred hand. We chose heterogeneous devices for the two-handed interface for two practical reasons. Firstly, we already have a Polhemus tracker available. Secondly, we think both PHANToM and Polhemus tracker are fit for their specific tasks. Fig 1 shows the architecture of the interface of the C-CAD system.

Under the framework of Guiard's three principles of two-handed tasks, different tasks are designated to the preferred hand and the nonpreferred hand respectively. The preferred hand holds the stylus of the PHANToM to perform the actual modelling task such as carving and drilling. The haptic feedback from the PHANToM device makes the modelling process more intuitive and natural; moreover, it also helps to import the metaphor of traditional modelling into the conceptual CAD system. Tasks of the nonpreferred hand mainly include manipulating and rotating the virtual object, therefore the nonpreferred hand forms a dynamic reference frame for the modelling action of the preferred hand. For these tasks, haptic feedback for nonpreferred hand seems less important and single-handed force feedback for the interface satisfies the requirement of these tasks.
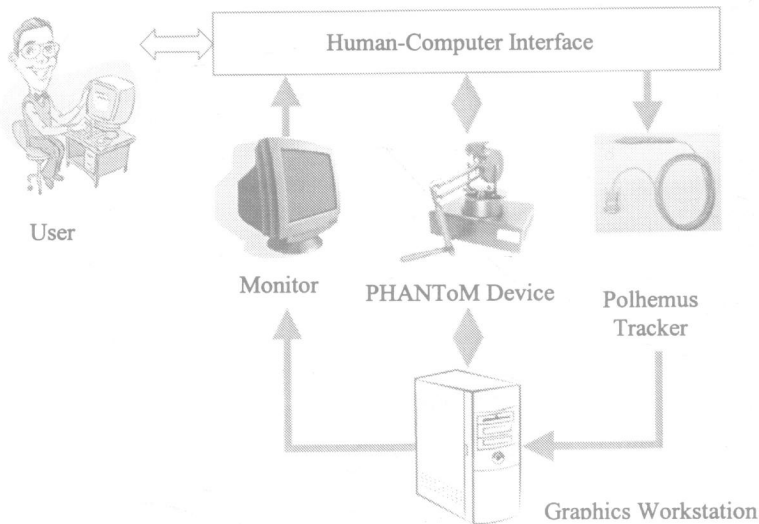

Fig 1. Architecture of the interface

However, in some other situations, where both hands cooperate to fulfil some integrated tasks such as scaling and curve editing (described later), haptic feedback for both hands probably is preferred because the force interaction can give users a realistic sense that they are working with two hands. For such integrated tasks, two hands are more reliant on each other than the situation where the preferred hand conducts the main modelling tasks with nonpreferred hand as dynamic reference frame. Because only one PHANToM device is available in the interface, the nonpreferred hand cannot feel the force feedback generated by the two-handed interaction and hence forms an artifact of unbalanced force feedback. Users may not be accustomed to this artifact. Two methods could potentially offset the artifact: 1) decrease the magnitude level of the force feedback during these tasks and therefore decrease the difference of haptic sense; 2) give some mass to the Polhemus tracker so that the tracker can provide some natural passive force feedback. How to satisfy these situations with the heterogeneous interface requires further study.

Although the SensAble supports dual PHANToM interface which is potentially better than the heterogeneous interface as we proposed, our interface is more practical and good enough to satisfy the requirement of conceptual modelling on the whole.

4.2 Building haptic simulation
Since PHANToM is chosen as the haptic device, we use GHOST (General Haptic Open Software Toolkit), a SDK bundled with PHANToM, to conduct force interaction. GHOST SDK is also supplied with OpenGL and GLUT libraries to manage graphics. The GHOST SDK is an object oriented toolkit in C++. It works as the "physics of touch" engine which takes care of the complex computations and allows developers to deal with high-level objects and physical properties like location, mass, friction and stiffness.

GHOST represents a haptic environment as a hierarchical collection of nodes, the basic building blocks used in GHOST. The GHOST library defines many types of nodes. The two essential node types are geometry and PHANToM nodes. These nodes represent the physical geometry of objects in the haptic scene and the PHANToM haptic interface respectively. Nodes are placed into a tree structure called a graph scene. In order to create a scene, the programmer must first define a root of the scene. Once all the nodes are defined and added to the scene, the haptic simulation process is started and performed by repeating the servo loop at a rate of 1 kHz. OpenGL properties are then defined, such as lighting and color, so that main loop can continuously display the scene.

4.3 Physical metaphor of traditional tools
While it is possible to interact with the CAD program using a 2D mouse or a 3D tracker without force feedback, the modelling experience would be much more realistic if the designer could use 'familiar tools' to input movements and actions to the program. In the conceptual CAD application, it's reasonable for the input devices to mimic the specific functions, the range of motion and the number of degrees of freedom relating to the traditional tools.

Different tools have different degrees of freedom, different modelling functions and different haptic feedback features. For clay modelling, shaving tools and spatulas are commonly used. Designers shave the extra clay away from the model with shaving tools while spatulas are used to lay lumps of clay onto the model. Tools have 6 DOFs before contact with a model.

When a shaving tool or a spatula is in contact with a model or a lump of clay, its movement is constrained by the shape of surface being edited. For these tools, the plane-probe haptic display model is ok. For other tools such as a hand drill, the situation is different. When a hand drill penetrates a model, its degrees of freedom are reduced to two and its movement is constrained on the axis of the drilling direction. Such a system should give accurate force simulation to user. Furthermore, pliers for both hands can be used to bend and edit curves.

4.4 Two-handed task integration
The two hands should be able to cooperate to do some integrated tasks.
1. Scaling. Two-handed interface enables a straightforward scaling operation that is analogous to stretching a piece of rubber in the physical world.
2. Line segment editing. Two-handed input enables additional flexibility when creating and editing line segments. Users are allowed to simultaneously manipulate two vertices of a line segment or a polyline. Although this interaction technique may seem simple, it can be very useful.
3. Curve editing. In conventional CAD systems, designers edit a curve usually by manipulating the curve's control points. This method is basically mathematical and the consequence of this method is hard to predict. Sometimes a movement of a single point leads to the modification of the whole curve. Two-handed curve editing could straightforwardly isolate the modified section so that it is between the two cursors. The position and tangent vector of the two cursors should be used to decide the shape of the modified curve section. Two-handed interface helps to hide the mathematic foundation away from the user.
4. Freeform deformation. Similar to two-handed scaling, this interface could also provide intuitive solid freeform deformation with the physical metaphor of elastomeric clay.

4.5 Current system structure
A basic two-handed interface has been set up. The program is built based on the GHOST structure. A child thread is generated from the main process to handle the information sent by the Polhemus tracker. The position and orientation information of the tracker is recorded by a program provided by Polhemus. A data pipe links the Polhemus program and the main process for the transference of Polhemus information. Two cursors are displayed on the screen; one is for Polhemus tracker and the other stands for PHANToM device. Since the program is based on GHOST, the user can feel the sense of touch when using the PHANToM cursor to explore the 3D space. Users can even feel the existence of the Polhemus tracker cursor by touching it with the PHANToM cursor. Fig 2 shows the current system structure.
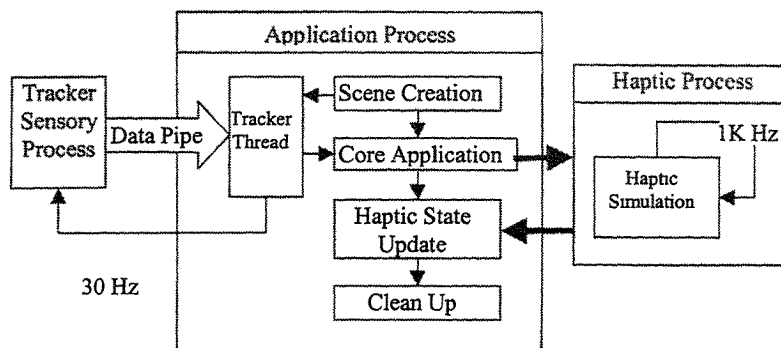
Fig 2. The current system structure

4.6 Future work

Two-handed curve editing will be researched and implemented on our two-handed environment. We will also implement some typical traditional modelling tools, like drill and scraper, etc. Some specific modelling methods with haptic feedback will be realized with respect to these tools. Experiments will be devised to study the usability of the two-handed and haptic interface. Complete usability studies might help to refine the interactions and determine the situations when they are most effective. Further, we believe that continued exploration of two-handed input for 3D modelling and editing operations with haptic feedback will result in additional effective interactions for C-CAD applications.

5 SUMMARY

Intuitive interaction and more sensory information such as haptic feedback could enhance performance of C-CAD. However, appropriate utilization of these new technologies to achieve a good result is a challenge. The WIMP (Window Icon Menu Pointer) interface mode has not changed a lot in the last decade. As our tasks become more complex and demanding we may benefit by devising some intuitive two-handed interface. There are many technical challenges associated with creating a good simulation system with haptic feedback. Currently there are very few haptic devices available and none is sufficiently inexpensive to be widely used. However, to enjoy the full benefits of a C-CAD system and employ the skills designers would use when interacting with a real model and tools, it is necessary to incorporate a sense of touch into CAD applications.

REFERENCES

[1] Bier EA, Stone MC, Pier K, Buxton W, DeRose TD, (1993) Toolglass And Magic Lenses: The See-Through Interface, Computer Graphics Proceedings. ACM, 73-80.
[2] Burdea, Grigore C. (1997) Force and Touch Feedback for Virtual Reality, John Wiley & Sons, Inc.

[3] Chen, E., & Marcus, B. (1998). Force Feedback for Surgical Simulation. Proceedings of the IEEE, 86(3), 524-530

[4] Cohen, J., M. Lin, D. Manocha and M. Ponamgi, (1995) I-COLLIDE: An Interactive and Exact Collision Detection System for Large-scale Environments, Proceedings of ACM Interactive 3D Graphics Conference, ACM NY, 189-196

[5] Dachille F, Qin H, Kaufman A, El-Sana J, (1999), Haptic Sculpting Of Dynamic Surfaces, Proceedings of the 1999 symposium on Interactive 3D graphics, ACM, 103-227.

[6] Dani, T H and Gadh, R. (1997) Creation Of Concept Shape Designs Via Virtual Reality Interface, Computer-Aided Design, vol. 29, no. 8, 555-563.

[7] Fitzmaurice GW, Ishii H; Buxton W, (1995), Bricks: Laying the Foundations for Graspable User Interfaces, CHI'95 Conference Proceedings. ACM, 442-449.

[8] Frank A O., Twombly I A, Barth and Smith J D. (2001) Finite Element Methods for Real-Time Haptic Feedback Of Soft-Tissue Models In Virtual Reality Simulators, Proceedings of the Virtual Reality 2001 Conference, IEEE, 257-263

[9] Gregory AD, Ehmann SA, Lin-MC, (2000) inTouch: Interactive Multiresolution Modelling And 3D Painting with a Haptic Interface, Proceedings IEEE Virtual Reality 2000, IEEE Computer. Soc, Los Alamitos, CA, USA, 45-52.

[10] Guiard, Y. (1987). Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. In Journal of Motor Behavior, 19(4), 486-517.

[11] Hinckley K, Pausch R, Goble J C., and Kassel NF (1994) Passive Real World Interface Props for Neurosurgical Visualization. Proceedings of CHI, 452-458.

[12] Jacob, R. J. K., Sibert, L., McFarlane, D. et Mullen, P. (1994) Integrality and Separability of Input Devices, ACM Transactions on Computer-Human Interaction. 1(1), 3-26.

[13] Kabbash, P., Buxton, W., and Sellen, A., (1994) Two-Handed Input in a Compound Task. Proceedings of the CHI'94 Conference on Human Factors in Computing Systems, 417-423.

[14] Maarten Gribnau, (1999) Two-handed Interaction in Computer Supported 3D Conceptual Modelling

[15] William M, Randolph S, Finch M, Verth JV, Russell MT, (1996) Adding Force Feedback to Graphics Systems: Issues and Solutions, Computer Graphics: Proceedings of SIGGRAPH '96, 447-452.

[16] McDonnell K T, Qin H and Wlodarczyk R A, (2001) Virtual Clay: A Real-time Sculpting System with Haptic Toolkits, Proceedings Of The 2001 Symposium On Interactive 3D Graphics, ACM, 179 - 190.

[17] Owen R, Kurtenbach G, Fitzmaurice G, Baudel T, Buxton W, (2001), Bimanual Manipulation in a Curve Editing Task, URL: http://www.billbuxton.com/Curve Match.html

[18] www.sensable.com

[19] Shaw CD, Green.M, (1997) Thred: A Two-Handed Design System, Multimedia Systems Journal, 2(5), 126-39.

[20] Shimoga, Karun B. (1993) A Survey of Perceptual Feedback Issues in Dexterous Telemanipulation: Part I. Finger Force Feedback, IEEE Virtual Reality Annual International Symposium VRAIS '93; Seattle, 263-270.

[21] Yamamoto K, Ishiguro A, Uchikawa-Y, (1993), A Development of Dynamic Deforming Algorithms For 3D Shape Modelling With Generation of Interactive Force Sensation, IEEE Virtual Reality Annual International Symposium. IEEE, 505-511.

[22] Zeleznik RC, Forsberg AS, Strauss PS, (1997), Two Pointer Input for 3D Interaction, Proceedings 1997 Symposium on Interactive 3D Graphics. ACM, 115-120.

Towards a systematic theory of axiomatic design review (STAR)

G Q HUANG
Department of Industrial and Manufacturing Systems Engineering, University of Hong Kong, Hong Kong
Z H JIANG
Department of Industrial Engineering, Shanghai Jiao Tong University, People's Republic of China

SYNOPSIS

This paper extends Suh's well-known Theory of Axiomatic Design (TAD) to formulate a Systematic Theory of Axiomatic design Review (STAR). Design Review is defined as the mapping process between the Design Objects {DO} domain and the Review Criteria {RC} domain. This resulting STAR framework enables us to develop a systematic and rigorous methodology for more efficient and effective DR, to develop and deploy design review web applications as enterprise portals on the Internet/intranets, and to generate specific design review methods and techniques.

Keywords: Axiomatic Design, Fuzzy Design Review, Enterprise Portal

1. INTRODUCTION

Design review (DR) is used throughout the product development process to evaluate the design in terms of costs, quality and delivery, to ensure that most suitable knowledge and technology are incorporated into the design, and to resolve possible problems instead of passing them downstream (Dillon, 1996; Voigt, 1996). DR is mandatory in ISO9001 for design verification across the entire product development process. A design review activity must take place at the end of the Design Output Stage (Schoonmaker, 1996). However, ISO9000 does not specify how DR should be conducted.

Despite the industrial importance of DR, our comprehensive literature search reveals that relatively little has been done as compared with other design-related activities such as concept generation and evaluation. Perhaps the most intensive coverage over the DR topic is the work by Ichida (1989). The basics of DR have been investigated in relative depth. One of the main strengths of this work is its inclusion of six case studies concerning the industrial DR practices. Although these case studies do not follow a common DR framework, they do indicate some common elements of good DR practices. This indicates the feasibility of developing such a common framework and incorporating it into computer software package that are sufficiently flexible for different organizations to customize.

Interestingly and surprisingly, majority of the Product Data Management (PDM) systems do not address explicitly and intensively the DR issue. Engineering change management

(ECM) and design releases are usually considered as essential parts of PDM systems. However, few facilities are provided explicitly to support the DR activities. Check-in and check-out of design documents using vaulting controls can hardly be considered as design release, review and revision management.

Some forward-looking companies have applied web sites to support design reviews. For example, Pfund (2001) reports that a web site is set up for each design review and members of the review committee assigned by the project manager are able to access the design documents at the web site. This type of design review practice is expected to prevail in the near future. Unfortunately, Such web sites usually adopt the ad-hoc DR approach, rather than rigorous DR theory or systematic DR framework.

The above discussions highlight the serious lack in systematic theories or frameworks for supporting DR activities and practices from the viewpoints of both the research community and the practitioner community. The central subject matter of the research reported in this paper is to tackle this. The long-term aim is three-folded: (1) to develop an overall methodology for more efficient and effective DR in new product development process, (2) to demonstrate the framework through a prototype web-based platform on the Internet/intranets using the web technology, and (3) to investigate and develop a suite of methods and techniques and to package them into ready-to-use tools in a systematic way. This paper presents an overview of the recent developments that have been made in the above directions.

2. STAR: SYSTEMATIC THEORY OF AXIOMATIC DESIGN REVIEW

In the Theory of Axiomatic Design (TAD), Suh (1990) defines design as the mapping process between the Functional Requirements (FRs) in the functional domain and the Design Parameters (DPs) in the physical domain. Conceptually, the design process can be interpreted as a process of involving choosing the right set of DPs to satisfy the given FRs. Mathematically, the mapping is expressed as Design: $\{FR\} \rightarrow \{DP\}$, or more specifically, $\{FR\} = [A] \times \{DP\}$, where $\{FR\}$ is the functional requirement vector, $\{DP\}$ is the design parameter vector, and $[A]$ is the design matrix.

Axiomatic design guides designers in using all of their existing design tools and software to arrive at a successful new design, or to diagnose and correct an existing design. The latter application in "diagnose and correct" has motivated the authors to explore the possibility of extending the TAD into a framework for design review. The resulting framework is called STAR (Systematic Theory for Axiomatic design Review).

In essence, STAR is defined as the reverse mapping from the DP domain to the FR domain, that is, Review: $\{DP\} \rightarrow \{FR\}$. Here, the FR domain should be generalized as the design review criterion domain (simply the RC domain). The RC domain can be established from the functional (FRs) and other types of design requirements (DRs), that is, $\{RC\} \subseteq \{DR\}$ and $\{FR\} \subseteq \{DR\}$.

Figure 1 is a diagrammatical representation of STAR. The left-hand side of Figure 1 shows TAD while the right-hand-side of Figure 1 shows STAR. The reason why TAD is included in STAR framework is simple. The outputs (descriptions of design objects and parameters) from and the process (background information associated with design objects) of TAD are inputs to STAR.

Basically speaking, TAD and STAR share the same design workspace (DW), namely $\{DO\}$ in STAR is the same as $\{DP\}$ in TAD except that $\{DP\}$ is dynamically changed after certain design tasks are performed. $\{DO\}$ in STAR is relatively static until redesign tasks are

approved and implemented. The main objective of TAD is to generate as many alternative design concepts as possible by mapping between the design requirements and design objects. All design concepts are feasible if they fall within the constraints of the design specifications {DS}. The aim of STAR is, however, to evaluate the capabilities {DC} of resulting design concepts and objects by effective mapping methods onto the design review criteria, and to identify the most capable design concept or to determine if a design concept is sufficiently capable up to a defined threshold based on a collection of given DR axioms.

For STAR to become pragmatic from industrial practitioners' viewpoint, its key abstract concepts must be clarified and explained:

- Domains,
- Hierarchical decomposition and representations,
- Zigzag Mapping, and
- Axioms.

Because STAR has been derived from Suh's TAD, the above concepts are therefore the same as those used in TAD. Because of limited space here, general discussions on the definitions and meanings of these terms are not included here. Interested readers would find further discussions elsewhere (http://www.axiomaticdesign.com/, Huang, 2001a).

DR	– Design Requirement
DS	– Design Specification / Constraint
DO	– Design Object / Parameter
RC	– Review Criteria
DC	– Design Capabilities / Performance Measures
m	– Number of Design Requirements
n	– Number of Design Objects
k	– Number of Review Criteria
	– Domains
	– Mapping

$$\{DR\} = [A]\times\{DO\}$$
TAD mapping to generate DOs

$$\{RC\} = [B]\times\{DO\}$$
STAR mapping to estimate DCs
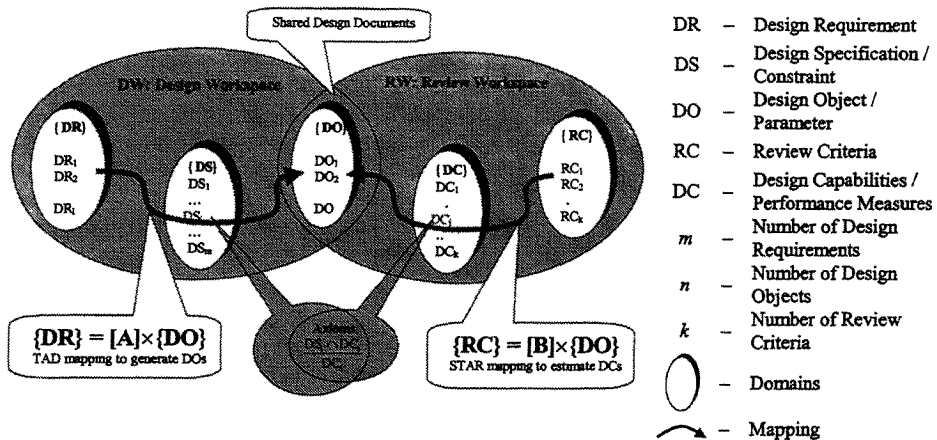
Figure 1 Overview of Systematic Framework of Axiomatic design Review.

3. *CyberReview* AS DESIGN REVIEW PORTAL

The STAR framework outlined in the preceding section has been used as a basis for the development of a web application, called CyberReview, dedicated to design review. Figure 2 shows an overview of the CyberReview system. CyberReview is deployed as a web site that serves two main groups of users. One is the group of designers or product development team who submit designs (in the form of documents) for review. The other group is the committee established for reviewing a design project at certain stage. CyberReview provides a repository

for archiving the design and review documents and supports activities involved in DR for both groups. The rest of this section briefly summsrises the components, design and implementation considerations, and the general procedure of CyberReview. Detailed discussions can be found elsewhere (Huang, 2001b).
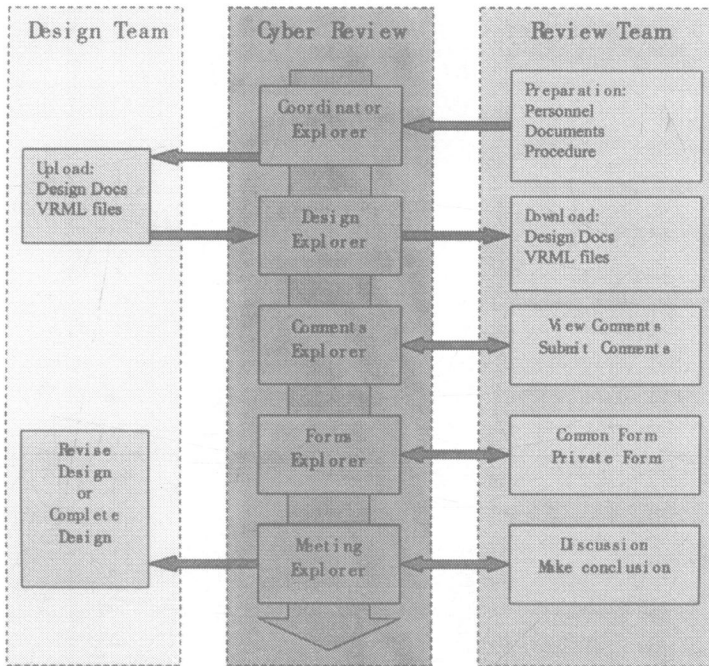
Figure 2 CyberReview framework.

As shown in the middle of Figure 2, a rich set of facilities is provided to facilitate the entire design review process. The system includes the following main components:

- CyberReview Explorer. This is the entry component to which all the other components and facilities are attached.
- Project Explorer. This lists all the design review projects that a particular user is involved.
- Review Coordinator. This component provides a set of facilities for the project manager to plan and manage the activities and resources involved in the design review process, in particular, for establishing the review committee, defining design documents, and preparing review documents.
- Design Explorer. This component is provided for the product development and design team to upload the product design onto the CyberReview database for future access. The review team uses the similar facilities to download design documents for review.
- Comments Explorer. This is an electronic discussion forum for review team members to submit their comments freely concerning the design documents submitted by the design team.

- Review Explorer. This collects all the online and offline forms and documents that are prepared during the review process.
- Meeting Explorer. This is a sophisticated module providing a variety of facilities to support holding review meetings for both the chairperson (project manager) and the team members before, during and after the meeting.
- BOM Explorer. This is a module associated with the design explorer. BOM is treated as a special type of design document required for design review when it becomes available. VRML files, comments and reviews may be directly related to BOM items.
- VRML Whiteboard. This is a whiteboard based on the VRML display of product features. This module would be necessary during the review process.
- CyberReview *Utilities*. In addition to the above main modules, CyberReview provides a set of general utility facilities such as user registration, searching for necessary information, etc.
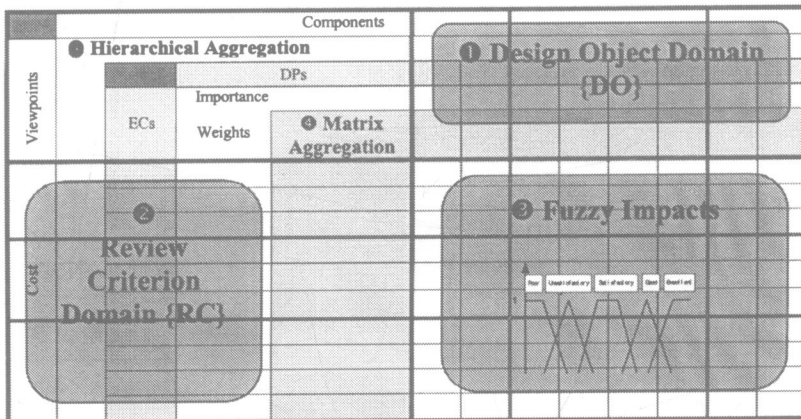
Figure 3 Overview and steps of FuzzySTAR.

4. FUZZYSTAR: FUZZY SET THEORY FOR AXIOMATIC DESIGN REVIEW

The STAR framework can be actually used to generate further specific design review techniques. One such generalization is to incorporate fuzzy set theory. The resulting method is termed as FuzzySTAR (Fuzzy Set Theory for Axiomatic design Review). Figure 3 presents the overview of the FuzzySTAR worksheet that also reflects the main components and the logical flow of the method. Two case studies have been carried out using the FuzzySTAR method (Huang and Jiang, 2001a,b) while detailed discussions are omitted because of limited space here. Five key steps involved in the method are summarized as follows:

Step 1 – Define the Design Objects {DO} Domain. A Design Object DO is defined as a triplet (DP, TV, DW). The first element in the DO triplet is a Design Parameter (DP) that describes and defines a product design object collectively with other DPs. The second element, TV, of the DO triplet is the value (or range of values) of the corresponding design

parameter DP. The third element, DW, in the DO triplet is the weight or rating based on certain aspect (e.g. the cost of changing the TV of the corresponding DP).

Step 2 – Define the design Review Criteria {RC} Domain. A Review Criterion RC is also expressed as a triplet (EC, DC, FW). The first element in the {RC} domain is an Evaluation Criterion (EC). It is a description of an objective that Design Objects intend to achieve. The second element, DC (Design Capability), of the RC triplet is the value (or range of values) of the corresponding evaluation criterion EC. This value or range of values describes the capability of the design objects to meet the intended design requirements. The third element, FW, in the RC triplet is the weight or rating that the design objects collectively achieve against this RC (or EC).

Step 3 – Evaluate the Impacts of DPs on Ecs. There are two ways of obtaining the FW of a RC. One method is to derive the FW from fuzzy relationships between DPs and the EC through certain aggregation algorithm. This method is called direct fuzzy mapping method because the FW is obtained directly from the membership functions established between the EC across the TVs of the DPs. The other method is to derive the FW from a membership function established for the EC based on its DC. This method is called indirect or tandem fuzzy mapping method.

Step 4 – Aggregate with Fuzzy Operators. Individual fuzzy impacts of DPs on ECs are processed to obtained intermediate and final results.

Step 5 – Defuzzificate the Results and Observe for Further Investigations. The results, whether individual, intermediate or final, from the aggregations are fuzzy numbers. They must be properly interpreted in order to draw meaningful conclusions.

5. CONCLUDING DISCUSSIONS

This paper has proposed a systematic framework for design review in the product development and realization process by extending the Suh's Theory of Axiomatic Design (TAD). The resulting Systematic Theory of Axiomatic design Review (STAR) is an important contribution to the scarce design review literature, bridging the gap between the industrial urgency and the lack of rigorous theoretical approaches.

STAR is well-structured for computerization. In fact, a prototype web-based system, CyberReview, has been developed for collaborative product design review in the extended enterprise environment. CyberReview provides various Internet and web facilities for supporting both the product design and design review teams to carry out design review activities. Once CyberReview is deployed as a centralized design review portal on the web, it allows simultaneously data access and processing while paper-based and standalone systems only allow single user access. Design review data are shared and communicated between all parties concerned immediately after they enter the system. Therefore, the practice of design review is reengineered from what is traditionally sequential to a parallel process. The parallel execution of design review activities reduces the review throughput time and the amount of paperwork dramatically to a minimum level. This parallel design review contrasts sharply with the traditionally sequential design review practice. This sequential process is usually very tedious and the review cycle time is very long, becoming very inefficient especially when some external members such as key customers and suppliers are involved from other geographical regions.

In addition to being well-structured, STAR is generic too. It can be not only easily and freely substantiated and/or instantiated to suit specific design problems/projects, but also be

able to accommodate a wide variety of specific design review methods and techniques. FuzzySTAR is just one of such methods developed under the STAR umbrella for the review team members to collaboratively evaluate a design. In FuzzySTAR, past experience and insights are expressed as membership functions between design parameters and evaluation criteria.

In order to maximize the benefits of this web-based STAR approach to collaborative design review, further investigations must be conducted in two general directions. One is to improve the usability and functionality of the web-based CyberReview framework to a professional level for real-life industrial applications. The other direction is the development and collection of more rigorous and pragmatic design review techniques and methodologies so that they can be formulated as tools or toolkits within the STAR and CyberReview environments.

ACKNOWLEDGMENTS

REFERENCES

Dhillon, B.S. (1996), *Engineering Design: A Modern Approach*, Richard D. Irwin, a Time Mirror Higher Education Group, Inc. Company.

http://axiom.mit.edu/

http://www.axiomaticdesign.com/

Huang, G.Q. (2001a) "STAR: Systematic Theory of Axiomatic design Review and Web-based design review portal", Submitted to *Robotics & Computer-Integrated Manufacturing*.

Huang, G.Q. (2001b) "CYBERREVIEW: Web-Based Collaborative Product Design Review", Submitted to Journal of Engineering Design.

Huang, G.Q., Jiang, Z. H. (2001a) "FuzzySTAR: Fuzzy Set Theory of Axiomatic Design Review", Submitted to International Journal of Engineering Design, Analysis and Manufacturing.

Huang, G.Q., Jiang, Z. H. (2001b) "Fuzzy Set Theoretic Approach to Design Review of Fuel Pumps", Accepted by IMechE Proceedings Part B Journal of Engineering Manufacture.

Huang, G.Q., Jiang, Z. H. (2001c) "Web-Based Design Review of Fuel Pumps Using Fuzzy Set Theory", To be submitted to ASME Transactions.

Ichida, T (1989), *Deizain Rebyu Jireishu*, Productivity Press, Portland, Oregon.

Pfund, A.P. (2001) "Design on an International Project", In: *Proceedings of 2001 ASME DETC, International Issues in Engineering Design*, Pittsburgh, USA, September 2001.

Schoonmaker, S.J. (1996*), ISO 9001 for Engineering and Designers*, The McGraw-Hill Companies, Inc.

Suh, N.P. (1990) *The Principles of Design*, Oxford University Press, New York.

Voigt, E.C. (1996), Product design review, a method for error-free product development, Productivity Press, Portland, Oregon.

Image-based modelling for reverse engineering of large objects

N M ALVES and **P J S BARTOLO**
Mechanical Engineering Department, School of Technology, Polytechnic Institute, Leiria, Portugal
J C FERREIRA
Mechanical Engineering Department, Instituto Superior Técnico, Lisbon, Portugal

SYNOPSIS

In the engineering field there is no effective process for the rapid generation of computer models from existing large objects. Commonly, engineers use contact probe or laser-scanning systems to perform reverse engineering activities, but these machines are not adequate for large objects such as buildings, cars, planes, etc. This study aims to solve this limitation through the replication of the human vision process into a computer programme, using an image-based modelling approach. The importance of this new tool for reverse engineering and design activities are highlighted here. This approach uses two-dimensional photos to generate three-dimensional computer models that can be used to produce sophisticated photo-realistic renderings or to generate surface models for simulation purposes and rapid prototyping applications. The proposed image-based modelling system, intends to be a very robust, versatile and flexible computer tool. The general structure of this computer tool is described in this paper as well an analogy with the human process of visual perception.

1 INTRODUCTION

Over the past two decades, computational design and redesign activities have been extensively studied and several computer-aided design implementations developed (1). However, the traditional methods to build digital models from large objects are too labour-intensive processes, implying the previous digitization of the object or existing drawings or the modification of existing CAD data, if available. Though, there is no effective process for the rapid generation of computer models from existing large objects. Currently, engineers use contact probe or laser-scanning systems to perform such work, but these machines are not adequate for large objects, besides being time-consuming and expensive (2).

This research intends to solve this limitation through the replication of the human vision process into a computer programme as a part of a major project called "the wheal of design" (3-6). The whole process can be described as the integration of both the human vision process and the comparable performance of computers operating similarly to a biological system,

through a process called Biomimetics (7). Broadly, the "wheal of design" project (see fig 1) is an interdisciplinary effort, aiming to develop computer-based procedures and evolutionary design strategies to complement and strengthen the design process, exploring possible improvements to the cooperative design environment that will benefit of the diverse opportunities provided by information technology.
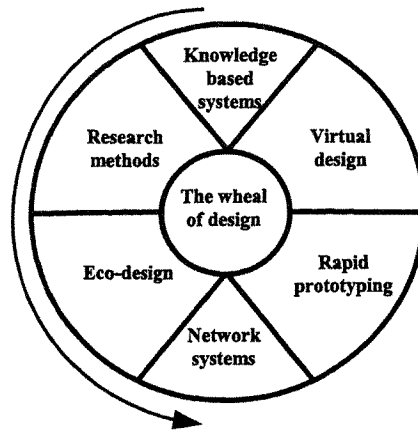
Figure 1. The wheal of design

2 VISUAL PERCEPTION

The word vision embraces many aspects stemming from our experience, as humans use vision to identify objects, shapes, colours, movement, etc. The theory of visual perception suggests that the human vision is organised as an information-processing system comprising different stages (8). The *initial* perception step takes place in the retina, a multilayered sheet of photoreceptors and neurons. The visual image is then focused most sharply on a small area of the retina called fovea. Afterwards, the retina collects the light and converts the light energy into electrical nerve impulses that are transmitted to the brain (9) that will process these impulses through the visual cortex giving information about what we are seeing. In some visual areas, the neurons are organized in an orderly fashion called *topographic or retinotopic mapping*, forming a 2D representation of the visual image formed on the retina in such a way that neighbouring regions of the image are represented by neighbouring regions of the visual area. Additionally, the corresponding points in two slightly different images on the retinas, left and right eyes, are determined by the visual cortex in a process called *binocular stereopsis* (8). Stereopsis corresponds to that extra sense of solidity and depth that is experienced by humans when they use the two eyes rather than only one. Each eye captures its particular view and the two separate images are sent on to the brain for processing. The two images will arrive simultaneously to the back of the brain and will be united into one picture. The mind combines the two images by matching up the similarities and adding in the small differences and the final image is a three-dimensional *stereo* picture. Thus, visual perception can be organised into three regions as indicated in fig 2:

- The **physical world** of objects (environment)
- The **visual stimuli**, which gives rise to visual events
- The **brain**
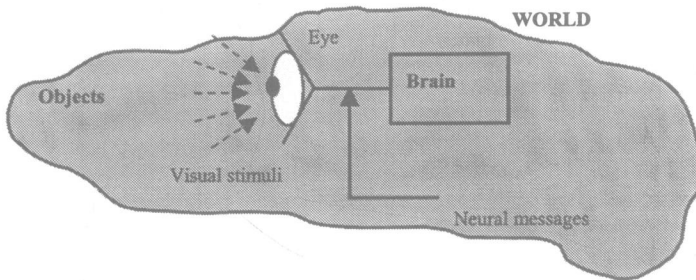
Figure 2. Visual perception

It is this visual perception's organisation scheme that is being computationally developed through a computer software briefly described in this paper.

3 COMPUTER IMPLEMENTATION

The computer reconstruction system has been developed in C++ under a computational code called Three-Dimensional Reconstruction System (3DRS), using a commercial digital camera as the retina of the computer to take two-dimensional (2D) photos of existing objects (3,10). These 2D photos are then used by a robust computer routine, working as the visual cortex of the computer that will convert observed photos (input) into tri-dimensional (3D) computer models (output). Thus, these 3D models can be used to produce very sophisticated photo-realistic renderings, to generate surface models or three-dimensional meshes in triangular STL format for simulation purposes and rapid prototyping applications.

Under the ongoing research project called the "wheal of design", this image-based modelling system coupled with rapid prototyping will allow the implementation of a reverse design methodology, facilitating the re-design of an existing product. This way, the shape of an existing product is captured and digitised through reverse engineering (see fig 3), creating the correspondent surface model. Finally, this model can be manipulated towards the production of the model of the new product using rapid prototyping techniques.

The flowchart indicated in fig 4 represents a plain conceptualisation of the reconstruction process from 2D digital photos that will enable to explain how the generated digital models can be used for simulation, visualisation and optimisation purposes during the redesign process. Fig 5 shows the working area of the software.

The first two initial parts of this image-based modelling computer code, aiming to create three-dimensional objects from a set of two-dimensional images taken by uncalibrated cameras, are described the camera calibration and the binocular stereopsis. The purpose of the camera calibration intends to establish the projection from the 3D world coordinates to the 2D image coordinates. The binocular stereopsis performed through the epipolar transformation,

will enable to identify the points in two or more images that are projections of the same point in the world. Both procedures are briefly described below from a mathematical view.
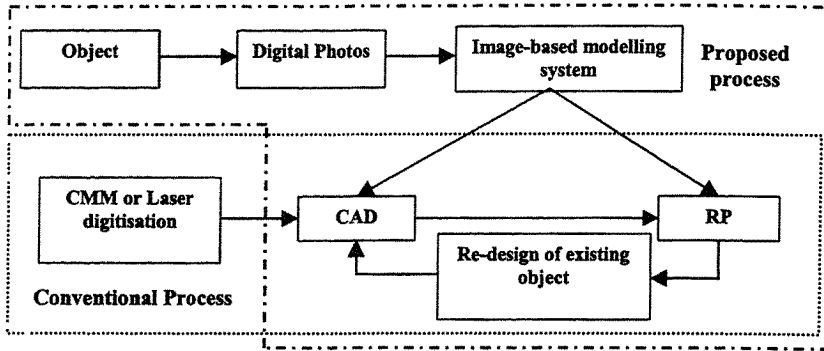
Figure 3. The reverse engineering process. Comparison between the conventional approach and the proposed methodology
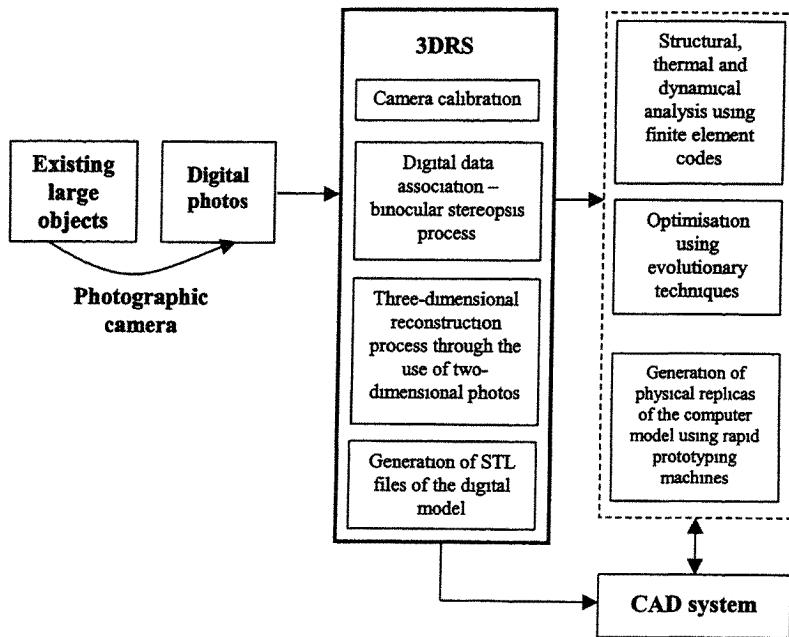
Figure 4. The structure of the proposed image-based modelling system and its relationships with other CAD and simulation codes

I002/010/2002 © IMechE 2002
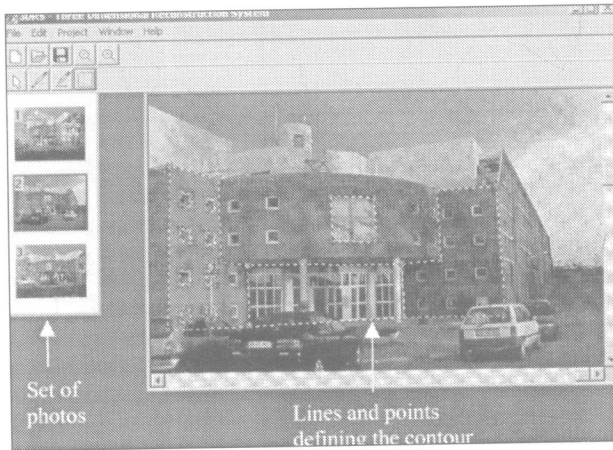
Figure 5. Working area of the proposed computational code

3.1 Camera calibration

The geometric model indicated in fig 6 is used for the camera calibration procedure. It consists of a plane called retinal plane or image plane in which the image is formed through an operation called perspective projection. A coordinate system (X,Y,Z) is defined for the three-dimensional space and (p_x, p_y) for the retinal plane, where (p_{xo}, p_{yo}) are the coordinates of the intersection between the optical axis and the image plane. Such point, located at a distance $\Im$, the focal distance between the centre of the image and the projection centre, is used to form the image p in the retinal plane of the 3D point P.

According to (3,11), the relationship between image coordinates and global 3D coordinates can be written in the following way:

$$\lambda \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} = [T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where λ represents a scale factor and [T], the perspective transformation matrix, can be expressed by the following matrix:

$$[T] = \begin{bmatrix} -\chi_{px}\Im & 0 & p_{xo} & 0 \\ 0 & \chi_{py}\Im & p_{xo} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2}$$

where χ_{px}, χ_{py} are the horizontal and vertical scale factors, respectively. For simplification purposes we can consider $\varepsilon_{px} = -\chi_{px}\Im$, $\varepsilon_{py} = \chi_{py}\Im$, so [T] becomes:

$$[T] = \begin{bmatrix} \varepsilon_{px} & 0 & p_{xo} & 0 \\ 0 & \varepsilon_{py} & p_{yo} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3}$$
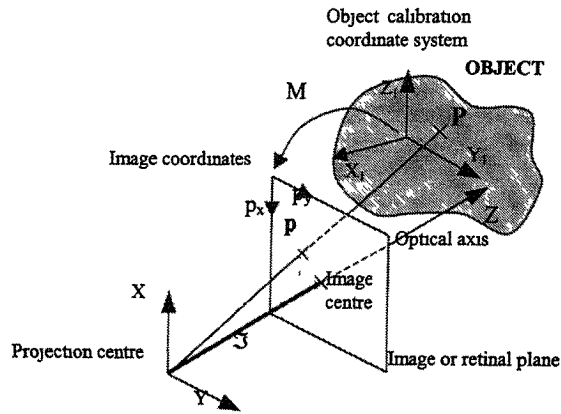
Figure 6. Generalised geometric model for the camera calibration

The parameters p_{xo}, p_{yo}, ε_{px} and ε_{py} are called intrinsic as they do not dependent on the position and orientation of the camera in space (11). However, for the camera calibration procedure, we will need to know this position and orientation. This means that the camera must be considered as a system that depends upon the intrinsic and extrinsic parameters (3,11). In order to include the extrinsic parameters we must study the relationship between the image coordinates and the 3D coordinates of the object, expressed in terms of a local coordinate system, (X_1, Y_1, Z_1), named the object calibration coordinated system (3). Thus, we must consider the transformation matrix [M], apart from relating the coordinates (X_1, Y_1, Z_1) with (p_x, p_y):

$$[M]=[T]\,[G] \tag{4}$$

$$[G]=\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where [G] is a matrix expressing the orientation and position of the camera, r represents the rotation and t the translation.

Using equations (3) and (5) in equation (4), we obtain

$$[M]=\begin{bmatrix} \varepsilon_{px}r_1+p_{xo}r_3 & \varepsilon_{px}t_1+p_{xo}t_3 \\ \varepsilon_{py}r_2+p_{yo}r_3 & \varepsilon_{py}t_2+p_{yo}t_3 \\ r_3 & t_3 \end{bmatrix} \tag{6}$$

with $\quad r_1=\begin{bmatrix} r_{11} & r_{12} & r_{13} \end{bmatrix} \qquad r_2=\begin{bmatrix} r_{21} & r_{22} & r_{23} \end{bmatrix} \qquad r_3=\begin{bmatrix} r_{31} & r_{32} & r_{33} \end{bmatrix}$

Finally, we obtain the following relationship between (X_1, Y_1, Z_1) and (p_x, p_x), similarly to equation (1):

$$\lambda \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} = [M] \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix} \qquad (7)$$

Then, the camera calibration can be depicted as the process of estimating the intrinsic and extrinsic parameters of the camera, a two-step process comprehending:

- evaluation of [M]
- estimation of the intrinsic and extrinsic parameters from [M].

3.2 Binocular stereopsis

In this study, the binocular stereopsis reconstruction process performed by the humans is simulated through the use of the "so-called" epipolar geometry, describing the relationships that exist between two images. The epipolar geometry considers two cameras as shown in fig 7.
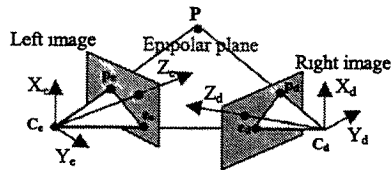
Figure 7. Epipolar geometry

The epipoles (e_e and e_d) are the intersection point of the line joining the optical centers (C_e and C_d) with the image planes. The epipole is the image in one camera of the optical centre of the other camera. Moreover, the plane defined by an image point coupled with the optical centres is named epipolar plane.

Each camera has a specific coordinate system, $X_e\, Y_e\, Z_e$ and $X_d\, Y_d\, Z_d$, respectively for the left and right side. Consequently, a generic image point P, has projections p_e and p_d on the left and right images, respectively. The evaluation of P is performed through the evaluation of the intersection between the lines defined by (C_e, p_e) and (C_d, p_d).

As observed above, the problem of recovering three-dimensional information from a set of photographs is essentially the correspondence problem. Given a point in one image, we should find the corresponding point in each of the other image. Once the necessary correspondence has been identified, the evaluation of the depth (the third dimension) is simply a matter of geometry. The automatic recovery of three-dimensional information from two-dimensional photographs is a major focus in this research and represents an important step forward in relation to conventional photogrammetry, which requires a large amount of human input to identify corresponding features in multiple photographs. Therefore, the correspondences are used to both recalculate the extrinsic parameters of the camera and determine the three-dimensional position of selected features.

4 CONCLUSIONS

The research work described here represents the initial development stage of a new computational tool for the rapid generation of computer models from existing large objects. The software will replicate the human vision process and is based on image-based modelling techniques. Until now, only the camera calibration and binocular stereopsis stages have been developed. This computational software is being further developed to become a very robust, versatile and flexible computer tool, making possible to perform reverse engineering of large objects. In the future, this tool can be applied to different fields, such as architecture, anthropology, medicine or engineer.

REFERENCES

1 **Jarvis, R.A.** (1983), A perspective on rage finding techniques for computer vision, IEEE Trans. Patter Anal. Machine Intell. PAMI-5, pp. 123-139.

2 **Bidanda, B., Motavalli, S.** and **Harding, K.** (1991) Reverse engineering: an evaluation of prospective non-contact technologies and applications in manufacturing systems, Int. J. Computer Integrated Manufacturing, 4, pp. 145-156.

3 **Alves, N.M.** and **Bartolo, P.J.S.** (2002) A new image-based modeling system to support architectural redesign activities, Proceedings of the 2nd International Postgraduate Research Conference in the Built and Human Environment, The University of Salford, pp. 322-333.

4 **Bartolo, P.J.S.** and **Bartolo, H.M.** (2001) Concurrence in design: a strategic approach through rapid prototyping, Proceedings of CIB World Building Congress, Wellington, NZ, 4, pp. 54-64.

5 **Bartolo P.J.S.** and **Bartolo, H.M.** (2001) The use of rapid prototyping techniques and internet tools to support a knowledge integrated design process, Proceedings of COBRA RICS, Glasgow Caledonian University, Glasgow, 2, pp. 742-51.

6 **Bartolo, P.J.S.** and **Galha, H.M.** (2000) A concurrent design methodology through the use of rapid prototyping, Research by Design, Delft University Press, Delft, pp. 134-140.

7 **Jeronimidis, G.** (2000) Biomimetics: lessons from nature for engineering, Proc. Instn. Mech. Engnrs.

8 **Gordon, I.E.** (1989) Theories of visual perception, John Wiley, Chichester.

9 **Audesirk, T.** and **Audesirk, G.** (1999) Biology: life on earth, Prentice-Hall, New Jersey.

10 **Alves, N.M.** and **Bartolo, P.J.S.** (2002) Rapid Generation of Computer Models From Large Objects, Time Compression Technologies, (to appear).

11 **Faugeras, O.** (1993) Three-Dimensional Computer Vision: A Geometric Viewpoint, MIT Press, CA.

Design animation for rapid product development

G LIU, H T LOH, A C BROMBACHER, and H S TAN
Design Technology Institute, National University of Singapore, Singapore

ABSTRACT

This paper proposed a methodology of exploiting design animation for rapid product development (RPD). The objective is to investigate a fast and economical communication media for all design participants. By using commercial software: *solidworks* and *multimedia studio pro*, a live design animation AVI file can be conveniently created and easily shared by the whole product development group. The efficacy of using this AVI file for improving the efficiency of design communication is demonstrated by a case study.

1 INTRODUCTION

The continual introduction of new products is the life-blood of a successful company as this is necessary for it to keep ahead of its competitors and withstand the competitive pressure. To do so in a timely manner requires the company to adopt a rapid and cost-efficient product development process. By its very nature, however, such a process is iterative and the ability to reduce the number of cycles is a key to reducing the overall process time.

In the development of new products, particularly where complicated assembly is introduced, the design tasks are normally divided up and performed by several groups. The developing process of a new product involves not only the work of designers, manufacturers but also the collaboration from suppliers and marketing department as well, which makes the development process time-consuming and labor-intensive. As time is presently the most challenging parameter [1][2], the development process needs to be accelerated and therefore, how to offer a fast and convenient communication environment for all design participants becomes important. Traditionally, collaborative design is mainly based on design meetings, where the feedbacks collected for the original design are summarized into documents and designers communicate their design ideas by reviewing the corresponding manuals and documents, as illustrated in Fig. 1. This document-based communication is not only time-consuming,

inconvenient but also tedious and of low efficiency, because it involves too many human and environmental factors, e.g. designers have to read piles of documents to understand the detail of the product design, the meeting place and the meeting schedule have to be properly planned so that the meeting is possible for all related members. All these aspects indirectly increase the difficulty for designers, professional requirements for technicians and thus delay the time to launch a new product. To speed up the design communication in new product development, in recent years, virtual reality (VR) technology has been proposed and receiving more and more attention [3][4]. However, applying VR technology is costly in terms of hardware and software. Moreover, there are still many problems to be solved [3]. Hence, in this paper, a design animation method is proposed to achieve this task. The objective is to investigate a fast and economical media tool for all design participants, with which all related people can freely communicate and share their ideas in an efficient way.
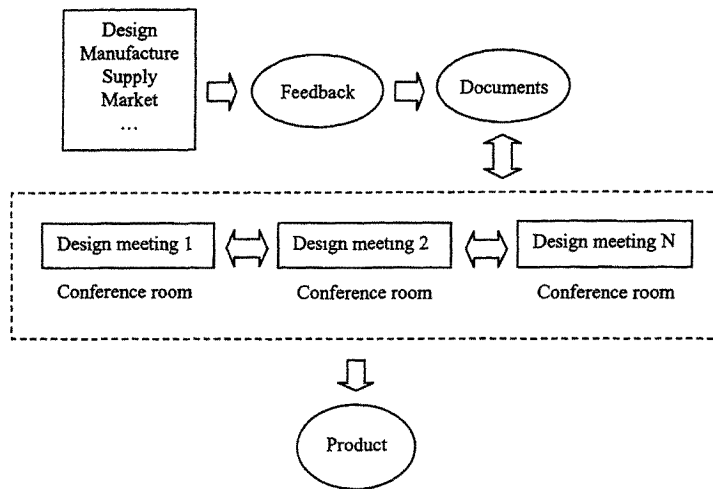
Fig. 1 Traditional design communication for new product development

2 DESIGN ANIMATION IN NEW PRODUCT DEVELOPMENT

In essence, design animation is the production of multimedia design AVI (audio visual interleave) files by capturing the product design from all angles and in motion so as to effectively demonstrate how the new product will look and perform. As such design AVI files can be conveniently emailed to the website or partners and easily viewed by anyone who has a PC, design animation allows designers to communicate their design concepts and collaborate more effectively, as illustrated in Fig. 2. In addition, the produced design AVI files can also act as a live training 'movies' for the technicians, particularly those for assembly and repairing. Hence, design animation would be an efficient and effective tool to help the

designers shorten the development cycle of new products. The advantages of design animation for new product development are summarized as follows [3][4][5]:

- Concurrent product design and marketing
  Using the AVI files produced via design animation, customers, designers, sales staff and engineers from different divisions can simultaneously evaluate a proposed product design and would no longer be affected by some human or environmental factors such as time and place. This not only allows for a more comprehensive assessment of design tradeoffs with a view to manufacturability, economics, parts availability, human factors consideration, maintenance and reliability but it also helps designers to arrive at an agreement on the basic design decisions earlier in the design cycle. On the other hand, as the produced design AVL files provide potential customers with the ability to visualize various aspects of the future product, it can be used for marketing studies prior to expensive prototype production.

- Live technical support documentation and training materials
  In design animation, the original design model can be flown around or revolved about 360° to show what it looks like from different angles; an assembly can be exploded or collapsed to show how the corresponding components will fit together and come apart; and the motion of moving parts of an assembly can be captured to show how they will interact. Hence, design animation can effectively demonstrate each feature of the new product in detail. The produced multimedia AVI files therefore become live technical support 'movies' for both designers and technicians.
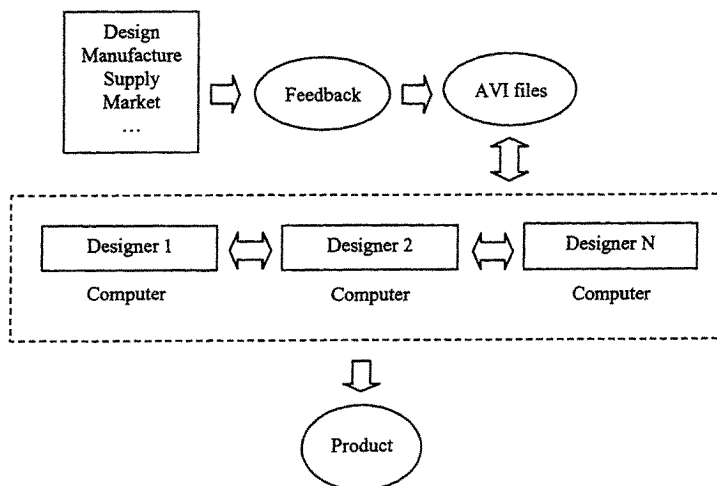
Fig. 2 Proposed design communication for new product development

3 THE PRODUCTION OF A MULTIMEDIA AVI FILE

The production of a multimedia design AVI file consists of two major steps: first to create AVI file from the product design and then edit the AVI file for multimedia effects. To implement this task, the two software *Solidworks* and *Multimedia Studio Pro* are applied in this paper.

3.1 The creating of AVI file

As mentioned, design animation is a powerful tool for communicating design engineering concepts more effectively. In the past, however, the high cost and time required to learn animation software made it impractical for most designers and engineers to adopt such technology [5]. *Solidworks* is a recently developed windows-based software for 3D modeling. Its embedded *Animator* is an easy-to-use and economical animation tool. Hence, we choose *Solidworks* for animation design.

Animating a product design includes the design for motion, the change of view-angle and the zooming of local feature details of the product. For assemblies, it also includes the design for assembly sequence and disassembly sequence. With *Solidworks*, these tasks can be performed in a simple way. For example, when creating a moving sequence, the designer needs to define only several key positions, as illustrated in Fig. 3. The software will automatically generate a smooth path passing through all these defined positions. Similar procedure applies for the design of view-angle change and local product feature zooming. As for the design of assembly and disassembly sequence, the software provides automatic exploding and collapsing function to fit and separate the components of the assembly, and the designer's main work is to schedule the time occurring sequence, as illustrated in Fig. 4. By combining the use of abovementioned functions, animation design such as flying and rolling for the product can be obtained, and this animation can be easily recorded and output as an AVI file by *Solidworks Animator* recorder.

Fig. 3 Automatic generation of smooth motion path in *Solidworks*

```
                          ┌───┐
                          │ A │
                          └───┘
┌───┬───┐  Explode  ┌───┐  Collapse  ┌───┬───┐
│ A │ B │    ⟹      │ B │     ⟹      │ A │ B │
├───┼───┤           └───┘            ├───┼───┤
│ C │ D │           ┌───┐            │ C │ D │
└───┴───┘           │ C │            └───┴───┘
                    └───┘
                    ┌───┐
                    │ D │
                    └───┘
```

Time 1   *Time 2*   *Time 3*   *Time 4*

ABCD ⇾ Explode A ⇾ Explode B ⇾ Explode C ⇾ Explode D
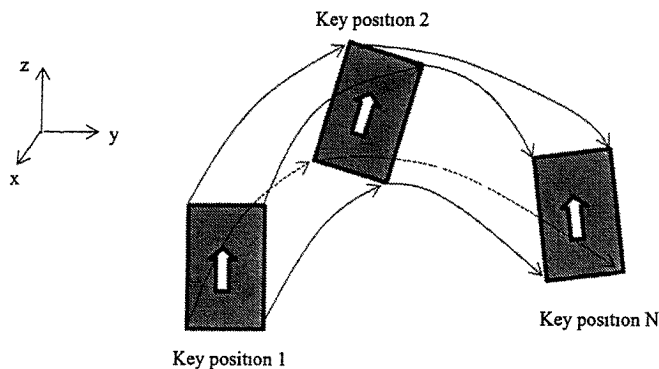
Time 10 ↑   ↓ *Time 5*

ABCD ⇽ Collapse A ⇽ Collapse B ⇽ Collapse D ⇽ Collapse C

     Time 9   *Time 8*   *Time 7*   *Time 6*
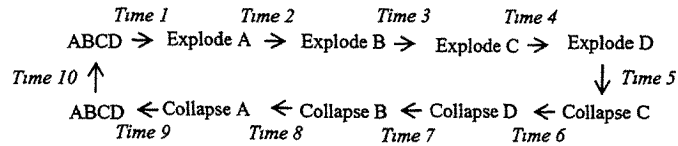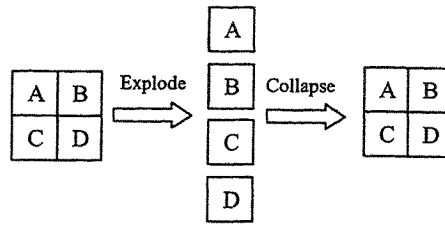
Fig.4 Automatically explode and collapse the assembly and schedule the time sequence

3.2 Editing AVI file for multimedia effects

Although the produced AVI file by *Animator* can effectively demonstrate how a new product will look and perform, the AVI file is not yet ready or effective for design demonstration. Without sound, sometimes, it is still hard for some design participants to understand the product design. For example, suppose an AVI file is presented to customers to show how a new product will function, if there is no description notes, it could be still difficult for customers to get the corresponding idea. Hence, we apply software *Multimedia Studio Pro* to further process the 'silent' AVI file for more impressive effects.

Editing AVI file mainly involves inserting notes files and audio files. The notes file is used to describe the animation design in detail in the format of Text and the audio file is in the format of Sound. These two kinds of files need to be prepared before insertion. *Multimedia Studio Pro* offers user a convenient *Timeline* tool to implement these tasks. As illustrated in Fig. 5, with *timeline*, notes file and audio file generated can be easily inserted into the AVI frames by drag-and-drop. Moreover, *Timline* allows user to insert more than 3 audio files. Hence, beside the audio file created to describe the animation design, we can also create an audio file for musical purpose so that the reviewing of the AVI would not be tedious. For assemblies, we can also create more audio files to simulate the assembly process, such as the 'sound' when components collide. With the inserted description notes and music sound, the 'silent' AVI file takes on multimedia effects and therefore would be more effective by a more 'life-like' appearance.

4 AN APPLICATION EXAMPLE

In this paper, we use an assembly product to illustrate the animation design for RPD. This assembly is a front door cabinet taken from an audio product. It comprises 19 components. A

multimedia design AVI file is produced to demonstrate how these components are fitted into the cabinet and how they are separated after assembly.

In the animation design, at first, we applied the automatic functions provided by *Solidworks* to explode and collapse the cabinet. Then the time sequence for the exploding and collapsing sequence is re-scheduled, as illustrated in Fig. 6. For each motion of the component, the motion time is defined with 3-5 seconds. To enhance the animation effects, some motions of the components are designed with flying and rolling path, as illustrated in Fig. 7, and some components are designed with local zooming to show the detail (Fig. 8). To create the AVI file, the whole animation process is recorded at a rate of 7.5 frames per second. For the produced AVI file, description notes and sound are added to be in conjunction with the corresponding animation effects for multimedia effects, as illustrated in Fig. 9. The size of the final produced AVI is about 73 MB and the time to play the file is about 3.5 minutes.

In the experiment, this multimedia design AVI file is emailed to technicians for assembly and repairing. They were able to rapidly capture the design idea by just playing the file several times, whereas with traditional documents, they would normally need several days to understand the assembly idea. This indicates that design animation is an effective tool for RPD and the multimedia design AVI files are live technical 'movies' for rapid design communication.
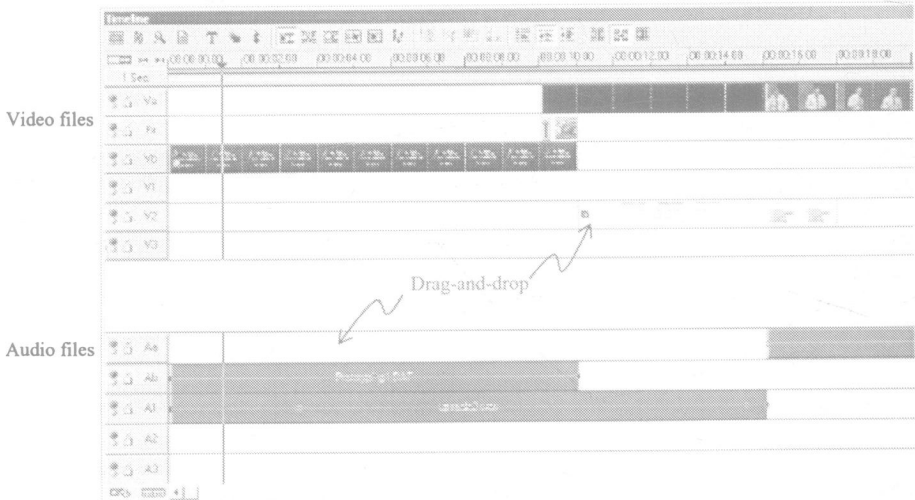
Fig.5 Timeline tool for editing AVI files in *Multimedia Studio Pro*

total-front-assy-mc170_asm	

Fig. 6 Animation sequence of the assembly

Fig.7 Animation design for rolling and flying effects
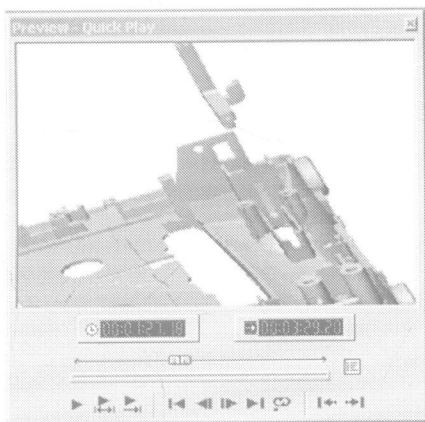
Fig.8 Animation design of local zooming
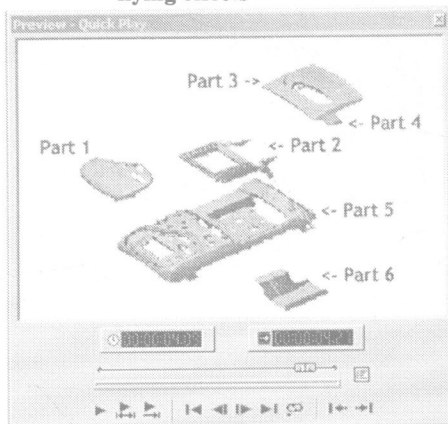
Fig. 9 The AVI file having description notes and sound

5 CONCLUSIONS

This paper proposed a methodology of exploiting design animation to enhance the design communication in the new product development. By using readily available and reliable

commercial software *Solidworks* and *Multimedia Studio Pro*, multimedia based design animation file (AVI file) can be conveniently created and easily shared by all the design participants. The generated AVI file has been demonstrated by an assembly case study to be not only a high efficient communication media for designers, but also a live technical support 'movie' for technicians.

REFERENCES

1. **H. J. Bullinger, J. Warschat** and **D. Fischer**. Rapid product development – an overview. Computers in Industry, Vol. 42, 2000, pp. 99-108.

2. **A. Bernard**. Rapid product development case studies and data integration analysis. Computers in Industry, Vol. 43, 2000, pp. 161-172.

3. **R. Gupta, D. Whitney** and **D. Zeltzer**. Prototyping and design for assembly analysis using multimodal virtual environments. Computer-Aided design, Vol. 29, No. 8, 1997, pp. 585-597.

4. **U. Roy, B. Bharadwaj, S. Kodkani** and **M. Cargain**. Product development in a collaborative design environment. Concurrent Engineering, Research and Applications, Vol. 5, No. 4, 1997, pp. 347-365.

5. www.solidworks.com

6. The manual of Solidworks 2000.

7. The manual of Multimedia Studio Pro 6.5.

Systematic approach for modelling the superplastic deformation process of 2024Al alloys under constant strain rate – use of FE technique

O F YENIHAYAT, H UNAL, A MIMAROGLU, and A OZEL
Faculty of Engineering, University of Sakarya, Adapazarı, Turkey

ABSTRACT

In recent years, there is a considerable interest in the application of superplastic forming for the fabrication of complex parts in the aircraft and automobile industries. In this field, the existing analytical theory is far from perfect and the results deduced from it do not describe the real conditions very well. Therefore numerical and modelling procedures are essential. In this study, ANSYS finite element code is used and a systematic approach was developed to carry out the forming analysis at constant strain rate. In this approach experimentally obtained material data at different cross- head speeds were used to model and analyse processes at constant strain rate. This was ensured by writing software using a parametric design language (PDL) to lead the analysis. Uniaxial tension and deep drawing processes for 2024 Al. alloy sheet were modelled and analysed. The results showed the effectiveness of this approach in describing the superplastic forming process at predefined process conditions. Furthermore it is established that in using the systematic approach it is possible to allocate the proper process conditions parallel with the complexity of the product.
Keywords: 2024 Al. alloy, superplasticity, finite element, constant strain rate

1 INTRODUCTION

Superplasticity phenomena have been observed for a wide range of materials including metallic alloys, ceramics, composites and minerals. In this process, the deformation is carried out under low pressure and a large elongation is reached without failure. Materials such as titanium, aluminium alloys, when subjected to the proper conditions, pressure, temperature and strain rate can exhibit the phenomena of superplasticity [1-2]. These conditions are summarised as a grain size less than $10\mu m$, low strain rate of less than 10^{-3} s^{-1} and temperatures of ≥ 0.5 T_m where T_m is the melting point of the material [3]. The main application fields of this process are aircraft and automobile industries. In fact, its application for aircraft structures have been expanding recently because of its superior characteristic, such as low-cost, light weight and short fabrication time [4-9].

Modelling of superplastic forming (SPF) has taken considerable interest over the years and many numerical algorithms have been suggested [10-13]. These algorithms generally restricted to simple geometry in plane strain and axisymmetric loading configurations. More recently, finite element method is on increase to simulate the SPF process because of its capability for variable such as strain rate to be examined more closely than can be done during experimental process. In fact the numerical simulation by finite element method of such highly non-linear processes is potentially expensive [14-18]. Moreover, the accurate

modelling of the complex product requires the use of meshes comprising large numbers of elements. Therefore, a successful model must hold the promise of accurately predicting both the behaviour of the material during forming and the characteristics of the finished work piece. In past several works has been carried out for SPF analysis using MARC, code. First, Rebelo and Wertheimer [14] showed general description of SPF by analyzing a two-step pan. Sadeghi and Pursell [10] modelled an aircraft door using the same code. Later Nihat ET. al., [19] modelled bulging of Titanium alloy tube to predict some process parameters such as time, pressure. They have concluded that the forming times are slow, and there is a critical need for optimizing forming pressures, strain rate and deflection.

The fundamental studies for evaluation of the superplastic deformation process usually begin with the uniaxial tensile behaviour of the materials. Therefore, the main aim of this study is to model and analyse the uniaxial tensile deformation of aluminium alloy 2024Al at constant strain rate. Analyses were carried out using commercially available finite element code ANSYS and an approach was developed to control the analysis to follow a systematic route, which ensure a constant strain rate deformation condition.

2 EXPERIMENTAL WORK

To obtain material data for the finite element analysis, tensile samples from 0.25mm thickness 2024Al alloy sheets were prepared with dimensions, see figure1a. Unaixial tensile tests were carried out on a DARTEC testing machine at 525°C temperature and at constant cross - head speeds range 0.002, 0.001, 0.02, 0.01, 0.1 mm/sec speeds (an initial strain rate range 5×10^{-5}, 1×10^{-4}, 5×10^{-4}, 1×10^{-3}, 5×10^{-3} sec^{-1}) conditions. The load- extension data were recorded and using the volume conservation, true stress and strain curves were obtained. Finally, these data were approximated by bilinear model and feed into F.E. analysis see, figure 2. In this figure, different cross-head speed results are identified as CHS_1, CHS_2, CHS_3, CHS_4 and CHS_5.

3 ANALYSIS

Due to symmetry, a quarter of uniaxial tensile sample was modelled using 96 quadratic (VISCO108) element type. The ends of the model were constrained and a prescribed displacement of 12mm (as a load) was applied at the gauge length end nodes, see figure 1b. To carry out the superplastic forming analysis for these models at a pre-defined constant strain rate condition, an F.E. code ANSYS was used and was systematically controlled. The control process was ensured by preparing a PDL (parametric design language) file, see figure 3. The function of the PDL software is to calculate the strain rate value for each element during the analysis time and to act as a control mechanism for the pre-defined strain rate value. Hence to lead the analysis programme to use different segments of the materials data curves among that of CHS_1 to CHS_5 which full fill the pre-defined strain rate value of the process condition, see figure 2 and 4. In F.E. analysis the overall process was analysed in several loading steps. This was done to ensure the convergence of the solutions.

Finally, in the PDL file two different algorithms for the definition of the strain rate were suggested and used. In the first definition, the strain rate of each element is calculated as the change in the strain value with time. In the second definition, an equation, which gives the strain rate in uniaxial tensile test according to its gauge length and cross- head speed, is used. These three definitions are:

$$\dot{\varepsilon} = \Delta\varepsilon_{seq} / \Delta t \quad \text{......... ....1}$$

$$\dot{\varepsilon} = v / \left[L_o (1 + e_t) \right] \quad \text{........ .....2}$$

Where $\dot{\varepsilon}_i$ is the strain rate for element i, Δu_i presents the displacement of element i, $\Delta\varepsilon_{ieq}$ is the change in strain, v is the cross- head speed, L_o is original gauge length, e is the engineering strain and Δt is the time interval.

Finally, the finite element analysis using the systematic approach and using different algorithms (equations 1 and 2) for strain rate were executed for tensile sample and for sheet forming models. The deformed shape, stress and strain contours were obtained and compared.

4 RESULTS AND DISCUSSION

Figure 5-7 presents the deform shape, equivalent strain and equivalent stress contours results for tensile sample model respectively. In figure 5, the numbering for the element from 1-5 presents the final material properties reached by each element at the end of the deformation process. For example, number 5 (CHS$_5$) means that, during the analysis, this element have followed material properties CHS$_1$ (at cross-head speed 0.001m/sec) then by CHS$_2$, CHS$_3$, CHS$_4$ and finally reached CHS$_5$ (cross head speed 0.1m/sec), see figure 2. By the systematic approach the analysis has applied segment from each curve, which provide the pre-defined constant strain rate value. In figure 5, the different and the expected numbering of the elements within the model suggest the effectiveness of the systematic approach. In fact the PDL programme has ensured the full control of the strain rate value during the whole deformation process. On the other hand the blank eye observation of the experimentally tested sample and the deformed shape obtained by the analysis indicate the matching of the dimensions. This is again a good support for the practicality of the systematic approach. Figure 6 present the equivalent strain contour obtained by using strain rate algorithm (equations1 and 2). In this figure, the largest strain value reached at the necking area is about 2.876 and 2.796 for using strain rate definitions eq1 and 2 respectively. This represents a 3% difference in the results of these two definitions. Figure 7 present the equivalent stress contours using (equations1 and2) for strain rate definitions. In this figure, the maximum stress value is 86 and 93.8 MPa. This presents a difference of 9% in the results of the two different definitions of the strain rate. This is because equation 2 does not consider the time factor in the deformation behaviour of the material. Generally, all the results are in good agreement with each other. The favourability of those deformation rate definitions is dependent on the practical process condition. In general, it is clear that F.E analysis with the new approach is quiet able to trace and explain the superplastic forming of the materials.

5 CONCLUSIONS

- The results showed the effectiveness of the systematic approach to analysis deformation processes (such as superplasticity) with large non-linearity and high strain rate sensitivity behaviours. Furthermore showed it's capability for continuously tracing the deformation process within the die.
- The results suggest the reliability of both algorithms of the deformation rate.
- The simplicity of the systematic approach in explaining and analyzing very complex problems and shapes.
- The ability of this approach to point out the precise process condition according to the product complexity. Furthermore to suggest the convenient material properties for manufacturing the particular product within the required accuracy.
- The systematic approach showed the possibility of using simply obtained experimental data to predict, trace and solve very complex problems
-

REFERENCES

1 **Hamilton, C.H.** and **Paton N.E.** (1988)(Eds) Superplasity and Superplastic Forming, Proc. Int.Conf. P.706. The Minerals, Metals and materials Society, U.S.A.

2 **P., ARGARD-LS-154** (1987) Superplasticity, Advisory Group for Aerospace Research and Development, NATO

3 **Watanate,** H. and **Mukai, T.** (1999) Scrip. Materialia, (40 N0.4), pp 477-484

4 **Plling, J.** and **Ridley, N.** (1989) Superplasticity in crystalline solids p159, London, The institute of Metals

5 **Yang S.H. Ahmed, H.K.** and **Roberts,W.T.**(1989) Mater. Sci.Eng. (A122), pp193-203

6 **Zhou, D.J. Lian, J.** and **Suery, M.**(1988), Mater. Sci. Technol., (4), pp 348-353

7 **Matuo, M.,** (1986), Japan Inst. Of Light Metals,(36 No.1), pp 43-50

8 **Hals, S.J. Bales, T.T. James,W.F.** and **Shinn, J.M.,** (1990), the Minerals, Metals, Materials Society, pp 167-185

9 **Hira,H.Hayami,K.H.I. Higashi, K.** and **Tanimura,T.S.,**(1991) JSTP, (32.No.362), pp 359-363

10 **Sadeghi,R.S.** and **Pursell,Z.S.,**(1994) Materials Science Forum, (170-172), pp.571-576

11 **Lee,D.** and **Backofen,W.A.,**(1968) .Int.J.Mech.Sci, (12), pp403-427

12 **Luo,Z.J. Guo,N.C.** and **Gong,Q.Y.,**(1986), Proc. 14th North American Manuf. Research Conf.,pp 420-425

13 **Gavriushin, S.S.** and **Zenkiewicz,O.C.,**(1986), Int.J.numer.methods eng., (23), pp 1179-11994

14 **Rebelo, N.** and **Wertheimer, T.B.**(1988) Trans. Of NAMRI of SME (16)

15 **Shah, K.N. Story, J.M.** and **Chandra, N.**(1988), Superplasticity in Aerospace, pp.135-147, Arizona, January 25-28.

16 **Bonet, J. Wood, R.D.** and **Wargadipura,H.S.,**(1990), Int. J. Numerical Methods in Eng, (30), pp.1719-1737

17 **Suzuki, N. Kohzu, M. Tanabe,S.** and **Higashi, H.,**(1999), Mat. Sci. Forum, 1999(304-306), pp.777-782.

18 **Yenihayat, O.F.**(2001), Computer aided new approach for analyzing superplastic forming process, PhD thesis, Sakarya University, Sakarya,Turkey.

19 **Nihat, N. Ken-ichi, M. Masanori, K** and **Hisashi, N.,**(1997),A finite element modelling for superplastic bulging of titanium alloy tube and pressure path optimisation, Mat. Sci. Forum, (243-245), pp 729-734
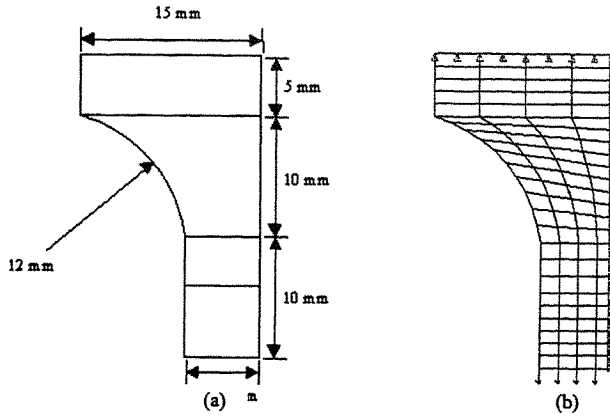
Figure1. Tensile Sample a) experimental quarter sample b) finite element model
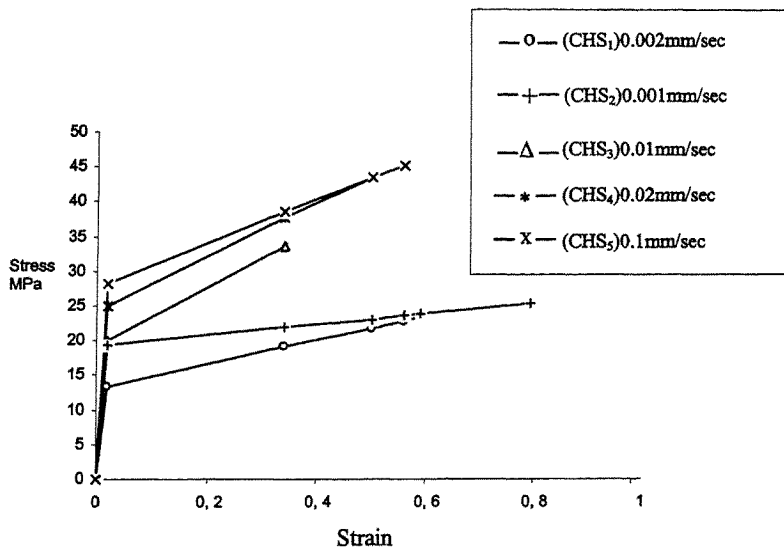
Figure 2 experimentally obtained and approximated true stress-strain curves at 525°C for 2024Al alloy tested at .002, 0.001, 0.02, 0.01, 0.1 mm/sec range cross-head speeds
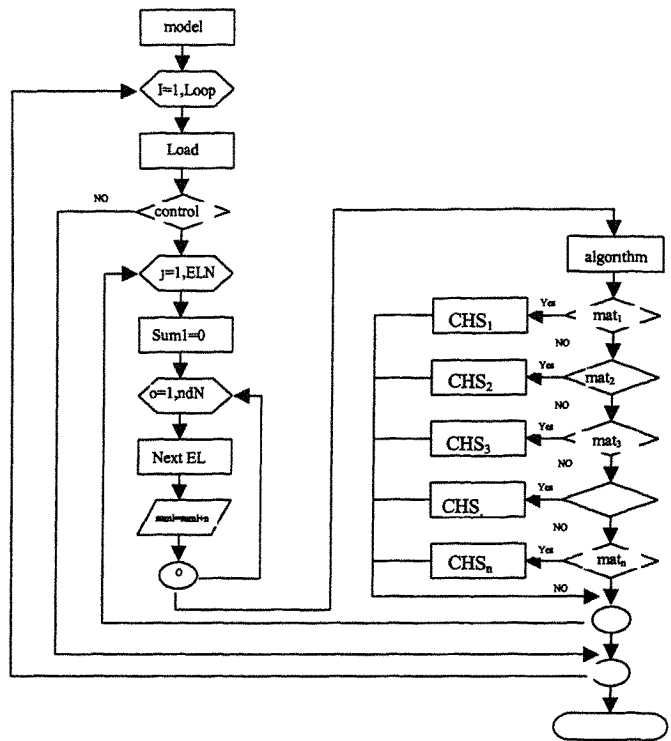
Figure 3: Flow chart for PDL file
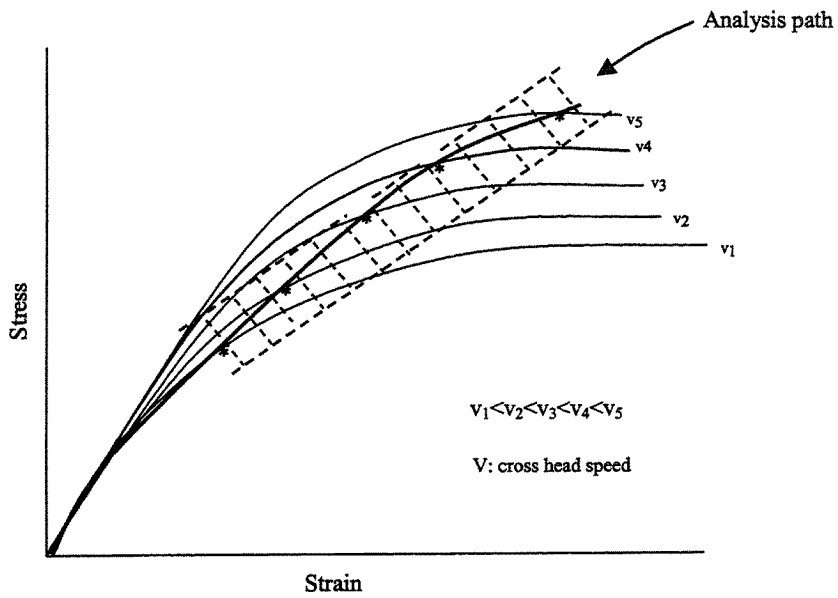
Figure 4 Schematic curves used for explaining the systematic approach
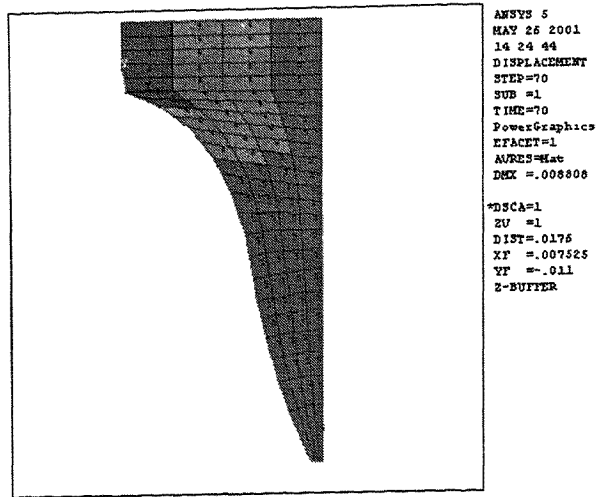
(a)
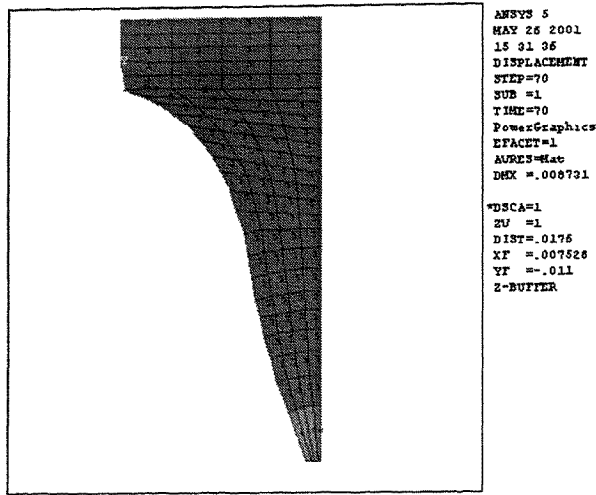
(b)

Figure 5 Deformed meshes for 2024Al Alloy at 525°C at 7mm applied load: a) using strain rate definition eq1 b) using strain rate definition eq2
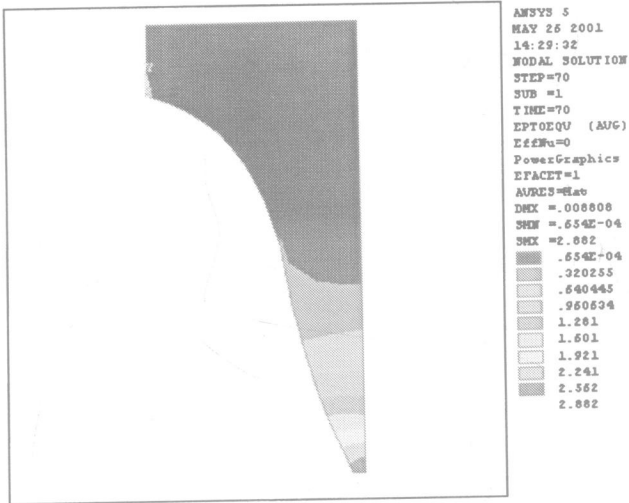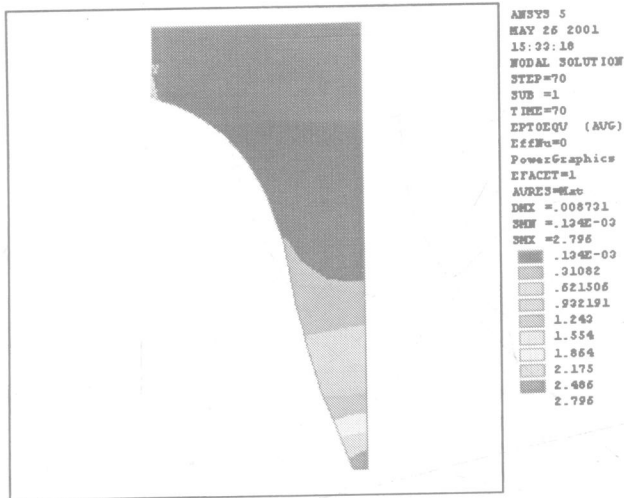
Figure 6 Equivalent strain contour for 2024 Al Alloy at 525°C and 7mm applied load: using strain rate definition eq1 b) using strain rate definition eq2
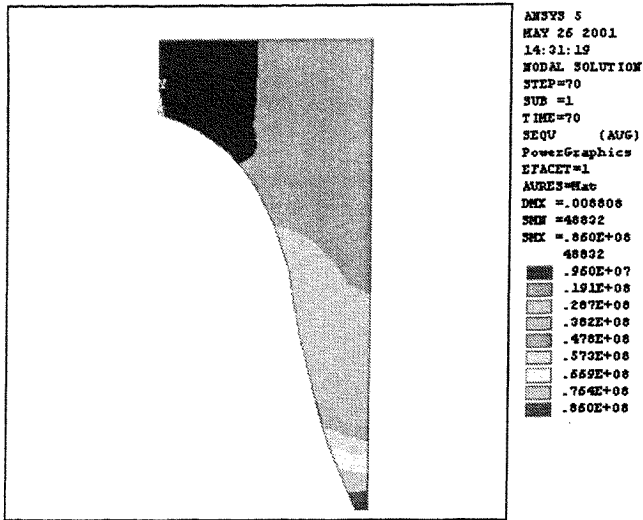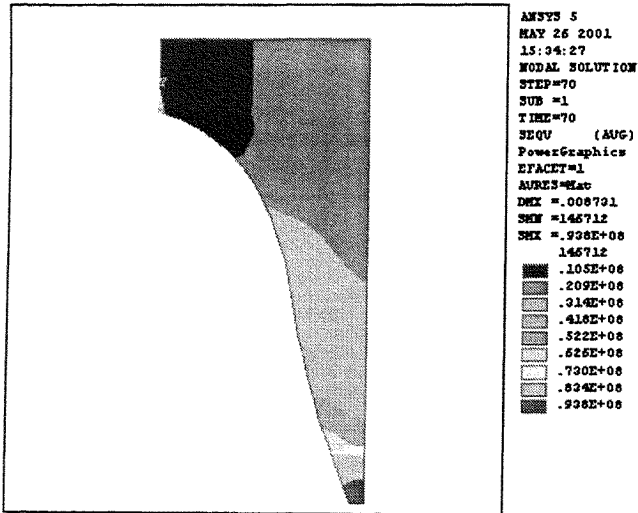
Figure 7 Equivalent stress contour for 2024 Al alloy at 525°C and 7mm applied load: a) using strain rate definition eq1 b) using strain rate definition eq2

Computer-aided design methods for additive fabrication of truss structures

H WANG and **D W ROSEN**
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

ABSTRACT

This paper presents computer-aided design methods to create truss structures manufactured with additive fabrication to enhance a part's mechanical properties. It describes the overall design process for the truss structure, and then discusses how parametric modeling and solid modeling techniques, finite element methods, and optimization approaches are applied to design the truss structure. Examples and some potential applications will be discussed.

1 INTRODUCTION

Trusses can be used to stiffen and strengthen structures and mechanisms, while reducing weight (7,11). The challenge is in their manufacture. In general, truss structures can either support an individual part surface shown in Figure 1 or could fill the entire volume of a part. They should fit in the part space and can achieve high strength and stiffness with less material. The computer-aided design methods presented in this paper are used to create such truss structures. The part surface is approximated by Bezier surface patches and then a truss structure is created between these Bezier patches and/or their offsets using parametric modeling technology. The truss structures are optimized with finite element methods and optimization techniques. Solid models of parts with the truss structure can be created after determining the topology of a truss structure. Parts are then manufactured using Additive Fabrication processes, in which parts are built by adding material, as opposed to subtracting material from a solid object; additive fabrication processes include Stereolithography, Selective Laser Sintering, and so on (6,8).
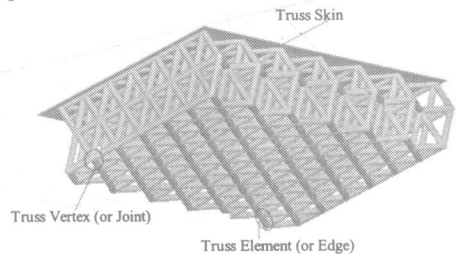
Figure 1. A Truss Structure to Enhance Mechanical/Dynamic Properties of a Part

This paper presents computer-aided design methods to create a conformal truss structure that conforms to the part shape. Daily, Lees, and McKitterick created a pattern of truss elements and then repeated it in every direction to form a uniform truss structure as shown in Figure 1 (5). However, if part boundaries are curved, some boundary truss joints may not be located on the part wall and all truss elements are oriented into a few fixed directions. An internal truss structure that conforms to the part's shape would fit in the part and would better distribute forces within the part. Figure 4 shows a conformal two-dimensional triangular truss, in which the boundary truss vertices are located on the part bounds and most truss elements can be oriented toward boundary loads. The conformal truss structure in Figure 4 would better enhance the mechanical properties of the part than the uniform truss in Figure 3. Secondly, the shapes of the individual elements in the uniformly patterned truss shown in Figure 3 are not changeable for adaptive material distribution to better enhance mechanical and dynamic properties. On the other hand, the individual element sizes of the conformal truss structure shown in Figure 4 can be adaptively adjusted to obtain the desired mechanical and dynamic properties. Therefore, there are two advantages of the conformal truss over the uniform truss: conforming to the part shape, and adaptively enhancing the mechanical and dynamic properties. Hence, a truss structure that conforms to part shape is desired.

The complex geometry of the conformal truss structures is far beyond that of typical CAD models. Parametric modeling technologies, finite element method, optimization approaches, and solid modeling techniques have to be investigated to design and represent the conformal truss structures; these investigations will allow us to better enhance the mechanical and dynamic properties of parts. Figure 2 shows the design process of truss structure, which consists of five sequential steps (11):

Step I   The part is shelled in the original geometric modeling package to obtain the STL model of the thin skin for the part that covers the truss structure.

Step II   The part surface is manually decomposed and approximated with a series of bicubic Bezier surface patches.

Step III The truss topology is created between Bezier patches in a Matlab program using parametric modeling techniques. Conformal truss topology contains information about the truss vertex coordinates (Vertex Topology) and the edge connections (Edge Topology).

Step IV The truss topology is optimized with finite element methods and engineering optimization techniques in ANSYS. The optimization objective is to minimize the material mass, but still get the required mechanical properties.
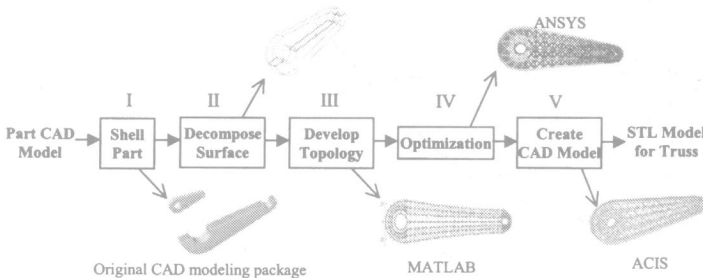
Figure 2. Design Process of Truss Structure

Step V   The solid modeling technology is applied to create the solid model (STL) of the truss structure with software developed using ACIS as the geometric modeling kernel.

2 DEVELOPING TRUSS TOPOLOGY

Parametric surface modeling techniques are utilized in Step II and III to decompose the part surface into bicubic Bezier surfaces and develop the conformal truss. Other parametric surfaces such as NURBS (9) are also applicable, but the bicubic Bezier surface is one of the simplest forms and commonly used in most CAD systems. Bicubic Bezier surface is a 3-order surface. The matrix form of bicubic Bezier surface is shown as Equation 1 and it has $4 \times 4 = 16$ control points. $p_{ij}(i,j = 0,1,2,3)$ are the control vertices, and u,w are the parameters for a Bezier patch. B is the matrix containing the sixteen control vertices.

$$p(u,w) = UM_B BM_B^T W^T \qquad \qquad \textbf{Equation 1}$$

Where, $U = [u^3, u^2, u, 1]$ $\qquad W = [w^3, w^2, w, 1]$

$$B = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ p_{30} & p_{31} & p_{32} & p_{33} \end{bmatrix} \quad M_B = \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

The presentation of Step III, creating the truss topology, will go from two-dimensional truss, to 3-D truss for a single volume, then to 3-D truss for multiple volumes. The 2-D conformal truss in Figure 4, on a bicubic Bezier surface, can be transformed from the two-dimensional uniform truss in Figure 3 by relocating the truss vertices. Vertex V_{ij} in the conformal truss corresponds to Vertex V_{ij} in the uniform truss $(i,j = 0,1,2,3)$ with the identical u and w values, and the edge connections among truss vertices are kept the same. The values for u and w can be determined as the uniform truss since all vertices are located on a grid shown as the dashed lines in Figure 3. With these u and w, the coordinates of all vertices in the triangular truss can be calculated using Equation 1. Besides calculating the vertex topology, we need to obtain the edge topology, which describes the connections between the truss vertices. Both the uniform truss and the conformal truss have the identical edge topology. The truss edges are constructed by linking one vertex by one vertex. For example, in Figure 3, each set of truss vertices $(V_{00}, V_{10}, V_{01}, V_{11}, V_{02})$, $(V_{20}, V_{10}, V_{21}, V_{11}, V_{22})$, (V_{10}, V_{11}), and (V_{20}, V_{21}, V_{22}) is
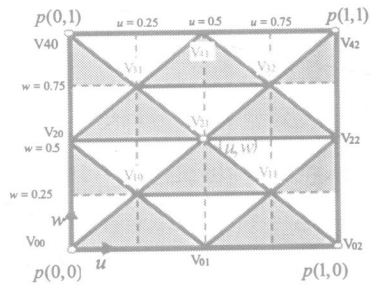
Figure 3.   Uniform Truss
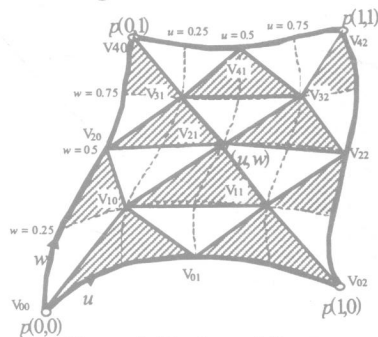
Figure 4. Conformal Truss

linked into Polyline $\overline{V_{00}V_{10}V_{01}V_{11}V_{02}}$, $\overline{V_{20}V_{10}V_{21}V_{11}V_{22}}$, $\overline{V_{10}V_{11}}$, and $\overline{V_{20}V_{21}V_{22}}$ as an edge topology.

The method to develop 3-D truss topology for a single volume is to start with two-dimensional truss topology. Figure 5 shows a truss structure developed between two bicubic Bezier surfaces, which enclose the space into a single volume. Their control vertices are represented by b_{ij_γ} ($\gamma = 0$) and b_{ij_γ} ($\gamma = 1$) respectively ($i = 0,1,2,3$; $j = 0,1,2,3$). The space between these two bicubic Bezier surfaces is considered as a $3 \times 3 \times 1$-order Bezier solid and the intermediate bicubic Bezier surfaces linearly interpolate those two boundary bicubic Bezier surfaces. Control vertex b_{ij_γ} ($i = 0,1,2,3$; $j = 0,1,2,3$; $\gamma \in [0,1]$) linearly interpolates the corresponding vertices b_{ij_0} and b_{ij_1} as well. Therefore, the two-dimensional truss topology can be developed for all intermediate Bezier patches. The edge topology between two neighboring layers, L_k and L_{k+1}, can be obtained by linking vertex $V_{i,j,k+1}$ on layer L_{k+1} with three vertices $V_{i-1,j,k}, V_{i,j,k}, V_{i,j+1,k}$. The bold dashed lines represent the truss edges on the Bezier surfaces, and the bold solid lines represent the truss edges between these Bezier patches.
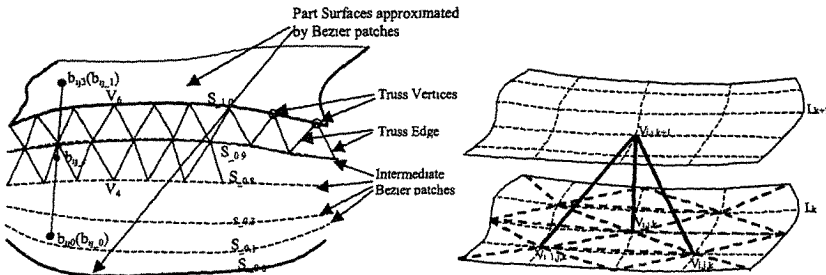
Figure 5. Developing Truss Topology for a Single Volume

Three-dimensional truss structure can also be created for multiple volumes. Figure 6 shows the trusses for two neighboring volumes represented only by their top and bottom surfaces. The corresponding control vertices of their common cubic Bezier curves (C1 and C2, C3 and C4) are identical. The vertex topology of the combined truss can be obtained by combining the two vertex topologies together by merging the coincident truss vertices along the coincident curves. G^1 continuity ensures the smoothness between two neighboring truss volumes (9). The tangents of the neighboring Bezier surfaces must be collinear at their intersection edges as shown in Figure 6.
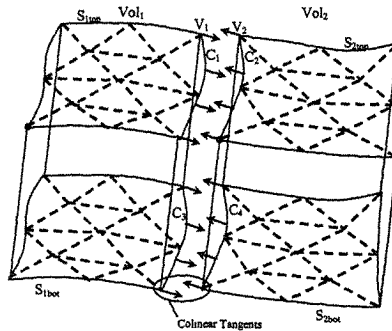
Figure 6. Composing Two Truss Topologies

I002/013/2002 © David W Rosen 2002

3 DECOMPOSING PART SURFACES AND CONSTRUCTING BEZIER SURFACE

In Step II, the part surfaces are decomposed to increase the approximation accuracy of Bezier surfaces and the Bezier surfaces are constructed from these decomposed surface patches. One Bezier surface patch usually cannot represent an entire part surface, so that additional Bezier patches are often required to increase the approximating accuracy. As a curve example shown in Figure 7, the variation of the original curve is beyond the representation capability of one cubic Bezier curve. The truss vertices, such as A, B, C and D located on the Bezier curve, are far away from the original curve. The approximation problem of a single Bezier surface is similar to that of this single Bezier curve. The purpose of decomposing the part surface is to increase the approximation accuracy of the bicubic Bezier surfaces, each of which will approximate one surface section. Figure 8 shows a composite Bezier curve (dashed curve) composed of two cubic Bezier curve segments. The approximation accuracy is increased significantly and almost all truss vertices are located on the original curve. The approximation accuracy of the bicubic Bezier surface can be increased similarly by using multiple bicubic Bezier surface patches.
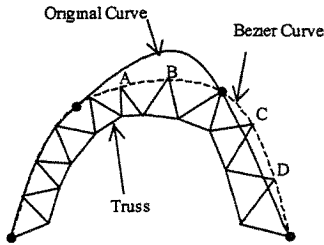
Figure 7. Inadequate Approximation Accuracy of Single Bezier Curve
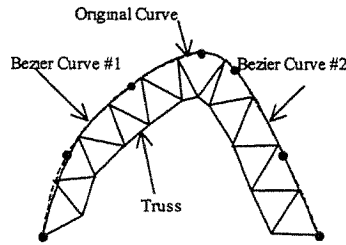
Figure 8. Increased Approximation Accuracy with Multiple Bezier Curves

After the decomposition of the part surface, 16 points should be determined for each bicubic surface patch as shown in Equation 1. Currently, this work is done manually. Figure 9 shows a cup (scaled down) on right and one quarter of this cup on left. For symmetry, only one quarter is considered and its internal surface $\overline{OACEFDB}$ is decomposed into such three surface patches as $\overline{OAB}$, $\overline{ACDB}$, and $\overline{CEFD}$. The largest patch $\overline{CEFD}$ is examined as an example to approximate with a bicubic Bezier surface. Sixteen surface points, $P_{ij}(i, j = 0,1,2,3)$, will be picked up and measured manually.

Figure 9. Constructing a Bezier Surface Patch

The sixteen control vertices, $p_{ij}(i, j = 0,1,2,3)$, of this Bezier patch are converted from these sixteen surface points (9,11).

4 OPTIMIZING TRUSS STRUCTURE WITH FINITE ELEMENT METHODS

In Step IV, Finite element methods and optimization approaches are utilized to analyze and optimize the truss structure to achieve maximum strength and stiffness with minimum material. The CAD model of truss structure is too complicated to analyze with finite element methods, and a better idea is to work on the truss topology. The truss topology is similar to the finite element model since the truss vertices can be used as the nodes in a finite element model, and the truss edge works as the element (10). This simplification method enables and significantly speeds up finite element analysis and optimization process (11). The optimization objective is to search optimal truss edge diameters to minimize the material mass, but with the desired mechanical properties, such as strength and stiffness. The truss edges are categorized into a certain number of groups according to their internal stress. Each group has one single truss edge diameter used as one design variable for optimization. ANSYS (3) is utilized to perform the finite element analysis and optimization. Figure 10 shows the finite element model of a robot arm with 8 different edge diameters (design variables). There are a variety of optimization methods provided by ANSYS. A direct search method is chosen to obtain the optimization result, and a gradient-based search method may be applied to find a more accurate solution (11).
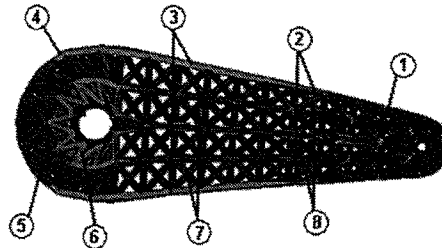
Figure 10. Finite Element Model in ANSYS

5 CREATING SOLID MODEL OF TRUSS STRUCTURE

After the truss topology is generated, it can be converted into a solid model by using the solid modeling techniques in Step V. The complex geometry of the truss structure is far beyond that of typical CAD models, and it is not possible to directly use commercial geometry modeling tools to create its solid model by adding truss edges one by one. A computer program must be developed to create the truss structure automatically without user's operation. The STL model for the truss is created directly, not converted from other solid models. STL file format is the de facto CAD model for Additive Fabrication processes (8). The STL model is much simpler than the other CAD models and requires much less computation resources (11).

A geometric modeling kernel, ACIS (4), is applied to perform Boolean operations on the truss edges (cylinders), since the intersection curves resulting from Boolean operations are still very hard to obtain. The concept, 'Vertex Group', is used to represent a sub-structure with a truss vertex and all its connected truss edges in a truss structure, as shown in Figure 11 (11). First, the solid model of each vertex group is created from uniting a series of cylinders using Boolean operations provided by ACIS. Only one half of each truss edge is created for every vertex group shown in Figure 11, in which each cylinder only represents half a truss edge. Then the planar end faces of all cylinders, Face F_1~F_4 in grey shown in Figure 11, are removed from the closed solid model to obtain the opened surface model of the vertex group. The STL model of the vertex group can result from the above surface

model using ACIS faceting tool. Finally, all the vertex groups' STL models are placed together directly without transformation shown in Figure 12.
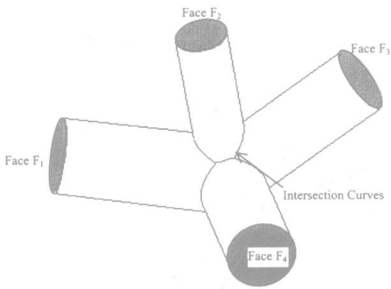
Figure 11. A Truss Vertex Group
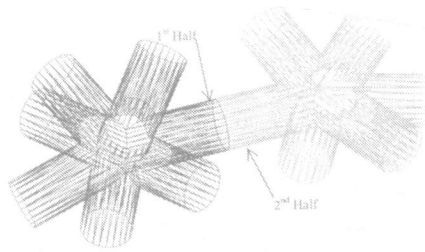
Figure 12. Creating STL Model Directly

6 POTENTIAL APPLICATIONS AND EXAMPLES

A truss structure can provide maximum strength and stiffness with minimum material mass. Additive Fabrication is able to implement virtually any internal structure and external shape. An optimized truss structure fabricated by Additive Fabrication can better enhance the mechanical and dynamic properties to achieve the mechanism with high strength, low mass, low inertia, high stiffness, desired resonant frequency, high surface area, minimum aging effects, and high porosity. Potential applications can be manufacturing, aerospace and automotive industries, medical engineering, and so on.

The strong, stiff and light mechanisms are highly desired in the manufacturing industry. For example, the robotic systems are used in all areas of manufacturing, including assembly, welding, spraying, material handling and various machining processes. The major portion of the

Figure 13. Robotic Link with Optimized Truss

actuator torques of serial manipulators is still used to support the manipulator's own weight (2). The truss structure manufactured with additive fabrications can be used to replace the internal material of the parts to achieve high strength and stiffness while reducing the overall mass/inertia and altering its resonant frequency as desired. Using the same material, the mass of each link in a robotic manipulator could be reduced by more than 30% through the use of truss structures without reducing its stiffness shown in Figure 13 (11).

There is a tremendous benefit to be obtained from least-weight structural design in all areas of our technological society. Some primary applications are sport utility vehicle, and aerospace structures. The advantages are better performance, fuel efficiency and better use of scarce material resources. Figure 14 shows the CAD model of a big triangle designed for an aerospace company. The truss structure is enclosed in the thin wall and the cylindrical edge diameter is $1.0mm$. Figure 15 shows the physical truss structure built on SLA 5000 (1). The

total material volume is $390cm^3$ with a saving of 80.7% compared to the original part. The total building time shows a saving of up to 41.7% (12).
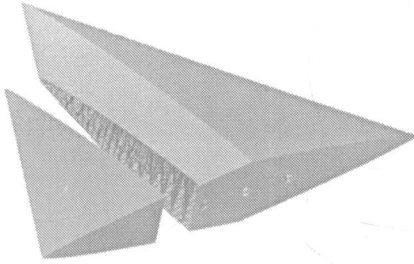
Figure 14. CAD Model of Truss Structure with Thin Wall
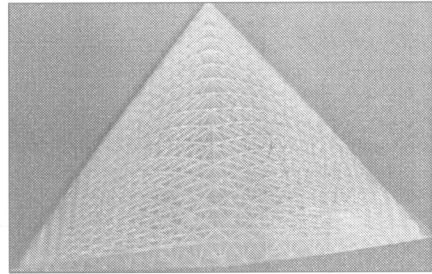
Figure 15. Truss Structure Built on SLA

7 CONCLUSIONS

Computer-aided design methods are presented to design and represent conformal truss structures. The proposed methods use parametric modeling techniques to decompose the part surface into bicubic Bezier surface patches and develop the conformal truss topology between these Bezier surfaces patches. Parametric modeling method can ensure that the truss structure fits in the part space and most truss elements are oriented towards loads on the part surface. Truss topologies can be created for two-dimensional truss, 3-D truss of a single volume and 3-D truss of multiple volumes. Decomposing the part surface and approximating it with more Bezier surfaces can increase the approximation accuracy. After converting the surface points into control vertices of Bezier surfaces, the bicubic Bezier surfaces can be constructed.

Finite element methods and engineering optimization technology can be utilized to analyze and optimize the truss structure. A simplification process is developed to simplify its finite element model to analyze the mechanical properties of the truss structures. The optimization methodology with the finite element methods can be used to optimize the truss structure for best enhancing the mechanical and dynamic properties. Solid modeling techniques are applied to create the STL model of the truss structure directly. The STL model of the truss structure is created as one vertex group by one vertex group, and then combined together directly into a whole STL model for truss structure. Creating STL models directly can effectively and efficiently generate the solid models of the truss structures in various shapes and sizes for additive fabrications.

The designed truss structure can be applied in aerospace, manufacturing, and other areas, where the light, strong and stiff parts are desired. Currently the part surface decomposition and the Bezier surface construction are done manually and can be time consuming. A new computer-aided design tool is under development to enable more automation.

ACKNOWLEDGEMENT

REFERENCES

1 **3D Systems, Inc.** (1995) "Stereolithography Basic Workstation Course Guide".

2 **An, C. H., Atkeson, C.G.** and **Hoolerbach, J.M.** (1988) Model-based Control of a Robot Manipulator, MIT Press.

3 **ANSYS, I.** (1999) "ANSYS Advanced Analysis Techniques Guide".

4 **Corney, J.** (1997) 3D Modeling with the ACIS Kernel and Toolkit, John Wiley & Sons.

5 **Daily, C., Lees, D., McKitterick, D.** (1997) "Truss Structure Design", United States Patent 6076324.

6 **Diez, J.** (2001) "Design for Additive Fabrication: Building Miniature Robotic Mechanisms", Master of Science, Department of Mechanical Engineering, Georgia Institute of Technology,

7 **Fuller, R. B.** (1975) Synergetics, Explorations in the Geometry of Thinking, Macmillan Publishing.

8 **Jacobs, P.** (1992) Rapid Prototyping & Manufacturing: Fundamentals of Stereolithography, SME, Dearborn, MI.

9 **Mortenson, M. E.** (1997) Geometric Modeling, John Wiley & Sons, Inc., New York.

10 **Reddy, J. N.** (1993) An Introduction to the Finite Element Method, McGraw-Hill, Inc.

11 **Wang, H.** (2001) "Computer-Aided Design Methods For The Additive Fabrication Of Truss Structure", Master Thesis, School of Mechanical Engineering, Georgia Institute Of Technology, Atlanta.

12 **Wang, H., Rosen, D.** (2002) "Parametric Modeling Method for Truss Structures ", ASME Computers and Information in Engineering Conference, Paper NO. DETC02/CIE-34495, Montreal.

A three-dimensional surface offset method for STL-format models

X QU
Industrial and Manufacturing Engineering, University of Rhode Island, Kingston, USA
B STUCKER
Mechanical and Aerospace Engineering, Utah State University, Logan, USA

ABSTRACT

This paper presents a new 3D offset method for modifying CAD model data in the STL format. In this method, vertices, instead of facets, are offset. The magnitude and direction of each vertex offset is calculated using the normals of the facets that are connected to each vertex. To facilitate the vertex offset calculation, topological information is generated from the collection of unordered triangular facets making up the STL file. A straightforward algorithm is used to calculate the vertex offset using the adjoining facet normals, as identified from the topological information. This newly developed technique can successfully generate inward or outward offsets for STL models. As with any offset methodology, this technique has benefits and drawbacks, which will be discussed in this paper. Finally, conclusions will be made regarding the applicability of the developed methodology.

1 INTRODUCTION

Offsets are widely used in many applications. These include tool path generation for 3D NC machining, rapid prototyping, hollowed or shelled model generation, and access space representations in robotics. In the rapid prototyping/manufacturing arena, 3D offsets are particularly important and useful as pre-process modifications to CAD geometry. Offsetting the CAD model allows one to compensate for errors caused by non-uniform shrinkage, inaccuracies in the process and the "stair-case effect" of the layer-by-layer build process. If the surface quality and accuracy of the fabricated parts is unacceptable, some kind of post processing is needed. If surface finishing is to be performed after part fabrication, a 3D offset of the model in the outward direction is required to ensure sufficient material for finishing to final dimensions. Since the STL file format is widely utilized, the development of an effective STL-based 3D offset method applicable to rapid prototyping is the goal of this effort.

Offset operations can be applied to curves, surfaces or entire 3D models. By definition, an offset means moving a point, a curve, or a surface of a 3D model by a constant distance 'd' in a direction normal to the surface of the model. Offset techniques for curves and surfaces have been extensively studied (1)(2). Generally, offsets of 3D models are achieved by first offsetting all surfaces of the model and then trimming or extending these offset surfaces to reconstruct closed 3D model (3)(4)(5). To offset a 3D model in the STL format, the most direct method would be to offset each triangular facet with the given offset distance in its corresponding normal direction. However, this will result in intersections or gaps between the

offset surfaces of two neighboring triangles. As shown in Fig.1a, a gap is formed between two offset surfaces F1 and F2 when the angle between them is convex. Conversely, an intersection or overlap occurs between offset surfaces, as shown in Fig.1b, when the angle between them is concave. In order to make closed 3D models from these triangular offset surfaces, it is necessary to identify all of the intersections, and then trim the surfaces on the line of intersection, and to identify all of the gaps and extend the surfaces to fill them. This can be quite complex, since thousands or millions of triangular facets may exist when representing complex 3D models using the STL format.
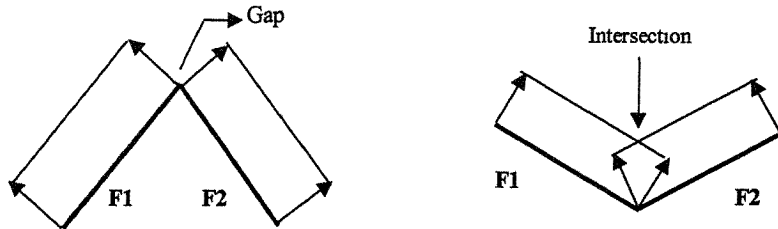
Figure 1a. Gap between offset surfaces   Figure 1b. Intersection between offset surfaces

This problem can be avoided if the vertices, instead of the triangular facets, are offset. As shown in Fig.2, when offsetting the vertices the relationship between facets will remain and there is no need to recalculate the triangle intersections. The challenge when utilizing this method is how to effectively calculate the offset vector for each vertex, taking into account the offset direction and magnitude, from all of its surrounding triangular facets.
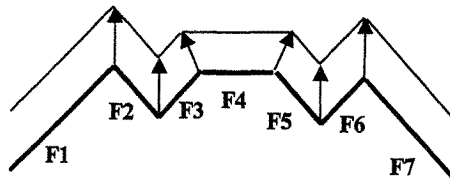
Figure 2. Offsetting Vertices

The most simple method for calculating vertex offsets is to calculate an averaged vector of the normal vectors of the triangles that are connected to this vertex. For example, the offset vector $\vec{V}_{Offset}$ at vertex V_1, where there are n triangles connected at vertex V_1, can be calculated using the following equation,

$$\vec{V}_{Offset} = \frac{\sum_{j=1}^{n} \vec{N}_{i,j}}{\left| \sum_{j=1}^{n} \vec{N}_{i,j} \right|} \tag{1}$$

where $\vec{N}_{i,j}$ are the normal vectors of the triangles that are connected to vertex V_1.

For a vertex located on a relatively smooth surface, this method works very well. However, for a vertex located on a sharp edge – the boundary curve of two or more surfaces with large normal vector differences between them – it will not work well and may lead to unacceptable

errors. As shown in Fig. 3, if the offset vector $\vec{V}_{Offset}$ of vertex V_i is calculated by averaging the normals $\vec{N}_{i,1}(1,0,0), \vec{N}_{i,2}(0,-1,0)$ and $\vec{N}_{i,3}(0,0,1)$, the result will be $(1/\sqrt{3},-1/\sqrt{3},1/\sqrt{3})$. But the correct offset vector should be $(1,-1,1)$ for vertex V_1, which can be easily derived by calculating the intersection of the offset facets of this cube.
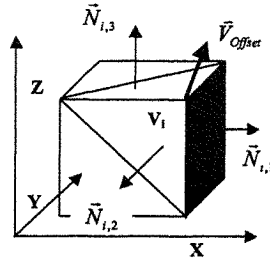
Figure 3. Offsetting vertex using averaged normal vectors

The same problem will occur for all remaining vertices of this cube, and the final offset will be $d/\sqrt{3}$ when using an offset distance d. This large error is unacceptable. In order to overcome this problem, this paper presents a new method for calculating the offset vector for a vertex by using the weighted sum of the normal vectors of the triangular facets connected to the vertex as follows:

$$\vec{V}_{Offset} = \sum_{j=1}^{n} W_j * \vec{N}_{i,j} \tag{2}$$

where W_j are the weighting coefficients associated with each triangular facet. A detailed method for calculating W_j and $\vec{V}_{Offset}$ will be discussed.

Since an STL file is simply a collection of unordered triangular facets of a model, topological information, particularly the identification of each unique vertex location and its neighboring triangles, needs to be generated. The remainder of this paper deals with the techniques for generating this topological information and a more detailed discussion of the algorithms used for calculating vertex offsets. Some examples of application of these algorithms to various geometries will be discussed and conclusions about the developed method will be given.

2 TOPOLOGICAL INFORMATION GENERATION

In order to facilitate the offset operation, topological information is derived from the unordered triangular facet list of the STL file. For a proper solid model, each triangle must have 3 neighboring triangles. The first task is to find those neighboring triangles and record their sequence numbers for each triangle in the 3D model. This can be done by finding matching edges for each edge of every triangle within the model. The implementation of this algorithm is simple but may be very time consuming if done as a one by one edge match operation. For example, if there are n triangles in the model, the number of edges will be 3n. Then the total match operations will be $C_{3n}^{2} = \dfrac{3n * (3n-1)}{2}$, which becomes a very large number and which will slow down the operation dramatically when offsetting a complex, high-accuracy 3D model. Thus, the hash technique is used to speed up the matching process.

After identifying the neighboring relationships among triangles, each vertex is identified and a reference to its connecting triangles is recorded. The algorithm begins with an initializing operation to mark the vertices of all the triangles as unchecked. Then all the vertices are checked one-by-one and corresponding operations are done based on the checking results as follows:

- If the vertex is unchecked, it means that this vertex is a newly identified vertex. It is added to the unique vertex list, and its status is marked as checked. The current triangle T_1 is registered as the first connected triangle, and neighboring information is used to find all other triangles that are connected to this vertex. As shown in Fig. 4, the triangle T_1 has two neighboring triangles T_2 and T_n that also share the current vertex V. The one on the right side (T_2) is taken as the next connected triangle, whose sequence number is recorded and the corresponding vertex (which is in the identical location to the original vertex and is not a unique vertex) is marked as checked. The right-side neighboring triangle of T_2 is then taken as the next connected triangle, and the process continues until the right-side neighbor is the first triangle, T_1.
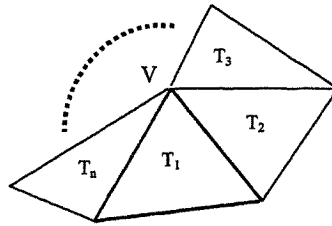
Figure 4. Neighboring relationship among triangles connected to vertex V

- If the vertex is marked as checked, this vertex is a duplicate vertex and the relationship between the vertex and adjacent triangles has already been established. This vertex is skipped and not added to the unique vertex list.

As a result of these operations, a unique set of vertices for the model is generated, along with a corresponding triangle sequence number list, which represents the connecting triangle information for each unique vertex.

3 VERTEX OFFSET CALCULATION

Equation (2) shows that the offset vector for each unique vertex can be calculated by using the weighted sum of the normal vectors of its connecting triangles. Before presenting the method for calculating the coefficients W_j, the relationship between $\vec{V}_{Offset}$ and $\vec{N}_{i,j}$ will be explored. As shown in Figure 5, $P_{i,original}$ is the vertex position before offset, and $P_{i,new}$ is the position after offset. Given the offset vector $\vec{V}_{Offset}$ and offset distance d_{Offset}, the following equation can be derived,

$$P_{i,new} = P_{i,original} + \vec{V}_{Offset} * d_{Offset} \qquad (3)$$
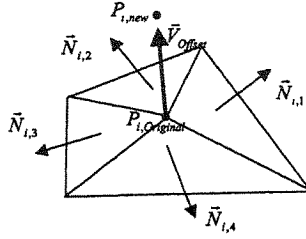
Figure 5. Illustration of the relationship between $\vec{V}_{Offset}$ and $\vec{N}_{i,j}$

From geometrical information, it is known that the perpendicular distance from $P_{i,new}$ to any original connecting triangular surface should be exactly the offset distance d_{Offset} since $P_{i,new}$ is located on the intersection point of the offset surfaces of all of the connecting triangular facets. That means that the following relational equation should be meet,

$$(P_{i,new} - P_{i,original}) \bullet \vec{N}_{i,j} = d_{Offset} \tag{4}$$

From equations (3) and (4), the following equation is derived,

$$\vec{V}_{Offset} \bullet \vec{N}_{i,j} = 1 \tag{5}$$

Substituting for $\vec{V}_{Offset}$ results in the following equation,

$$(\sum_{j=1}^{n} W_j * \vec{N}_{i,j}) \bullet \vec{N}_{i,k} = 1 \quad (k = 1 \text{ to } n) \tag{6}$$

Expanding and rearranging the equations in a matrix form yields,

$$\begin{bmatrix} \vec{N}_{i,1} \bullet \vec{N}_{i,1} & \vec{N}_{i,2} \bullet \vec{N}_{i,1} & \cdots & \vec{N}_{i,n} \bullet \vec{N}_{i,1} \\ \vec{N}_{i,1} \bullet \vec{N}_{i,2} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \vec{N}_{i,1} \bullet \vec{N}_{i,n} & \cdots & \vec{N}_{i,n-1} \bullet \vec{N}_{i,n} & \vec{N}_{i,n} \bullet \vec{N}_{i,n} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

These equations can then be used to find W_j (j=1 to n). However, when directly using the normals $\vec{N}_{i,j}$ of all the connecting triangular facets to calculate those equations, there may be no solution. For example, if two of these normals are identical, which is quite common in STL models, then two equations will be exactly the same, and there will be no way to solve for W_j. Thus it is necessary to eliminate the duplicated normals for each vertex before proceeding with the calculation. The equations will be solved using a computer numerical method. Because normals with similar vectors cause problems, due to computational accuracy, two normal vectors are treated as same if dot product between them is between a specified value and 1 – where the specified value is slightly smaller than 1. Under this situation, a vector, which is the averaged vectors of original similar normal vectors, is calculated to replace the original normal vectors. If the dot product of two normals is or closes to –1, then it means their corresponding triangles have opposite direction. This will not happen for the good STL model since any two triangles connected to same vertex will never

have opposite direction. Thus these two normals have opposite direction are eliminated as to make them no contribution for the vertex offset calculation.

Further analysis shows that the equations may have no solution if the number of normal vectors n is still larger than 3 after eliminating duplicated normals. From the viewpoint of offsetting operations, this would mean that there is no common intersection point among these triangular facets after offsetting each individual facet the given distance d_{Offset}. In order to solve this problem, an approximation method is presented when $n > 3$. In this method, the n normals are used to generate $(n+1)$ subgroups of normal vectors $(\vec{N}_{i,1}, \vec{N}_{i,2}, \vec{N}_{i,3})$, $(\vec{N}_{i,2}, \vec{N}_{i,3}, \vec{N}_{i,4}) \ldots (\vec{N}_{i,n}, \vec{N}_{i,1}, \vec{N}_{i,2})$. Each subgroup includes exactly 3 normal vectors, and can be solved to find its corresponding $\vec{V}_{Offset}$. The final $\vec{V}_{Offset}$ for the vertex will be the arithmetic mean of the subgroup $\vec{V}_{Offset}$.

After the offset vector $\vec{V}_{Offset}$ for each unique vertex is found, the new position of each vertex can be calculated using equation (3). Depending on whether an outward or inward offset is desired, d_{Offset} can be either positive or negative respectively. All of the new position values for each unique vertex are then used to construct a new set of triangular facets which represent the desired offset model in the STL format.

4   IMPLEMENTATION

The developed technique has been implemented using Visual C++ as a module in an integrated software package for finish machining of rapid manufactured parts. To investigate the performance of this offset method, some example parts were tested.

Fig. 6a shows a benchmark part from 3D Systems Corporation. It is a simple STL format benchmark model for machine accuracy studies. The developed software successfully generated both outward and inward offset models for it using an offset distance of 0.1 in., as shown in Fig.6b and Fig.6c.

(a)   (b)   (c)
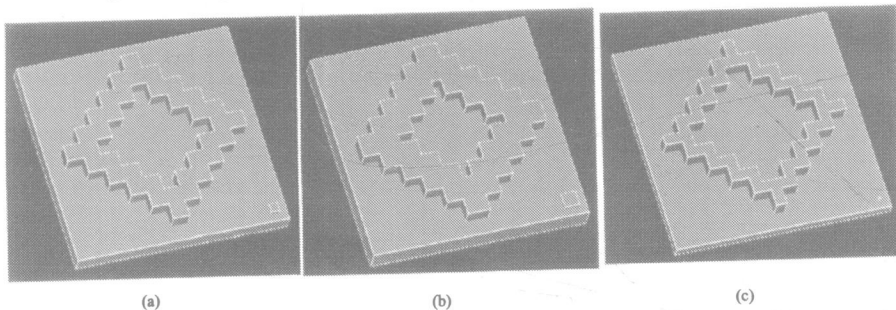
Figure 6. Illustration of offsetting first example part, (a). original model, (b) outward offset, (c) inward offset

The second example part was designed using Pro/Engineer, and then converted into the STL file format. This part includes a variety of features, such as different types of surfaces,

complicated sharp edges and holes in different orientations. The model was offset 0.03 in. outward and inward, and the results shown as Figure 7.

The third example part, shown in Fig.8a, is an injection molding tool cavity design provided by LaserFare. An offset value of 0.02 in. was applied, and the results are show in Fig.8b and Fig.8c.

(a)

(b)

(c)
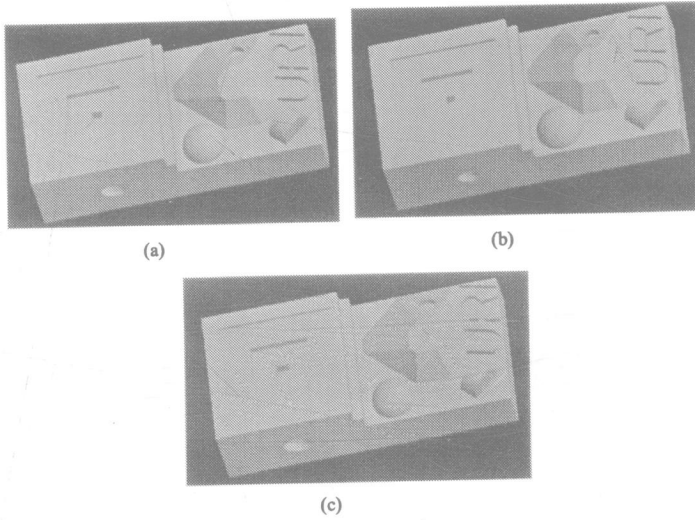
Figure 7. Second example part. (a). Original model, (b) outward offset, (c) inward offset

(a)

(b)

(c)
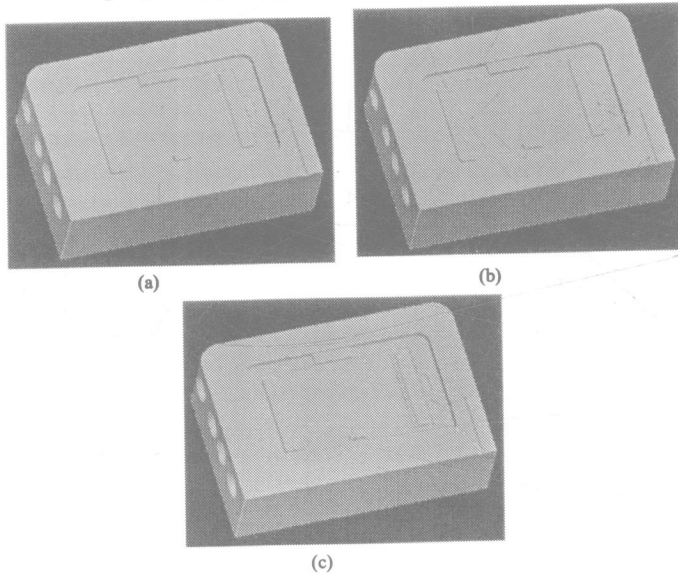
Figure 8. Third example part. (a). Original model, (b) outward offset, (c) inward offset

5 DISCUSSION

Since the offset vector for each vertex is calculated using a weighted sum of the normal vectors of those triangular facets connected to each vertex, in some situation, as shown in Fig. 9a, the length of the offset vector is much larger than one, and the vertex offset value becomes very large compared to the offset distance d_{Offset}. This is obviously undesirable. In order to solve this problem, a vertex splitting technique can be used. In this method, the original vertex is first divided into two vertices by adding a zero area triangular facet, then two vertices which include two different groups of triangular facets are offset individually using the above presented method. Fig. 9b shows the offset results by applying this technique to the same geometry shown in Fig. 9a.
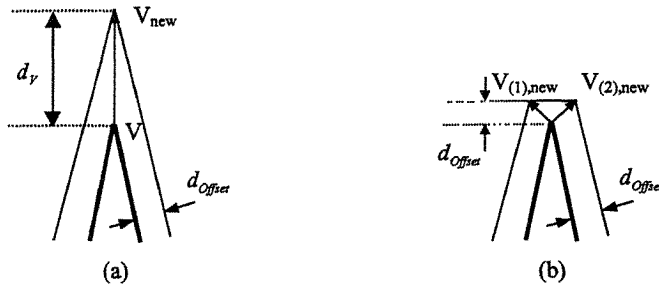
Figure 9. Illustration of vertex splitting method, (a). Normal offset, (b) Offset with vertex splitting

The methodology presented in this paper has focused on just solving the problem of offsetting all individual vertices of an STL model. It works well for small offset values, where local and global self-intersections normally do not happen. To handle the self-intersection problem while offsetting the model with relatively large values, some post processing needs to be done.

When there are no self-intersections, the topology of the offset triangular facets should be the same as the topology of the original triangular facets. However, when self-intersections exist, the outside surface of the model will be distorted. Under such circumstances, if each newly generated triangular facet is checked, it will be found that some of facets have experienced a dramatic change in their normal vector direction, even reversing the directions, and overlaps and undesirable intersections among those newly generated triangular facets will exist. By checking for unusual facets and facet intersections, these problems can be fixed. There are two possible ways to remove these self-intersections. One is to remove the self-intersection in 3D space. The possible method could be first finding all closed volumes by using the facet intersection information, then removing those closed volumes containing unusual facets. This is a complicated process. A different way is to try to remove self-intersection in 2D space, which is a suitable method for layer based manufacturing process. The method works by first slicing the model then removing the self-intersections from the 2D cross-sectional curves. This 2D method is relatively easy to implement. A lot of work has been done in the area of removing self-intersections from 2D offset curves (6)(7)(8), which can be effectively used.

6 CONCLUSION

In this paper, a new offset method is presented for the offsetting of 3D models in the STL file format. This method uses a technique for offsetting each individual vertex of the model, instead of offsetting triangular facets. This eliminates the complicated and time-consuming trimming and extending operations necessary for the offsetting of facets. Through analysis it was found that a method of calculating vertex offset vectors using the average of the normal vectors of triangles connected to the vertex is not accurate enough for vertices on sharp edges. Thus a method was developed to calculate offset vectors for vertices using the weighted sum of normal vectors of its connecting triangles. Detailed equations and algorithms were given to calculate the weighted coefficients and offset vectors. Computer implementation showed that the developed techniques can successfully generate inward or outward offsets for STL 3D models. Some of the limitations of this methodology were presented along with some possible solution methods.

REFERENCES

1   **Pham, B.** (1992), Offset curves and surfaces: a brief survey, *Computer-aided Design*, 24(4), pp.223-229.

2   **Maekawa, T.** (1999), An overview of offset curves ad surfaces, *Computer-Aided Design*, 31(3), pp.165-173.

3   **Farouki, R. T.** (1985), Exact offset procedures for simple solids, *Computer-Aided Design*, 2(4), pp.257-279.

4   **Satoh, T., Chiyokura, H.** (1991), Boolean operations on sets using surface data, *ACM SIGGRAPH: Syposium on solid modeling foundations and CAD/Cam applications*, Austin, Texas, USA, pp.119-127.

5   **Forsyth, M.** (1995), Shelling and offsetting bodies, *Proceedings of Third Symposium on Solid Modeling and Applications*, Salt Lake City, Utah, USA, pp.373-381.

6   **Tiller, W., Hanson, E.G.** (1984), Offset of two-dimensional profiles, *IEEE Computer Graphics and Applications*, Vol.4, pp.36-46.

7   **Bala, M., Chang, T.C.** (1991), Automatic cutter selection and optimal cutter path generation for prismatic parts, *International Journal of Production Research*, Vol.29(11), pp.2163-2176.

8   **Rohmfeld, R.F.** (1998), IGB-offset for plane curves - loop removal by scanning of interval sequences, *Computer Aided Geometric Design*, 15(4), pp. 339-375

The effective way of doing computer-aided reverse engineering

I E POPOV and **F M M CHAN**
School of Engineering, Cardiff University, UK

ABSTRACT

The purpose of this paper is to report the general procedures required to ensure appropriate methodology to enhance overall performance of doing computer-aided reverse engineering. Examples in the paper illustrate the key points of the procedures for reverse engineering especially on scanning, surface fitting, and 3D CAD modelling. The major advantages and disadvantages of the digitising techniques based on point cloud are discussed. Pre-processing of data acquired by scanning using either contact or non-contact method is explained including discussing some reverse engineering software.

1 INTRODUCTION

Over the years, various Reverse Engineering (RE) techniques have been introduced and several software packages have been released. These translate a captured point cloud into a 3D Computer Aided Design (CAD) model in the required format [1] in order to design, build, and support new processes and products. Even with state-of-the-art laser digitisers we cannot say whether or not digitising with a touch probe or a laser scanning head is superior and preferable. In recent years software development has frequently aimed at taking advantage of the on-going improvements in computing performance to do surface fitting of point clouds. Many companies already have the two basic tools needed for doing the RE task, namely a Co-ordinating Measuring Machine (CMM) and the CAD software. Unfortunately, few companies have the right CMM/CAD interface for the kind of RE design capabilities required in today's industrial environment. In addition, there is considerable overselling of hardware and software capabilities and benefits, resulting in a few companies having some bad experiences. Most of the RE related work concentrates on scanning, registration, data merging, segmentation, and surface fitting [2, 3] and there is very limited work on the RE strategies and RE software capability. Chan and Popov [4] compared various RE software for surface fitting and found, for instance, that most software was capable of smooth surface fitting but only a few were designed for obtaining sharp textured free-form surfaces. Also, existing CAD/CAM software would be used to translate point cloud capture into 3D model.

This paper demonstrates the appropriate methodology to enhance the performance within an RE environment in order to give users a better understanding of the requirements and

available strategies in RE as well as to enable the hardware and software developers to improve their products. Examples are presented to illustrate the shortcomings especially when fitting sharp textured free-form surfaces based on point clouds in order for the scanned real objects to be displayed in realistic images from arbitrary viewpoints.

2. RE PROCESS

Unlike the traditional design process, which transforms design concepts and models into products, RE builds products based on existing physical objects and exploits the advantages of CAD technologies. In many instances, capturing a complex geometric shape and translating it into a 3D CAD model that is an accurate representation of the physical object is not easy. The process of RE generally comprises five major steps (see fig 1). Each step is very important and if neglected can cause severe problems later on.

Figure 1. General stages of the RE process

The flowchart contains the following:

- Physical object
- Step 1: Estimate geometric configuration
- Step 2: Digitisation (Data acquisition)
- Step 3: Pre-processing of data
- Step 4: CAD model reconstruction — No
- Step 5: $P_{model} - P_{nominal} < Tolerance$ — Yes
- Final solid CAD model

3. STEP 1 - ESTIMATE GEOMETRIC CONFIGURATION

In the traditional RE approach, manual division and segmentation of the surfaces of the physical object represent the first operation. Essentially, the physical objects are divided into three configurations based on their geometric shape.

- Regular shaped surfaces
- Relatively smooth free-form surfaces
- Sharp textured free-form surfaces

The first geometric configuration comprises objects with regular shaped surfaces such as planes, cylinders, cones, etc, (see fig 2). These can be easily reverse engineered using ordinary measuring instruments such as callipers, micrometers, protractors, etc. In this case, most CAD modelling software packages can be used to create the object's CAD model. The second geometric configuration comprises objects with relatively smooth free-form surfaces (see fig 3). For these, a contact probe can be used to acquire both the perimeter and cross-section contours representing precisely the object's geometric shape. In this case, the scanning process is simple and can be carried out with a contact probe mounted on a CMM. This provides sufficient 2D scan lines to represent precisely the object's geometric shape

instead of scanning the entire surface. Thus, the number of data points will be significantly reduced and this will facilitate RE processes easier. Most CAD software including ProEngineer, Solid Works, AutoCAD, CATIA, etc is capable of using the curve-based approach to blend the curves created (containing the data points) to fit into a smooth free-form surface. Any symmetry or patterns in the geometry would allow utilisation of mirror/copy pattern functions of the software, which would contribute to further improvement of the efficiency of the RE process.
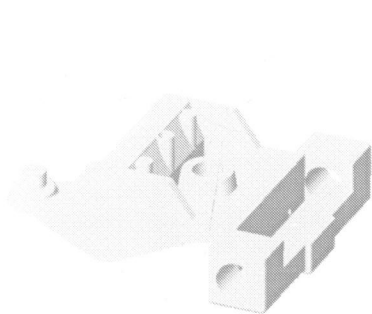
Figure 2. CAD model with regular shape surfaces only
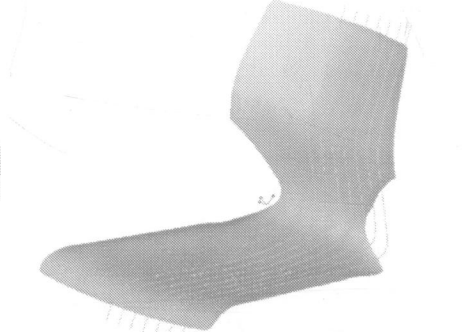
Figure 3. CAD model of a chair with smooth free-form surface

Figure 4. A prototype of a parrot insert with textured free-form surface

The third geometric configuration comprises objects with irregular shapes (see fig 4). These are time-consuming to measure by contact probe and non-contact digitising techniques must be used. Thus a huge number of points (or point cloud) is required to represent the entire surface geometry. Consequently, data acquisition and handling is made harder. As a result, probably the biggest challenge for RE is when tooling of sharp textured, free-form surfaces (see fig 4) is involved. This is because, apart from the huge amount of data and high dimensional accuracy demands, a features recognition function is required to build matching surfaces which fit properly and correctly. A study reported in [5] shows that at present most CAD software packages are capable of fitting smooth surface, but only a few are designed for fitting sharp textured free-form surface. Only RE software such as GeoMagic and RapidForm 2001 are fully capable of creating 3D models automatically with sharp texturing, otherwise the texture can be smeared or even totally lost (see fig 5).

4. STEP 2 - DATA ACQUISITION EFFICIENCY

It is vital to select the right hardware, as illustrated by fig 4, fig 5 and fig 6, and the right tool for the specific application that determines the digitiser system best suited for data acquisition efficiency. Depending on the application, either a contact digitiser such as a touch trigger probe or a non-contact digitiser such as a laser scanning head can be mounted on a CMM.

Figure 5. A Cad model created using non-contact digitising point cloud
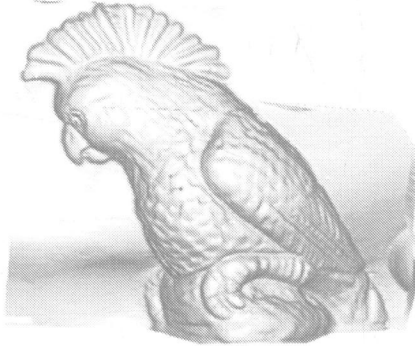
Figure 6. A CAD model created using contact digitising point cloud

The touch trigger probe functions by creating individual points as it makes contact with the object's surface. It is a very accurate digitising technique and is very flexible to reach boreholes and undercuts. However it is not suitable for digitising soft or elastic components, and is less efficient in terms of scanning speed than the non-contact digitiser. Hence, the touch probe is effectively utilised only for the first two geometric configurations as mentioned in section 3. Also, it is possible to do fast reverse engineering of extremely large components if the object has relatively smooth free-form surfaces. Hence a number of 2D scanned contours representing the object's geometric shape will be sufficient to create the required 3D CAD model. Also, it is feasible to do scanning of textured surfaces using a touch probe although some surface texture definition will be lost due to the stylus ball diameter. Fig. 6 shows a CAD model created by ProEngineer software using point cloud obtained with contact digitising. The texture obtained is very sharp because the surface is fitted from manually constructed style curves containing the data points. However, this is an extremely time consuming and tedious procedure.

In general, contact digitising should be used for scanning engineering parts such as cavity and core parts in order to achieve high accuracy for alignment and matching. As well as when a structured point cloud is required to construct sharp textured free-form surface with minimum errors, for example to study tool wear, etc. In such cases, CAD/CAM software such as ProEngineer can be used to convert the scanned data points into the required CAD model. The advantages of using contact digitising techniques include a known and common coordinate system, and a structured point cloud with an even pitch between the points.

On the other hand, using a non-contact digitiser is obviously the fastest way for data acquisition as some laser scanners can acquire up to 10,000 points per second. That is why non-contact digitising techniques should be used for fast scanning of sculptured objects such as dental products, jewellery, and works of art. However to capture the whole object, more

viewing frames are needed since the effective inclined angle of such non-contact equipment is limited. Alternatively, the position and orientation of the object must be changed several times. It may also be difficult to keep the scanning direction identical when the object orientation is changed. To deal with multiple sets of scan data, three spheres-to-spheres are preferred attached to the object in order to provide reference for coordinate transformation (so as to unify the coordinate systems of the two sets of data). However, its limitation is that all three spheres must be measured again whenever the object position is changed. Disadvantages include an unknown datum with respect to the coordinate system, the need to eliminate a huge amount of redundant data, and insufficient information obtained along edges and holes (due to discontinuities). Also, data acquired are either unstructured or appear as a grid with uneven pitch between the points. Non-contact digitisers are also sensitive to surface properties such as shininess and colour. Laser scanners based on interferometry are very accurate, but those using the triangulation principle (the so called "flying spot") are limited up to 30-40μm.

5. STEP 3 - PRE-PROCESSING OF DATA

Pre-processing of data is the most important stage as it is related to the efficiency of RE. Proper data pre-processing facilitates the RE process, whereas a poor pre-processing can make surface fitting impossible. It is important to remove any undesirable points in the regions in order to remove noise, reduce error accumulation, and improve computational efficiency.

Figure 7. Filtering in scan sets

Non-contact digitisers such as laser scanners and cameras allow only one view of an object at a time, usually limited in range, to be digitised. Therefore, several scans from different viewing frames, part positions, or orientations are frequently required to scan the whole object. In principle, registration rotates and translates a point cloud in order to align it accurately with another point cloud using overlapping areas (in some cases using three spheres-to-spheres as reference locator). Some of the RE software packages offer initial manual alignment. Pairwise automatic registration is preferred to establish reliable point correspondence between the two scans using the ICP (iterative closest point) method that was reported in [5] and verified in [6]. It is worth noting that the registration accuracy will be greatly increased if a large number of points are used for ICP registration and that this process is repeated once the full measurement data set is obtained.

Several types of data filtering and tolerancing have been used in RE. For example, when an object is scanned several times from different settings some data points will overlap with one

another after merging. The overlapping/duplicating points need to be removed or filtered within a carefully selected tolerance (see fig 7). Also, it is important to specify tolerance to the data point. This is to filter out noise as well as to remove the spikes lying outside a given tolerance, thus smoothing the raw data. Further filtering/tolerancing can be achieved by considering the scan line curvature (see fig 8). However, smoothing, filtering, and tolerancing should be carried out very carefully because it could result in loss of surface texture information. At a later stage, tolerancing a triangulated model will reduce the number of triangles while maintaining the shape to the specified tolerance.
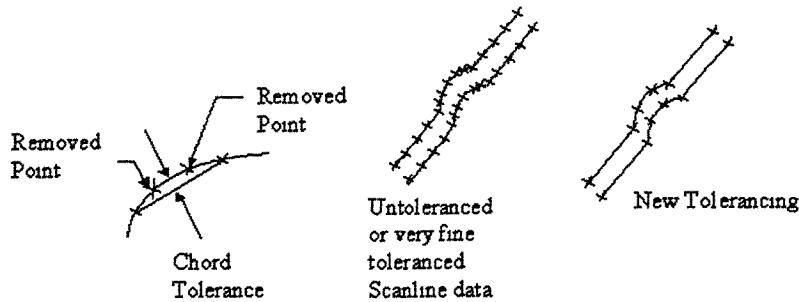
Figure 8. Tolerancing of scan lines

6. STEP 4 - CAD MODEL RECONSTRUCTION

The process of converting these data points into a useful 3D geometric model is often referred to as surface reconstruction. To create a 3D CAD surface model of a textured free-form surface, the point cloud needs to be transformed to polygons-to-surface. When converting point cloud to polygonal meshes, there is usually some bad data such as topological holes, duplicated triangles, degenerated triangles, degenerated edges, and inconsistent edges. Similarly, when converting polygonal meshes to surface models, surface anomalies occur such as orientation of surface normals, topological holes, duplicated or degenerated polygons and vertices, and unconnected meshes. Detection and repair of these bad data is desirable in order to obtain watertight polygonal meshes and surface models.

Surfaces are usually reconstructed by approximation methods using specialised algorithms in order to overcome noise errors and reduce computational complexity. It is important that the calculation must be sufficiently detailed to provide a good representation of the actual surface texture. Consequently, the fitting technique employed in the geometric modelling must be able to perform 3D wrapping of surfaces based on the point cloud. Subdivision is suitable because it proceeds from a coarse mesh to finer ones and thus can approximate the data points at multiple levels of details. On the commercial front, software that focuses on using dense (scanned) data of arbitrary topology to produce accurate surfaces can be semi or fully automated. The semi-automated techniques take point cloud data sets as input. These approaches begin by identifying a subset of points that are to be approximated, and require a user-guided process such as projection of the points to a manually constructed base plane or surface. These approaches can be difficult to use because the user must have specialised skills. Another common problem with semi-automated systems is lack of global control with

local modifications often resulting in global inconsistencies. The automated techniques take the arbitrary point cloud data sets as input and reconstruct the surface from the entire point cloud without any human intervention. One example is an RE software called Geomagic which automatically rebuilds a surface around a given point set. Depending on the point distribution, the initial surface may or may not be the final surface the user has in mind. The typical cases of disagreement are tunnels through the shape that are not supposed to be filled, or webs connecting finger-like extensions. Geomagic offers a few semi-automated editing tools that locally and globally modify the shape and create satisfactory solutions with ease. It is important to highlight that not all RE software is capable of accurately translating dense point cloud to 3D CAD model with detailed texturing during surface fitting, as shown in fig 9(a) and 9(b). Fig 9(a) shows a 3D CAD model with sharp textures on the surface constructed by rapid fitting. If another RE software is used to fit a surface with the same point cloud, the sharp textures are missing from the 3D CAD model as shown in fig 9(b). To get the sharp textures, too many man-hours of lofting technique [7] are needed in order to create and manipulate primitive shapes and surfaces. Software with lofting technique will not work well for surface design fitting and is mostly used for surface styling, surface modelling, and aesthetic designing of products in the aircraft, automotive, and footwear sectors. A study on the performance of RE software to do surface fitting is reported in [4].
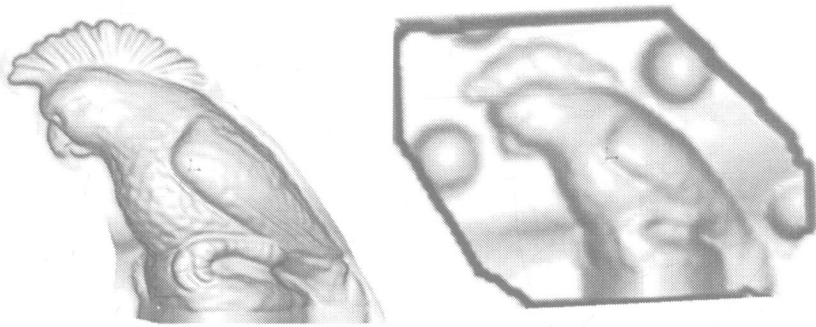
Figure 9. Cad models created using (a) 3D and (b) 2.5D wrapping of point cloud

7. STEP 5 – VERIFICATION OF CAD MODELS TO POINT CLOUD

When the point cloud is converted to polygons-to-surface model, cummulative errors such as scanning accuracy, overlapping errors, and modelling approximations and truncated errors, are inevitably introduced. Analysing the deviations between nominal polygonal meshes and/or surface models to the actual point cloud (after removing the outliers) is referred to as part-to-CAD verification. This is to maintain all detailed textures and to obtain the required reconstruction accuracy without losing much of the original data. This is because the digitised points can be manipulated by decimation and/or filtration to remove overlapping points and duplicate information. Also, by shifting (pooling or pushing) the control points of the surface created, further refinement of the CAD model can be done to maintain some detailed features and repair the local deviations.

8. CONCLUSIONS

The effective way for RE is a strategy for creating a CAD model from a physical object that cannot has a ready recipe. This is due to the diversity of the geometric descriptions and customers requirements as explained. In order to increase the effectiveness of RE, it is suggested that the geometric configuration of the objects should be divided into three types: regular shape, smooth free-form surface, and textured free-form surface. The general strategies and means for effective RE for each of these types, including relevant examples, are set out in the paper. The whole process for RE has been divided into five stages as follows: estimation of the geometric configuration, data acquisition, data pre-processing, CAD model reconstruction, and evaluation of the CAD model.

This paper demonstrates that choosing the right RE strategy, together with thorough consideration and careful selection of hardware and software can effectively enhance performance.

REFERENCES

1 **Menq C, Chen FL.** (1996) Curve and surface approximation from CMM measurement data. Computers and Industrial Engineering, Vol. 30, No. 2, pp. 211-225.

2 **Varady T, Martin RR, Cox J.** (1977) Reverse engineering of geometric models – an introduction. Computer Aided Design, Vol. 29, No. 4, pp. 255-268.

3 **Milroy MJ, Bradley C, Vickers GW.** (1977) Segmentation of a wrap-around model using an active contour. Computer Aided Design, Vol. 29, No. 4, pp. 299-320.

4 **Chan FMM, Popov IE,** (2001) A study of some reverse engineering software. Proceedings of National Conference on Manufacturing Research, pp. 225-230.

5 **Ristic M, Brujic D.** (1997) Efficient registration of NURBS geometry. International Journal of Image and Vision Computing, Vol. 15, pp. 925-935.

6 **Brujic D, Ristic M.** (1997) Analysis of free form surface registration. Proceedings of the Institution of Mechanical Engineers-Part B, Vol. 211, pp. 605-617.

7 **Filip D, Ball T.** (1990) Procedurally representing lofted surfaces. IEEE Computer Graphics Applications Vol. 9, No. 2, pp. 27-33.

A new approach for addition of draft angles on well-rounded polyhedral

Y YAN and S T TAN
Department of Mechanical Engineering, The University of Hong Kong, Hong Kong

ABSTRACT

Moulding is a common process to mass-produce parts for complex structures in the industry. The automation of mould design and manufacture using computer aided design and computer aided manufacturing (CAD/CAM) techniques draws the attention of many researchers. However, little attention is given to the automation of draft angles addition process, and existing algorithms are still having limitations. In this paper, a new approach for draft angles addition is introduced.

1. INTRODUCTION

Moulding has been used for centuries to mass-produce parts. With the development of computers in recent days, computer aided design and computer aided manufacturing (CAD/CAM) techniques have been used by industry to increase the efficiency of mould design and manufacturing. Many CAD/CAM systems have been developed to deal with the moulding process, including die-casting, metal casting and injection moulding.

In the moulding process, the moulded part must be easily removed from the mould. The ease in removing the moulded part from the mould depends on factors like, parting direction, parting line, undercuts and draft angles. A large volume of research has been carried out in determining the above factors. In the area of draft angles addition, only a few studies were conducted to investigate the problem, and draft angles addition is still manually designed in most cases.

Mitchell (10) pointed out that mouldability depends on draft angles. He stated that the surface parallel to the drawing direction must be tapered in order to help the ejection of parts. This taper is called the draft. Campbell (2), in his book, stated that sufficient inclination on the contact surfaces between a moulded part and its mould cavity or mould core is the primary concern in mould design. Smith and Lee (11) proposed a framework of a computer-aided pattern design system for the moulding process. In their proposal, they stated that draft angles could be added on a moulded part by changing the normal vector of a planar surface through variational geometry. Lee and Lee (8) introduced the major steps in a draft angles addition

process, including an idea of adding draft angles on cylindrical and conical surfaces. Ganter and Gudmundsson (5) implemented a draft angles addition system for simple polyhedral using the idea discussed in (11). Tokuyama and Bae (12) developed an approximate method to add draft angles on free-form surfaces by using the isocline curve and the isocline surface.

In most of the research done on this topic, the focus is on single part geometry only. In manufacturing, moulded parts will be assembled with other components to form a product. The fitting of a moulded part with other components will have a great influence on the quality of final product. Addition of draft angles causes the original part geometry to be deformed in one way or another. As a result, parts with draft angles added may not have a good fitting with other components. A system that allows draft angles addition with in an assembly is required to solve this problem. Such a system should be able to add draft angles to the required surfaces automatically. It should prevent potential misalignment, misfit of joints and interference. A proper draft angles addition algorithm will form the basis of this system, however there are many limitations in existing algorithms. Problems may appear in cases containing concave blending surfaces. In this paper, a new approach for adding draft angles to a single part is first introduced with a view to solving the assembly problem in the long run. This new approach can be applied to the common mechanical components rounded with constant radius.

In the moulding process, the moulded parts must be rounded to assist mould removal and reduce stress concentration (2, 10). In normal design process, addition of draft angles should be done before the edges and corners are rounded. However, the draft angles are often omitted by the designers (10). The manufacturers, as a result, have to add draft angles to a rounded part. Unfortunately, rounded part may contain concave blending surface, which cannot be handled by existing algorithms. Manual adjustment is required to add the draft angles. Problems with concave blending surfaces then become a critical issue for the automatic draft angles addition system. The proposed new approach is focused on geometry that contains well-rounded vertices and edges.

2. A NEW APPROACH FOR DRAFT ANGLES ADDITION

2.1. Terminology

Some special terms have been used in this paper, and they are introduced as follows. A blending surface is a type of surface that is formed by a sweep operation. This type of surface is mainly used to represent the rounding and fillet feature of a model, which may appear in the form of a cylindrical, toroidal or spherical surface. An offset surface intersection curve (OSIC) is the intersection of the offset of two base surfaces of a blending surface. This intersection curve is used as the guide path to form the blending surface in the sweeping operation. A network of blended entities is a group of edges and vertices that will be blended (1). A blending network is the network of blending surfaces that is formed by a network of blended entities. Control points are a set of points that have the characteristic of governing the shape of a model. They can be divided into two groups, the moving control points and the fixed control points. A translation vector is the direction that a moving control point will be translated. For each control point, there may be more than one translation vector, and the combination of these translation vectors is called the 'combined translation vector'.

2.2. Basic Idea

(a) A model without rounding

(b) A rounded model with constant radius

(c) The OSIC of a blending network

(d) The control points of a model

(e) The translation of control points

(f) The resultant shape after translation of control points
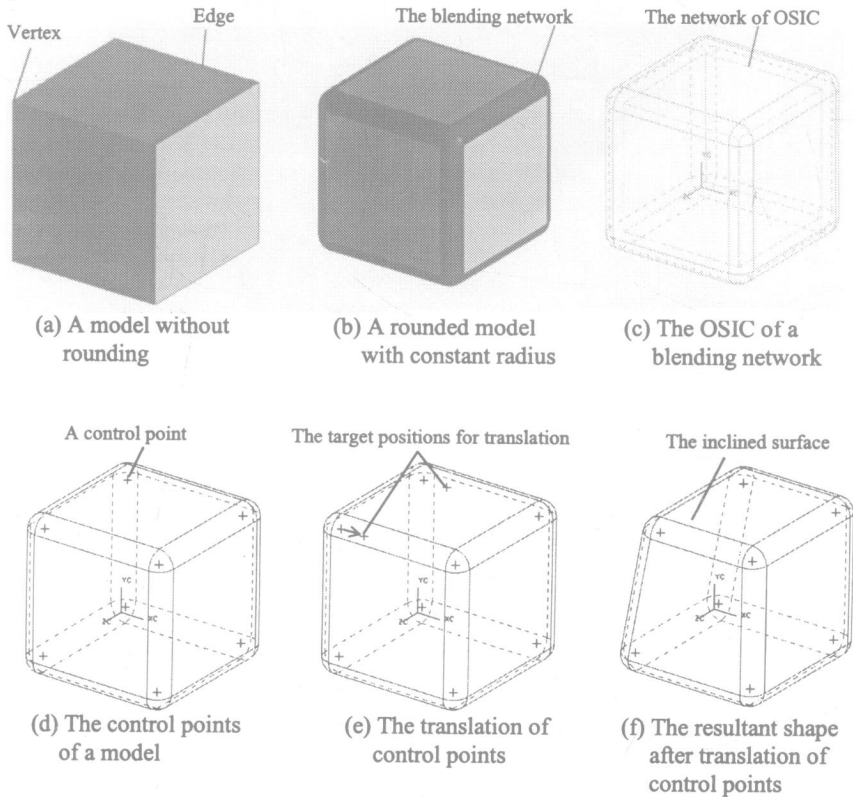
Figure 1: The process of control point translation

In this study, mechanical components are restricted to geometry that is formed by a combination of spherical, cylindrical, conical and planar surfaces before blending. For a mechanical component without any rounding, the geometry is governed by the edges and the vertices of a part (Figure 1a). After a part is rounded with a constant radius, the edges and the vertices become a blending network (Figure 1b). By controlling the shape of the blending network, the part can be deformed so as to have draft angles added. For constant radius rounding, the blending network is formed by the rolling ball method (3, 4). The rolling ball method produces a blend by rolling a ball such that the ball is always in contact with the two base faces. The locus of the centre of the rolling ball is the OSIC, as shown in (Figure 1c). The OSIC formed is a combination of mainly straight lines and arcs for mechanical components that are targeted in this study. Therefore, transforming the endpoints (Figure 1d) of the lines and arcs can induce the deformation of the OSIC (Figure 1e). This leads to the deformation of the blending network, which governs the shape of the rounded mechanical component (Figure 1f). The endpoints of lines are called the moving control points of a part.

The centre points of the arcs in a network of OSIC are also important in controlling the geometry of a part. They are never translated and they act as reference point only, so they are called the fixed control points. A moving control point, on the other hand, will be attached to all the surfaces around it. The attachment criteria of the moving control point are dependent on the surface adjacencies. A moving control point is one of the endpoints of a line or arc segment of the OSIC. This line or arc segment is the guide path of a cylindrical or toroidal surface in the blending network. Therefore, the moving control point is attached to this cylindrical or toroidal surface. It is also attached to all the surfaces adjacent to this cylindrical or toroidal surface. The proposed draft angles addition algorithm is based on translating these control points and changing the radius of the cylindrical or toroidal surfaces.

(a) The original model

The reference plane

The parting line

(b) The control points formed

A line segment of OSIC

A control point

(c) The translation of control points

The original network of OSIC

The deformed network of OSIC

A translated control point

(d) The reformation of surfaces

A reformed cylindrical surface

A spherical surface

The curve where two surfaces intersect

A cylindrical surface

(e) The trimmed surfaces

A trimmed cylindrical surface

(f) The deformed model

Figure 2: Deformation of a model having draft angles added

2.3. Deforming the geometry

Draft angles are added to a surface that is parallel to the parting direction and to a surface that makes an inclination less than the draft angle with the parting direction. The draft angles can be added by translating the position of moving control points as discussed in section 2.2. Figure 2 shows how the geometry of a 3D model is deformed in order to have draft angles added. It uses a rounded cube as an example. The parting line of the model is assumed to be on the mid-plane, and this plane is used as the reference plane (Figure 2a). Figure 2b shows the network of OSIC and the control points formed. All control points in this example are moving control points because the network of OSIC is composed of line segment only. Because of the mid-plane parting line, all the vertical surfaces in the example are split into two, and so are the vertical OSIC line segments. Additional moving control points are formed at the split locations. Figure 2c shows that a deformed network of OSIC is produced by translating the control points. Surfaces of a deformed model is reformed according to the position of the translated control points (Figure 2d). A spherical surface is centred at its attached control point. A cylindrical surface is reformed along the axis that passing through the two attached control points. The planar surfaces in Figure 2d and 2e are omitted in order to show the internal of a model. After reforming the surfaces, the surfaces are trimmed up by surfaces adjacent to them. In the enlarged view of Figure 2d, the spherical surface at the centre is trimmed up by three cylindrical surfaces adjacent to it. It is because only the shaded part of the spherical surface forms part of the deformed model. The surfaces are trimmed and are shown in Figure 2e. The trimmed surfaces are connected to form the deformed model in Figure 2f.

Figure 3 is used to demonstrate how a moving control point is translated. It shows a rounded 3D solid model without draft angles added. The parting direction is in the Y direction. S_A and S_B are two vertical planar surfaces that intersect with an angle of 70°. S_D is a cylindrical surface that does not belong to the blending network. S_C is a cylindrical surface and S_E and S_F are toroidal surfaces that belong to the blending network. Point S and P are the centres of S_E and S_F respectively. They are the fixed control points. M, N, Q, T are the moving control points. In Figure 3b, solid lines in black represent the section of the original model, and solid lines in grey represent the section of the model after draft angles are added. Dashed centre lines in black are the centre lines of cylindrical surfaces in the original model. The centre lines of the cylindrical surfaces in the model after draft angles are added are shown in grey. The following variables are used in the calculation of control point translation.

θ	Draft angle
β	Angle between two translation vectors
h	Distance of the control point from the reference plane
r	Radius of blend for the blending network
Δ	Distance translated by a moving control point
ε	Deviation of the intersection of an axis of a cylindrical surface with the reference plane before and after draft angles are added
$\bar{p}$	Parting direction
$\bar{n}$	Normal vector of a planar surface
$\bar{v}$	Translation vector formed by a planar surface
$\bar{u}$	Combined translation vector of a moving control point
R	Original radius of a cylindrical surface
R'	Modified radius of a cylindrical surface

R$_{major}$ Original major radius of a toroidal surface
R'$_{major}$ Modified major radius of a toroidal surface

(a) Projection on XZ plane

(c) Original Model

(b) Projection on XY plane
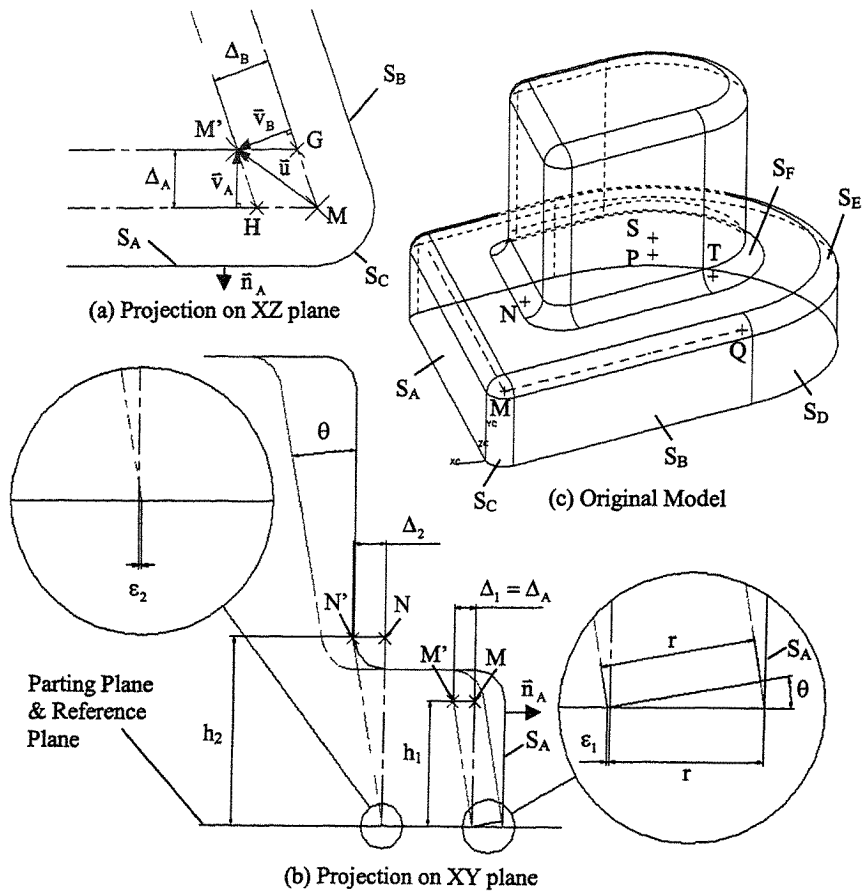
Figure 3: The translation of control points

For the moving control point M in Figure 3b, it is the centre of an arc in the original section with distance h$_1$ from the reference plane. Its position moves to M' after a draft angle is added. A vector $\bar{v}_A$ is used to represent the translation of the moving control point M to M' in the XY plane. The direction of $\bar{v}_A$ is controlled by the normal vector $\bar{n}_A$ of S$_A$ and the parting direction $\bar{p}$, with the following generalised relation:

$$\frac{\bar{v}}{|\bar{v}|} = \pm(\bar{n} \times \bar{p}) \times \bar{p}$$

The plus/minus sign is respectively used for the protrusion and depression portion of the model.

When looking at the enlarged area on the reference plane under M, surface S_A before (in black) and after (in grey) inclination is offset by the radius of blending r. The offset surfaces are drawn in dashed centre lines. A small displacement ε_1 exists between the intersections of the offset surfaces with the reference plane. This ε_1 is equal to $r(\sec\theta - 1)$. The displacement of the surface S_A at a distance of h_1 from the reference plane is equal to $h_1 \tan\theta$. The distance between M and M' is Δ_1, which is a combination of $h_1 \tan\theta$ and ε_1. Because ε_1 is displaced in the direction of $\bar{v}_A$, therefore Δ_1 is equal to $h_1 \tan\theta + \varepsilon_1$, and Δ_1 is the magnitude of $\bar{v}_A$.

In comparing the situation of M and N in Figure 3b, M corresponds to the geometry of a convex blending surface, and N corresponds to the geometry of a concave blending surface. The displacement ε_1 for M is displaced in the direction of $\overline{MM'}$, so that $\Delta_1 = h_1 \tan\theta + \varepsilon_1$. Similarly, ε_2 for N is displaced opposite to the direction of $\overline{NN'}$, where $\Delta_2 = h_2 \tan\theta - \varepsilon_2$. The magnitude of the translation vector $\bar{v}$ for moving control points on convex and concave blending surfaces can be generalised as:

$$|\bar{v}| = h \tan\theta \pm r(\sec\theta - 1)$$

The plus sign is employed when the moving control point is formed by a convex blending surface, and minus for a concave blending surface.

A planar surface that is parallel to the parting direction or makes an angle less than the draft angle with the parting direction forms a translation vector. Only this type of planar surface can form the translation vector, and only one translation vector is formed by each planar surface. The translation vector of a planar surface affects the translation of moving control points that are attached to the planar surface. A moving control point may be attached to more than one planar surfaces that have a translation vector. The combination of the translation vectors that affect the moving control point is called the combined translation vector $\bar{u}$.

Figure 3a shows how the translation vectors are combined. The translation of M to M' is affected by two translation vectors $\bar{v}_A$ and $\bar{v}_B$ of S_A and S_B respectively. The combined translation vector $\bar{u}$ is not just a simple addition of $\bar{v}_A$ and $\bar{v}_B$. The new position M' of the control point M is located at the intersection of the offset centre axes of the cylindrical surfaces above S_A and S_B, which are grey in colour. The original centre axes in black are offset by vectors $\bar{v}_A$ and $\bar{v}_B$. The magnitude of $\bar{v}_A$ and $\bar{v}_B$ are Δ_A and Δ_B respectively. In order to obtain $\bar{u}$, point G and H are employed. M, G, M' and H form a parallelogram, with $\bar{u} = \overline{MG} + \overline{MH} = \overline{HM'} + \overline{GM'}$, where $\overline{MG} = \overline{HM'}$ and $\overline{MH} = \overline{GM'}$. Therefore, $\bar{u}$ can be found by adding $\overline{GM'}$ and $\overline{HM'}$. $\overline{GM'}$ and $\overline{HM'}$ can be found by using $\bar{v}_A$, $\bar{v}_B$ and the angle β, which is the angle between $\bar{v}_A$ and $\bar{v}_B$.

$$\frac{\overrightarrow{GM'}}{\left|\overrightarrow{GM'}\right|} = \bar{v}_A \times \left(\bar{v}_B \times \bar{v}_A\right), \qquad \left|\overrightarrow{GM'}\right| = \frac{\left|\bar{v}_B\right|}{\cos(\beta - 90°)} = \frac{\left|\bar{v}_B\right|}{\sin\beta} = \frac{\Delta_2}{\sin\beta}$$

$$\frac{\overrightarrow{HM'}}{\left|\overrightarrow{HM'}\right|} = \left(\bar{v}_B \times \bar{v}_A\right) \times \bar{v}_B, \qquad \left|\overrightarrow{HM'}\right| = \frac{\left|\bar{v}_B\right|}{\cos(\beta - 90°)} = \frac{\left|\bar{v}_B\right|}{\sin\beta} = \frac{\Delta_1}{\sin\beta}$$

where β is the angle between the translation vectors $\bar{v}_A$ and $\bar{v}_B$, where $\bar{v}_A \cdot \bar{v}_B = \left|\bar{v}_A\right|\left|\bar{v}_B\right|\cos\beta$.

For Q and T, their translations are due to one single translation vector only, because only planar surfaces can produce translation vectors. Using Q as an example, the translation is governed by the planar surface S_B and the cylindrical surface S_D. Only S_B provides a translation vector $\bar{v}_B$ that affects Q. The cylindrical surface S_D and the toroidal surface S_E are deformed by changing the radii. For the toroidal surface S_E, the major radius R_{major} is the distance between P and Q. As Q is translated with $\bar{v}_B$, the major radius of S_E will be reduced by $\left|\bar{v}_B\right|$. The new radius R'_{major} of S_E is equal to R_{major} - $\left|\bar{v}_B\right|$. The same situation will also apply to S_F The generalised relationship between the deformation of the toroidal surface and the translation vector is:

$$R'_{major} = R_{major} \pm \left|\bar{v}\right|$$

The deformation of the surface S_D is obtained by changing its top and bottom radius, R, according to the following equation.

$$R' = R \pm h(\tan\theta)$$

For both of the equations, the use of plus or minus sign is dependent on the cylindrical surface. The plus sign is used when the cylindrical surface or the cylindrical surface that adjacent to the toroidal surface is concave, and the minus sign is used when it is convex. The value h is the distance of the top or bottom centre of a cylindrical surface from the reference plane, which is the distance of P and S from the reference plane.

It can be observed that the change in radii and the displacement translated by the control points are dependent on the draft angle θ, and the distance from the reference plane h. For the same θ, when the control point is further away from the reference plane, the displacement for translation increases. This produces an inclination on the deformed moulded part, which have the same results as if a draft angle is added to the moulded part.

3. DISCUSSIONS AND RESULTS

The above problem for draft angles addition on moulded part is developed and implemented on the UniGraphic V17 CAD/CAM system.

Moulded parts composing of planar, spherical, cylindrical, conical, toroidal and blending surfaces were tested with the procedures. All the moulded parts are assumed to be divided into two mould halves, the upper mould and the lower mould. The parting line is the closed-loop in thick dark line. Surfaces above the parting line belong to the upper mould half and those below the parting line belong to the lower mould half. The parting direction for the upper mould is assumed to be pointing upwards and downwards for the lower mould.

Test part 1 (Figure 4a) shows moulded parts with a complicated blending network. Test part 2 (Figure 4b) shows a transition of surfaces from cylindrical surface to conical surfaces. The result shows that all the cylindrical surfaces are transformed to conical surfaces. The toroidal surfaces due to rounding can match with the conical surfaces perfectly. Test part 3 (Figure 4c) shows a moulded part with both horizontal and vertical concave surfaces. The horizontal concave surfaces are the type of surfaces that cause problems in existing draft angles addition algorithms. The concave surfaces are mainly located in the depression of the part. Test part 4 (Figure 4d) demonstrates the case where the parting lines are not lying on the same plane. The reference plane of both the upper and lower mould is kept at the bottom of the model, in order to ensure that the surfaces are properly connected. Test part 1 and 4 are originally obtained from NIST's design repository.

The details of the test parts are listed in the table below, which shows the number of control points generated and the number of translations done for each part. The number of radii modified is also shown. The CPU time taken for each of the examples is also tabulated. A computer with Pentium 4 1.7 GHz CPU, 256 MB RAM is used to test the parts. The result shows that the CPU time is highly dependent on the number of moving control point, and the effect of the number of fixed control point is not significant.

Detail of the tested parts

Part	Dimensions	Radius of Blend	No. of Fixed Control Point	No. of Moving Control Point	No. of Modified Radius	No. of Translation	No. of Surfaces	CPU Time (s)
1	7000 × 15000 × 6000	150	0	72	0	136	114	34.73
2	68 × 68 × 30	1	11	0	31	0	32	3.84
4	60 × 100 × 30	1	52	68	88	68	97	29.68
5	4990 × 5260 × 1500	20	8	72	14	132	114	36.49

Original model   Model with draft angles added

(a) Test part 1

(b) Test part 2

(c) Test part 3

(d) Test part 4
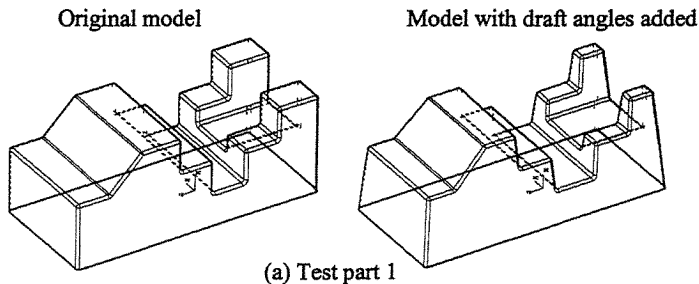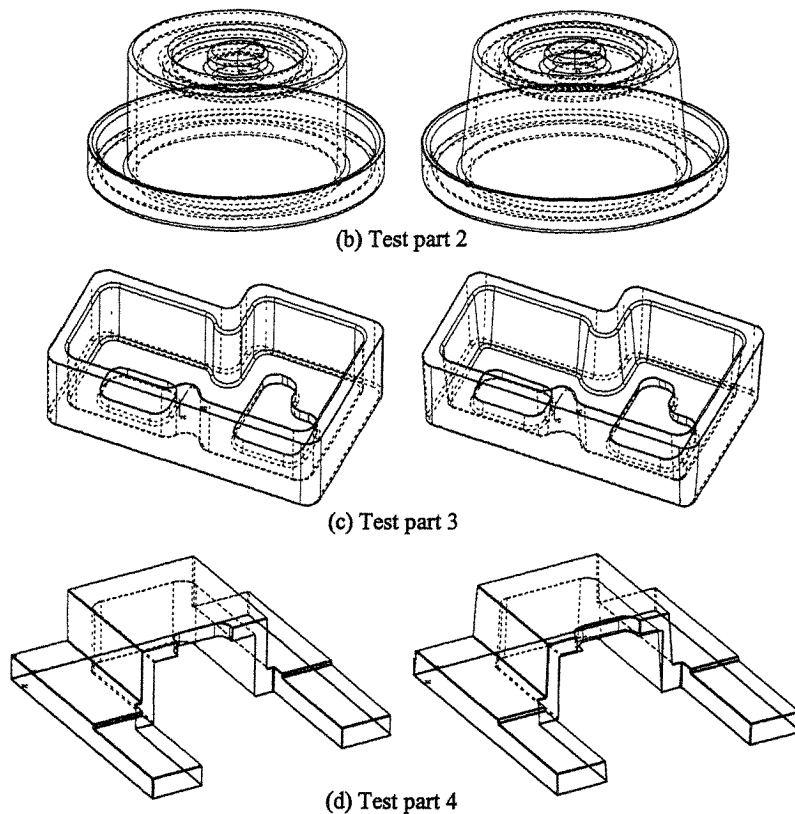
Figure 4: Test Results

4. CONCLUSION

The results in the above section show the effectiveness of the new approach in performing the operation of draft angles addition. They also illustrate that some of the geometry problems besetting existing draft angles addition algorithms are now solved by the proposed algorithm. However, there are still limitations in this approach. Cases with too complicated surfaces are still not solved, because some of these are free-form surfaces, as the case shown in Figure 5. Some areas are also not yet tested with the new approach. The idea of this approach is to transform the control points of the geometry, in order to induce the inclination. This is similar to the idea of NURBS surfaces where the shape of the surfaces can be deformed by moving the control points.

The approach discussed in this paper deals mainly with geometry that is rounded with constant radius. However, geometry with sharp edges and vertices can also be taken as those rounded with a constant radius of zero unit. The applicability of this approach on such type of geometry will need detail investigation.
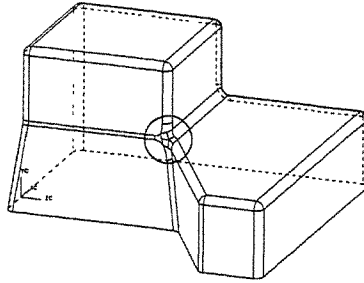
Figure 5: Complex sub-divided blending surface

In the approach proposed in this paper, the user is no longer required to select which surfaces need inclination or tapering. Further automation can be done by automating the selection of the suitable reference plane location. As stated in section 2, the distance of the control points from the reference plane governs the distance of translation and further controls the deformed shape of the moulded part. Special care has been taken to choose the position of the reference plane for all the tested parts, in order to ensure that no misalignments and disjoints occurs at the position of parting lines. Further studies can be done in selecting proper reference planes location automatically. This can increase the degree of automation of the system for a single part geometry.

The deformation of a moulded part after draft angles addition may induce misalignment and misfit of the parts in the assembly. Further studies are required to identify these potential problems and provide adjustment for the parameters needed. This can further lead to the development of a system that can provide draft angles addition for moulded parts in an assembly model.

ACKNOWLEDGEMENTS

REFERENCES

(1) **Braid, I. C.** (1997) Non-local blending of boundary models. *Computer Aided Design.* Vol. 29, pp. 89 – 100

(2) **Campbell, Paul D. Q.** (1996) *Plastic Component Design.* New York: Industrial Press.

(3) **Chen, Q., Barnhill, R. E.** and **Farin, G. E.** (1993) Constant-Radius Blending of Parametric Surfaces. *Computing Supplementum.* Vol. 8, pp. 1 – 20

(4) **Choi, B. K.** and **Ju, S. Y.** (1989) Constant-radius blending in surface modelling. *Computer Aided Design.* Vol. 21, pp. 213 – 220

(5) **Ganter, M. A.** and **Gudmundsson, B. B.** (1994) Computer-aided drafting of Mold Patterns. *AFS Transactions.* Vol. 101, pp. 891 – 895

(6) **Gavankar, K.** and **Henderson, M. R.** (1982) Graph-based extraction of protrusions and depressions from boundary representations. *Computer Aided Design*. Vol. 22 pp. 442 – 450

(7) **Gossard, D.** and **Light, R. A.** (1982) Modification of geometric models through variational geometry. *Computer Aided Design*. Vol. 14, pp. 209 – 214

(8) **Lee, S. H.** and **Lee, K.** (1998) An integrated CAD system for mold design in injection molding processes. In *Computer-aided design and manufacturing of dies and molds*. ASME, PED-Vol. 32, pp. 257 – 271

(9) **Lee, K. S., Ye, X. G.** and **Fuh, J. Y. H.** (2001) A hybrid method for recognition of undercut features from moulded parts. *Computer Aided Design*. Vol. 33, pp. 1023 – 1034

(10) **Mitchell, P.** (ed.) (1988) *Tool and Manufacturing Engineers Handbook, vol. 8, Plastic Part Manufacturing*. Fourth edition. Society of Manufacturing Engineer.

(11) **Smith, C. T.** and **Lee, K.** (1986) Computer aided pattern design for casting process. *AFS Transactions*. Vol. 94, pp. 21 – 28

(12) **Tokuyama, Y.** and **Bae, S.** (1999) An approximate method for generation draft on a free-form surface. *The Visual Computer*. Vol. 1, pp.1 – 8

Machining Technology

Radial force and hole oversize prediction in drilling using traditional and neural networks

V KARRI and **T KIATCHAROENPOL**
School of Engineering, University of Tasmania, Australia

ABSTRACT

A reliable method to predict drilling performance is the traditional mechanics of cutting approach, which mathematically relates various oblique cutting theories. However, the quantitative reliability of such conventional models is determined by reliable information of the numerous geometric features of the drill and quantitative accuracy of the data bank for a given work material. It is noted that when inevitable eccentricity and drill deflections are incorporated into the analysis, the complexity of such models is increased. In this paper, neural networks are used to perform prediction tasks due to their abilities to model the non-linear problem and their abilities to handle noisy data. A set of comprehensive drilling tests is carried out to train and test the architectures. This work shows a clear quantitative superiority of two entirely different feed forward neural networks models for prediction purpose to estimate drilling performance. A novel network that optimizes layer by layer is used as a predictive tool and compared with a more established back propagation network.

1 INTRODUCTION

Drilling is an important and indispensable manufacturing operation. The asymmetry in the drill point geometry as a result from manufacturing errors and the sharpening process are inevitable; consequently, unbalanced radial forces and oversized holes are created. To predict these forces and hole oversize in the past, traditional mechanics of cutting approaches and empirical approaches have been used. In recent years, the applications of artificial neural networks are used for cutting tool wear estimation, tool condition monitoring, vibration control and surface finish detection [1]. While these predictions are comparable with

conventional models, 'simultaneous' estimation of more than one performance feature is quite often necessary for 'on-line' control of a machining process. The conventional mechanics of cutting models are efficient to the extent of predicting individual performance but can not estimate different performance features 'simultaneously'. A brief description of unified mechanics of cutting and empirical approaches to drilling performance prediction are carried out in this work before the capabilities of neural network models are presented.

2 UNIFIED MECHANICS OF CUTTING ANALYSIS AND EMPIRICAL MODELS FOR DRILLING PERFORMANCE

The thin shear zone (plane) analysis for drilling [2-12] uses an elemental technique adopted to allow for changes in tool geometry and cutting speed with radius for different points on the lips and chisel edge. The geometry for general-purpose drills is shown in Fig.1. Figure 1 shows basic geometry and variables involved in a general-purpose drill. The cutting action in the lip region was treated as a number of elemental 'classical' oblique cutting elements [12-13], each with different normal rake angle (α_n), inclination angle (i) and resultant cutting velocity (V_w) depending on the mean radius of the element as shown in Fig.2.

It is usual practice when predicting the forces in drilling to use the oblique model to represent the lip edge and the orthogonal model to represent the chisel edge. Hence the chisel edge can be modelled in two dimensions but the added complexity of three dimensions is required for the lip region. The angles α_n and i were found from the commonly specified drill point features 2p, 2W, δ_o and D and the mean radius of the element r (Fig.1).
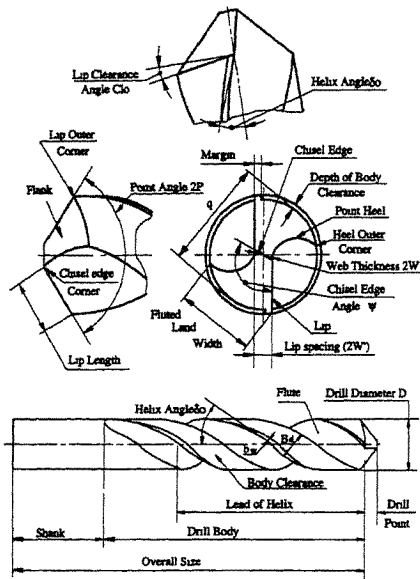
Figure 1. Geometry of general drill   Figure 2. Elemental deformation forces [3]
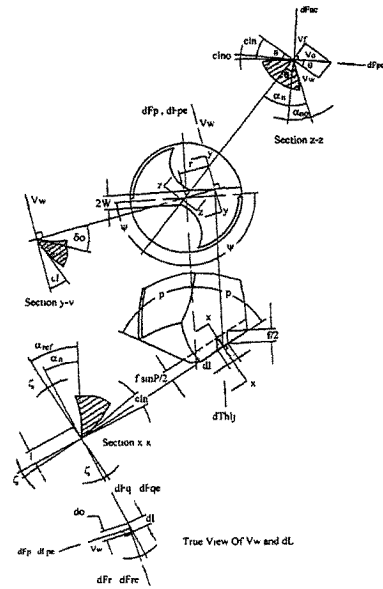
The elemental deformation forces dF_P, dF_Q and dF_R (Fig.2) were then evaluated from the 'classical' oblique cutting equations [12,14-15] given the elemental area of cut dA and the basic cutting data such as shear stress τ and the chip length ratio r/l. The edge forces [14-15] were also evaluated to give the total force on each element. The forces thus found were used to establish the elemental side forces in x and y directions. Summing up the elemental values of side forces, the total side forces generated by the lips during drilling were then predicted [14-15].

The cutting edge in the chisel edge region was also divided into a number of elements. The chisel edge was approximated to a straight line perpendicular to the drill axis and the elemental static chisel edge normal rake angles α_{nc} [12,14-15] were treated as constant for all points on the chisel edge and numerically equal to the half of the wedge angle at the chisel edge at the drill 'dead centre'. The chisel edge wedge angle could be obtained from measurement of the drill, for the unspecified flank shape of a general-purpose drill. Due to the high negative rake angles and low cutting velocities encountered at the chisel edge, a discontinuous orthogonal cutting model was applied [12,13].

The elemental chisel edge length dLc, the mean radius r, dynamic angles and cut thickness at each element for the selected number of elements could be obtained, hence the elemental side forces on the chisel edge could be determined by summation of the elemental side force values. The total side forces on the drill as a whole were found by summing the corresponding values in the lip and chisel edge regions.

The different between one half drill and another half in x and y directions are combined to formulate total unbalanced radial force. Besides, based on sophisticated investigations of drill geometry and material removal, the predictive model for hole oversize can be built by calculating these drill geometric features [13,14-18].

It should be noted the commonly specified drill point features 2p, 2W, δ_0 and D and r, together with cutting conditions N and f should be given to predict the unbalanced radial force and hole oversize in drilling. The accuracy of the mechanics of cutting approach depended on these features along with the orthogonal cutting data bank. Therefore the accuracy of this traditional approach to predictions was found to be dependent on the reliable orthogonal cutting data bank, the accuracy of the edge forces and the reliable estimation of drilling geometrical features. Neural network modelling involved fewer parameters for the simultaneous prediction of unbalanced radial force and hole oversize as discussed below.

3 BRIEF DESCRIPTION OF BACK PROPAGATION (BP) AND OPTIMIZED LAYER BY LAYER (OLL) NEURAL NETWORK ARCHITECTURES

While the specified literature provides adequate theory on the neural network models studied in this paper, it is useful here to consider the basic theory associated with each of these neural networks with an understanding of the industrial application applied. It is important to note that while the objective of each neural network is to predict the values of unbalanced radial force and hole oversize in drilling, the architecture and algorithms used by each network to achieve this are significantly different. A brief note on the BP and OLL networks is discussed below.

The standard back propagation network [19-24] comprises 3 layers of processing elements, fully feed-forward connected (Fig.3). With the sigmoid on the hidden layer, only the basic equations are :

$$y_k = \sum_{j=1}^{H} u_{kj} z_j \qquad\qquad 1$$

$$z_j = \frac{1}{1 + \exp(-\theta_j)} \qquad\qquad 2$$

$$\theta_j = \sum_{i=0}^{N} w_{ji} x_i \qquad\qquad 3$$

The Least Mean Square error

$$E = \frac{1}{2} \sum_{k=1}^{M} (y_k - t_k)^2 \qquad\qquad 4$$

All the data are scaled between 0..1 but it can be scaled between -1..1 to standardise all the inputs with various dimensions. The weights w_{ji} and u_{kj} are assigned random numbers in the range -1..1, and a random pair of input / output vectors are picked from the training set. The input vector is fed through the network to get an output vector (feed-forward process); this is then compared with the output vector and an error is found.

This error is then passed back through the neural network (back propagation process) to modify the weights using the following equations

$$u_{kj}^{new} = u_{kj}^{old} + \Delta u_{kj} \qquad\qquad 5$$

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji} \qquad\qquad 6$$

The gradient descent optimization technique is used to calculate the change in each weight. This is then repeated by picking another random pair of input / output vectors and continuing until the error is at a minimum. This was done 1.5 million times which was sufficient for the network to reach a suitable minimum.
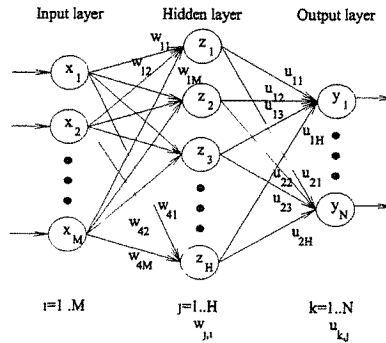
Figure 3. Multi-layered back propagation neural network

Momentum can be used to decrease times in training and the chance of the network getting stuck in a shallow minimum. This is done by accelerating the convergence of the error but is not applied in this situation.

The architecture of an OLL network [25], shown in Figure 4. It consists of an input layer, one or more hidden layers and an output layer. All input nodes are connected to all hidden nodes through weighted connections, W_{ji}, and all hidden nodes are connected to all output nodes through weighted connections, V_{kj}.
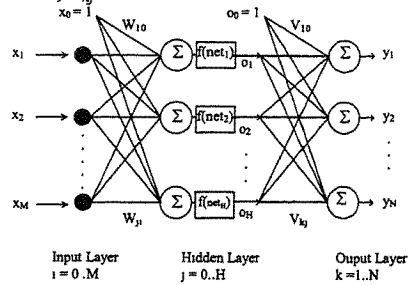
Figure 4. Basic structure of OLL with one hidden layer

The basic ideas of OLL learning algorithm are that the weights in each layer are modified dependent on each other, but separately from all other layers and the optimization of the hidden layer is reduced to a linear problem [25]. The algorithm of one hidden layer is summarised as following:

Training Algorithm of OLL with one hidden layer in Figure 4.

Step 1 Initialize weights W_{ji}, V_{kj} (Fig. 4) : set all weights to small random value range(-1,1)
    Set weight factor $\mu = 0.0001$, set Bias value $= 1$ to all hidden and output nodes.

Step 2 Calculate weights $V_{jk} = A^{-1}.b$ (Optimization of output-hidden layer weights)
    where the A and b matrix are given by:
    $A_{(j,j1)} = $ matrix $\{ a_{jj1} \}$; $a_{jj1} = \sum^{p}[o_j o_{j1}]$: $j,j1 = 0..H$
    $b_{(j,k)} = $ matrix $\{ b_{jk} \}$; $b_{jk} = \sum^{p}[t_k o_j]$: $k = 1..N$
    where t_k = target output for node k
        p = Number of iterations (1..Max of iterations)
        o_j = the output value of hidden layer node j

Step 3 Calculate Root Mean Square error (RMS)
    $RMS = \sum^{p}\sum^{k}1/2(t_k - y_k)^2$
    where y_k = calculated output for node k

Step 4.1 Calculate linearized weights in each output layer node :
{ Optimization of the input-hidden layer weights, step 4 to 6 }
    For $k=1$ to the last node at output layer (1..N)
        $Vlin_{kj} = \sum^{j} [f'(net_j) V_{kj}]$
    where $f'(net)$ = derivative of the sigmoidal function ($f(net)(1-f(net))$)

Step 4.2 Calculate weight correction term ($\Delta Wopt$) :
    $\Delta Wopt = Au^{-1}.bu$
    $Au_{(s,s)} = $ matrix $\{ a_{(ij,hm)} \}$; SxS dimensions square matrix
    $a_{(ji,hm)}$: for $(j \neq h) = \sum^{p} \sum^{k}[(Vlin_{kj}x_i)(Vlin_{kh}x_m)]$
        : for $(j = h) = \sum^{p} \sum^{k}(Vlin_{kj}x_i)(Vlin_{kh}x_m) + \mu/H^* abs(V_{kj})f''(net_j) x_i x_m$
    $bu_{(s)}$ = vector $\{ b_{(ji)} \}$; S dimension vector
        $b_{ji} = \sum^{p} \sum^{k}[(t_k - y_k) Vlin_{kj} x_i]$
    where $S = H \times (M+1)$ dimensions (Fig.4)

Step 5 Calculate Root Mean Square error (RMStest) by using Wtest

$Wtest_{ji} (new) = W_{ji} (old) + \Delta Wopt$

$RMStest = \sum^P \sum^k 1/2(t_k - y_k)^2$

Step 6 Compare between RMS and RMStest :

If (RMSTest > RMS) then

$\quad \mu \quad = \mu*1.2$ (increase μ) and Go back to Step 4.2

else $\quad W_{ji} \quad = Wtest_{ji}$

$\quad RMS \quad = RMStest$

$\quad \mu \quad = \mu*0.9$ (decrease μ)

Step 7 Do Step 2 to Step 6 until Test stop condition is True.

4 DEVELOPMENT OF TRAINING DATA RESULTS AND DISCUSSION

In order to train the network on a comprehensive range of cutting conditions and process variables, drilling experiments were carried out on the work material, S1214 free machining steel. An ANCA automatic drilling machine was used to carry out the experiments. The forces were measured using a three component dynamometer and associated data acquisition system. The hole diameter was measured by a probe with the highest resolution of 0.1 μm. Taking the handbook recommendations and associated feasible drill geometrical features a total of 72 experiments were carried out. The training of the network was carried out for the 57 cutting conditions above. All the input variables were scaled between 0-1 and the training was carried out over 57 combinations of cutting conditions.

The training was found to yield excellent accuracy with a small error at training stage indicating that the network was well trained with only 10 inputs and met the target unbalanced radial force and hole oversize accurately. The ten inputs were P1, P2, dP, ψ, δ_o and D and W/R, together with cutting conditions eR and eθ and f. It can be seen that the BP and OLL neural networks were trained well with great quantitative accuracy highlighting the predictive capability of the networks.

The error was calculated using the deviation formula (Predicted-Exp./Exp.)*100 and the percentage deviations at the training stage were excellent as shown in Table 1 and the line graph of training stage as shown in Fig 5.a, Fig5.b for the unbalanced radial forces (URF) and hole oversize (HO), respectively. It can be seen from Table 1 that at the training stage for both URF and HO there is no significant bias either for over prediction or under prediction for both BP and OLL neural network models. The neural network architecture was tested over 15 various conditions. The multi-layer perceptron with back propagation program was run to check the predictability of the neural network model for the testing stage.

Table 1. Average percentage deviations of the conventional method, BP and OLL

Method	% Deviation	
	Unbalanced Radial Force Training data	Hole oversize Training data
Conv.	-38.88	-61 88
BP	0.18	1.04
OLL	-0.19	-1.30
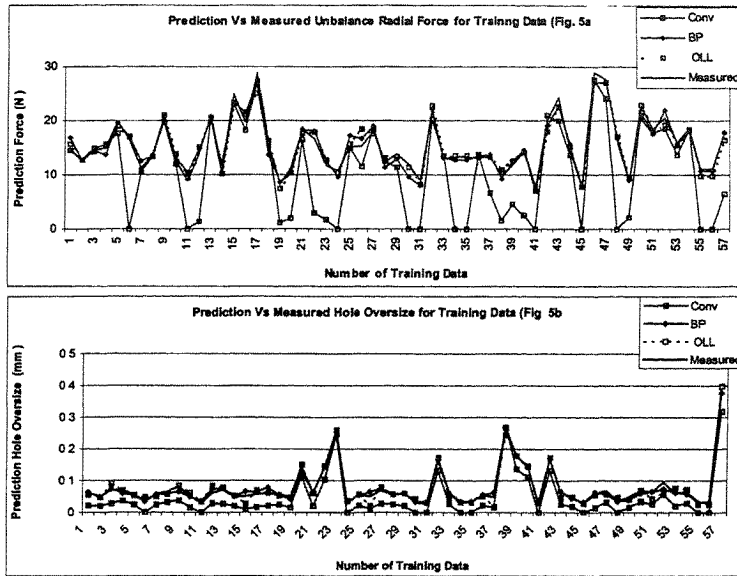
Figures 5a,5b. Predictions of the conventional method, BP and OLL at training stage
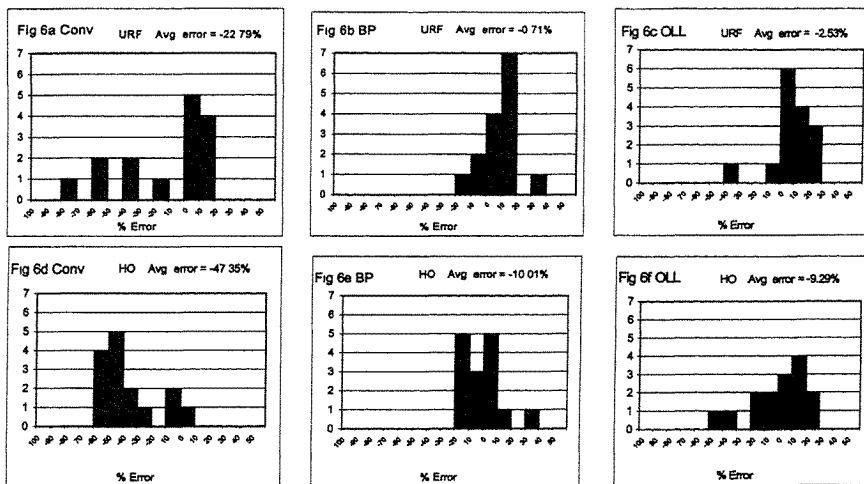
Figures 6a-f. Histograms of %error of the conventional method, BP and OLL
at testing stage

From a testing point of view, the histograms in Figs.6a-f highlight the comparison of the quantitative accuracy of three different approaches to unbalanced radial force (URF) and hole oversize (HO) prediction. The average percentage deviations of predicted URF and HO by conventional methods are -22.79% and -47.35% respectively (Figs.6a and 6d). For neural network based predictive models, BP had average percentage deviations of -0.71 and -10.01% (Figs.6b and 6e) and OLL had average percentage deviations of -2.53% and -9.29% (Figs.6c and 6f) for URF and HO, respectively. Comparing to conventional method, neural network approaches not only are able to simultaneously predict unbalanced radial force and hole oversize, but also yield higher quantitative accuracy.

5 CONCLUSION

The importance of reliable quantitative estimation of drilling performance is discussed. The unbalanced radial force contributing to the hole oversize and the mechanics of cutting approach to measure those two parameters is shown. It has been shown that the traditional approach to thrust and torque prediction using elemental force estimation and subsequent integration for total force estimation can be extended to the unbalanced radial force. This fundamental mechanics of cutting approach to radial force prediction in drilling operation is shown to be complex with numerous process variables involved. Two multi-layer feed forward neural networks have been chosen as proposed architectures. One of them is the established back propagation neural network model and the other is an optimized layer by layer feed forward network. Data gathered covered a range of cutting conditions for radial force measurements in drilling operation. A range of drilling conditions covering 72 different cutting conditions and tool geometrical features were selected as a training set. It has been shown that neural networks have superior predictive capability. The percentage deviations of unbalanced radial force and hole oversize predictions were -0.71 and -10.01% using the back propagation (BP) and -2.5% and -9.2% using optimized layer by layer neural network (OLL) compared to -22.8% and -47.3% using the traditional mechanics of cutting approach. These quantitative values are obtained using statistical routines such as average percentage deviations, cumulative frequency analysis and histograms.

REFERENCES

1. **Huang, S. H.** and **Zhang, H. C.**, "Artificial Neural Networks in Manufacturing: Concepts, Applications and Perspectives", IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, vol. 17, no. 2, 1994, pp. 212-228
2. **SOCIETY OF MANUFACTURING ENGINEERS**, Tool and Manufacturing Engineers Handbook, 3rd Edition, McGraw-Hill, New York, 1976.
3. **Armarego, E. J. A.** and **Brown, R. H.**, The Machining of Metals, Prentice-Hall, 1969.
4. **AMERICAN STANDARD**, U.S.A.S, B94.11-1967.
5. Chinese National Standards of Measuring and Cutting Tools, Chinese Standard Publishing House, 1990.
6. **Galloway, D. F.**, 'Some Experiments on the Influence of Various Factors on Drill Performance', Trans. A.S.M.E.., Vol. 79, 1957, p 191.
7. **Wright, J. D.**, 'A Study of the Geometrical Variability of Manufactured Twist Drills', M. Eng. Sc. Thesis, University of Melbourne, 1975.

8. **Micheletti, G. F. and Levi, R.** 'The Effect of Several Parameters on Twist Drill Performance', Proc of 8th Int. M.T.D.R. Conf., University of Manchester, Sept., 1967.

9. **Pramanik, D. K.** 'Variables Affecting Drill Performance', PhD Thesis, University of Melbourne, 1988.

10. **GALLOWAY, D.F. and MORTON, I. S.** 'Practical Drilling Tests', Research Dept. The Inst. of Prod. Eng (U.K.), 1946.

11. **Armarego, E. J. A. and Whitfield, R. C.** 'Computer Based Modelling of Popular Machining Operations for Force Prediction', Annals CIRP, Vol. 34, 1985, p 65.

12. **Whitfield, R. C.** 'Force Prediction in Machining', PhD Thesis, University of Melbourne, 1986.

13. **Karri, V.** 'Fundamental Studies of Rotary Tool Cutting Process', PhD Thesis, University of Melbourne, 1991.

14. **Zhao, H.** "Predictive Models for Forces, Power and Hole Oversize in Drilling Operations", PhD Thesis, University of Melbourne, 1994.

15. **Moore, T.** "Neural Networks to Estimate Performance in Drilling", M.Eng. Thesis, University of Tasmania, 2000.

16. **Boston O. W., and GILBERT, W. W.** 'The Torque and Thrust of Small Drills Operating in Various Metals', Trans. ASME., Vol. 58, N2, 1936, p 79.

17. **Pal, A. K., Bhattacharyya, A. and Sen, G. C.** 'Investigation of the Torque in drilling ductile materials, Int. J. Mach. Tool Des. Res., Vol. 4, 1965, p 205.

18. **Williams, R. A.,** 'A Study of the Mechanics of the Drilling Process', Harold Armstrong Conf. on Prod. Sci. Inst. of Engrs., Australia, Melbourne, 1971.

19. **Caudill, M. and Butler, C.** "Understanding Neural Networks -Computer Explorations", vol. 1: Basic Networks, Massachusetts Institute of Technology, 1992.

20. **Dayhoff, J. E.** "Neural Network Architectures - An Introduction", Van Nostrand Reinhold, 1990.

21. **Zurada, J. M.** "Introduction to Artificial Neural Systems", West Publishing Company, 1992.

22. **Yu. X., Loh, N. K., Jullien, G. A. and Miller, W. C.** "Comparisons of Four Learning Algorithms for Training the Multi-Layer Feed Forward Neural Networks with Hard Limiting Neurons", Neural Networks Theory, IEEE, New York, 1996.

23. **Masters, T.** "Practical Neural Network Recipes in C++", Academic Press Inc., 1993.

24. **D.E.Rumelhart. and J.L.Mcclelland.** "Parallel Distributed Processing :Explorations in the Microstructure of Cognition", MITPress, (1988).

25. **Ergezinger, S. And Thomsen, E.** "An accelerated learning algorithm for multilayer perceptrons: optimization layer by layer", IEEE transactions on neural networks, V.6, No.1, Jan.1995.

Intelligent process planning system for optimal CNC programming – a step towards complete automation of CNC programming

M K YEUNG
Integrated Manufacturing Technology Institute, National Research Council Canada, Ontario, Canada

ABSTRACT

One of the bottlenecks of CNC machining is the CNC programming. It relies on the experience and skills of the CNC programmer for the generation of the CNC program. The intelligent process planning system described in this paper generates a process plan automatically for CNC programming. It utilizes artificial intelligent technologies such as knowledge base, blackboard system and machine learning to extract machine-able features, proposes and selects optimal tools for the machining of the given part. Its flexibility and simplicity provides a convenient way to include new techniques and knowledge. The incorporation of this system with other CAD/CAM tools could effectively automate the CNC programming process.

Keywords: CNC programming, process planning, artificial intelligence, knowledge base, blackboard system, machine learning.

1 INTRODUCTION

CNC machining is one of the most important developments for manufacturing technologies in the 20th Century. It provides high efficiency for mass production of consumer products and flexibility for low quantity production of specialized parts and components. However, one of the bottlenecks of CNC machining is the CNC programming. It requires a skillful programmer who should not only be a CAD/CAM and computer literate but also a machining expert. While CNC tool paths can be generated by most CAD/CAM systems efficiently, planning for the machining process is tedious and almost completely relies on the expertise of the programmer. Decisions such as tools used for roughing and for finishing, number of passes and sequence of cuts are completely dependent on the knowledge of the programmer. Recent

publications have started to address this shortcoming of CNC programming and many research works have been conducted to find ways to reduce this deficiency but they mostly dealt with specific environments and conditions [1-9]. This paper proposes an intelligent process planning system based on a generic methodology and algorithm to automate and optimize the planning process for CNC machining.

The system utilizes the knowledge base, blackboard system and generic learning technologies. The knowledge base captures the experience of the programmer and facts relevant to the machine tools. The combination of the blackboard system, learning structure and algorithm defines the process plan and optimizes it for the operation. The implemented system would be very flexible. It has the capability to store and utilize new and specialized knowledge and facts or rules. Optimization is achieved within the defined knowledge set. With the capability and flexibility to expand and learn, the system could fully automate and optimize the CNC planning process within its given machining domain.

2 THE SYSTEM

The system divides the CNC planning process into three modules. The first module "Feature Extraction" is a knowledge base [10] that captures and stores part features, common machining techniques and logics for the given CNC machine. The second module "Tool Competition" is a blackboard system [11] that considers each cutting tool preset on the CNC machine as a 'competitor'. It competes the 'job' to machine the part that is posted onto the blackboard. The third module "Tool Optimization" is a learning system [12] that evaluates and selects the optimal cutting tools from the successful competitors to machine the part. An optimal cutting tool is selected for each pass of the tool path and therefore the tool path is optimized.

2.1 The Feature Extraction module

The core of this module is a knowledge base that consisted of a series of parallel nodes. Each of these nodes represents a family or class of machine-able parts (or sections of a part). Rules, facts, topological machining logics and constraints for the machining of the part are stored in a tree under the node. The leaf of the tree is the feature to be machined. To plan a machining job, the part to be machined is divided into features according to their machining criteria and procedures. Subsequently, the features are extracted in a topological order for the actual machining. The granularity of feature or part division can be adjusted for specific preference and application with appropriate design and implementation. The knowledge base S is a set of tree structures that represents different classes of parts.

$$S = \left\{ x_1, x_2, x_3, \cdots, x_n \right\};$$

Where $x_i \cap x_j = \phi$; $i, j = 0, 1, 2, \cdots, n$; $i \neq j$ and x_i is the root of a tree such that

$$x_i = \left(y_1, y_2, \cdots, y_k \right);$$

Where y_i is a sub-tree that represents a sub-assembly of features or a leaf that represents a feature to be machined, k is the number of sub-assemblies or features.

The output F of this module is a set of ordered pairs of machine-able feature z_i and updated blank configuration b_i for the part to be machined.

$$F = \left\{ (z_1, b_1), (z_2, b_2), (z_3, b_3), \cdots, (z_m, b_m) \right\};$$
$$and \; z_i \cap z_j = \phi \;; i \neq j \;; b_p \in b_q \;; b_q \notin b_p \;; p < q \;; i, j, p, q = 0, 1, 2, \cdots, m;$$

The operation starts by setting z_0 to null, b_0 to the blank and p_0 to the part. Then

$$z_0 = \phi \;;$$
$$b_0 = blank\_configuration \;; \quad size, material, shape, etc.$$
$$p_0 = \left\{ f_1, f_2, \cdots, f_m \right\}; \quad the\ part,\ represented\ by\ features$$
$$z_{i+1} = g(S, p_i);$$
$$b_{i+1} = b_i + z_i;$$
$$p_{i+1} = p_i - z_i;$$
$$and \; f_i \subseteq z_j \;; i, j = 0, 1, 2, \cdots, m \;;$$

Where g is a function (the inference engine) that extracts the machine-able feature f from the part P according to the machining topology and knowledge stored in S and outputs it along with the specific rules (if there are any) to z. It can be as simple as

$$g(S, p_i) = extract\_leaf(x_i \leftarrow p_i \; or \; f_j \in S);$$

The procedure re-iterates until all the features are extracted.

2.2 The Tool Competition module
The main element of this module is a blackboard or bulletin board structure where the feature and requirements extracted from the Feature Extraction module are posted. Each of the cutting tools preset on the machine is represented by a 'vendor' (a bidder to the job listed on the bulletin-board) or tool-node that will 'compete' to machine the feature based on its capabilities and constraints. Each tool-node is equipped with a knowledge base that stores these capabilities and constraints. The structure of the knowledge base is similar to the one in the Feature Extraction module. The capabilities are represented by a set of parallel nodes and their constraints are stored in tree structures under the nodes. A tool T is represented by a set of capabilities C.

$$T \supset C;$$
$$C = \left\{ c_1, c_2, c_3, \cdots, c_k \right\};$$

Where $c_i \cap c_j = \phi \;; i, j = 0, 1, 2, \cdots, k; i \neq j$ and c_i is the root of a tree such that

$$c_i = (r_1, r_2, \cdots, r_h);$$

Where r_i is a sub-tree that represents an intermediate constraint or a leaf. The leaf could be restrictions, limits of the functional/operational range, quantified finishing qualities, etc.

To machine a feature F being posted in the bulletin board, tool T examines its capabilities and decides whether to compete the machining work or not. The output Q of this module would be a set of tools that can machine F.

The algorithm is relatively simple for this process.

$$F = (f, b);$$
$$Q = \phi;$$
$$t_i (f, b) = \begin{cases} T_i; & \text{if } c_i \rightarrow f \text{ ; i.e. } T_i \text{ is to compete} \\ \phi; & \text{Otherwise ; i.e. } T_i \text{ is not to compete} \end{cases}$$
$$Q = Q + t_i (f, b);$$
$$\text{for } i = 1, 2, \cdots, k;$$

Where f is the feature to be machined, b is the current configuration of the blank, t_i is the inference engine of the ith tool that decides if T_i is capable to machine F and k is the number of tools preset for the machine.

2.3 The Tool Optimization module

The Tool Optimization module works in conjunction with the Tool Competition module to select optimal tools for the machining of the part. At this early design of the system, the environment for the evaluation is the simulation of machining path/pass of the selected tool and the goal is to achieve the fastest cycle time. A modified Stochastic Learning Automata [13] structure with generic learning algorithm is chosen to be the foundation of the implementation because of its flexibility and simplicity. Future design of the learning module will include multiple goals in various properties such as surface finishing, machining dynamics, tool characteristics, tool longevity, etc. The standard Stochastic Learning Automata structure is defined as the following quintuple.

$$SLA = \{ \alpha, \beta, p, T, c \};$$

Where α and β represent the sets of input and output states, p is the corresponding probability vectors, T is the learning algorithm and c is the corresponding penalty probability defining the environment. The modified SLA for this system is much simpler with only three elements.

$$SLA = \{ \beta, P, T \};$$

Where β is the output state, P is the probability vector for the cutting tools of the machine and T is the learning algorithm. A tool is represented by an element p_i of the probability vector P.

$$P = \{ p_1, p_2, \cdots, p_n \};$$

Where n = number of tools preset on the machine that is qualified for the job.

T evaluates the return from the simulation and updates P accordingly. After the ith iteration, T updates P^i to P^{i+1} for the next simulation based on the β^i.

$$P^{i+1} = T\{ P^i, \beta^i \};$$

The expansion of T can be described in two reactions to counter β^i.

Favorable reaction is when the return from the ith simulation is positive and k^{th} tool was chosen:

$$p_j^{i+1} = p_j^i - a p_j^i ; \quad \forall j, j \neq k, 0 < a < 1$$

$$p_k^{i+1} = p_k^i + \sum_{j=1, j \neq k}^{n} a p_j^i ;$$

Unfavorable reaction is when the return from the ith simulation is negative and k^{th} tool was chosen:

$$p_j^{i+1} = p_j^i + \left\{ \frac{b}{(n-1)} - b p_j^i \right\}; \quad \forall j, j \neq k, 0 \leq b < 1$$

$$p_k^{i+1} = p_k^i - \sum_{j=1, j \neq k}^{n} \left\{ \frac{b}{(n-1)} - b p_j^i \right\};$$

Where a is the reward parameter and b is the penalty parameter. They can be strategically set to offset the weights of the favorable or unfavorable return according to the desired scenario.

Initially, all qualified tools are set to have equal probabilities so that they would have an equal chance to be chosen to perform the next pass of the cutting path. Based on the result from the simulation, the tool returns the best performance (e.g. shortest cycle time) is chosen for the pass and its probability is asymptotically increased while probabilities for other tools are decreased. On the other hand, if the performance of the chosen tool is degraded from its previous pass, the opposite actions are applied. Note that the reward would be greater than the penalty if the penalty parameter were set to equal to the reward parameter. This gives the degraded tool other chances to prove itself until its probability is penalized to below the level of the others. This tool selection iterates until the path is completed.

3 MACHINING PARTS

To machine a part, the Feature Extraction module divides the part into a number of machine-able features and arranges them into an ordered set along with the corresponding blank

configurations according to the machining topology or operational preferences. For each of this ordered pair, the Tool Competition module recommends a set of capable tools to machine the feature and the Tool Optimization module, through simulation, evaluates and selects the optimal tools among the recommended tools. The process of tool recommendation and selection iterates until tools for all the features are selected. The outcome is an optimal plan for the CNC programming. Figure 1.0 shows the schematic of the operation of the system.
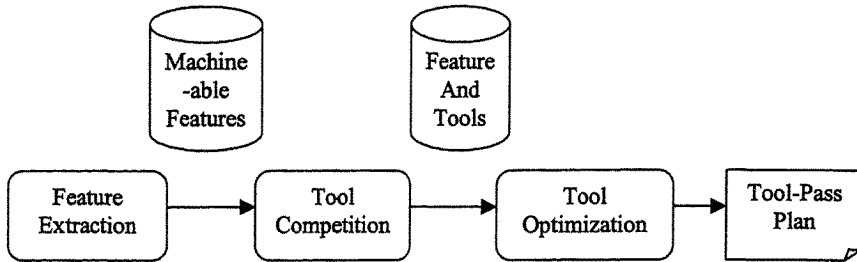
Figure 1.0 Operation of the Intelligent Process Planning System.

Following is an example to illustrate the operation of this intelligent planning process for CNC programming and to verify its feasibility.

The part: 2" $\emptyset$ x 2" long cylindrical sleeve with a 1½" $\emptyset$ through hole.

$$p_0 = \{ f_1 = bore, f_2 = o.d., f_3 = front\_face, f_4 = back\_face \};$$

The blank: 2¼" $\emptyset$ x 4" cold rolled steel bar with 1½" chucking length.

$$B = as\_described ;$$

The machine: 2-axis CNC lathe with 8-tool turret ATC; spindle speed and feed rate are negligent for this example.

The K-base:
$$S = \left\{ x_1 \left| \begin{pmatrix} tubular\_part; \\ y_1 \begin{pmatrix} rough\_cut; \\ z_1 = bore, \\ z_2 = o.d., \\ z_3 = front\_face \end{pmatrix}, y_2 \begin{pmatrix} finish\_cut; \\ z_4 = front\_face, \\ z_5 = bore, \\ z_6 = o.d., \\ z_7 = back\_face, \\ z_8 = parting \end{pmatrix} \end{pmatrix}, \cdots, x_n(\cdots) \right. \right\};$$

Preset tools: 1" $\emptyset$ drill with 4" drilling depth

$$T_1 = \left\{ c_1 \begin{pmatrix} drill\_hole; \\ r_1 = 1"\phi \ hole, \\ r_2 = \leq 4" \ deep \end{pmatrix} \right\};$$

½" $\emptyset$ x 4" finish-cut boring bar; optimal cutting depth 0.005" – 0.01".

$$T_2 = \left\{ c_1 \begin{pmatrix} machine\_through\_bore; \\ r_1 = requires \geq 0.625" \, \varnothing \, existing \, hole, \\ r_2 = 0.005"-0.01" \, optimal \, cutting \, depth, \\ r_3 = bore \leq 4" \, long, \\ r_4 = finish \, cut \, quality \, ; \, 0" \, under\_\phi\_size \end{pmatrix} \right\};$$

¾" $\varnothing$ x 4" rough-cut boring bar; optimal cutting depth 0.02" – 0.05".

$$T_3 = \left\{ c_1 \begin{pmatrix} machine\_through\_bore; \\ r_1 = requires \geq 0.875" \, \varnothing \, existing \, hole, \\ r_2 = 0.02"-0.05" \, optimal \, cutting \, depth, \\ r_3 = bore \leq 4" \, long, \\ r_4 = rough \, cut \, quality \, ; \, \cong 0.01" \, under\_\phi\_size \end{pmatrix} \right\};$$

Left side rough-cut tool; optimal cutting depth 0.02" – 0.1".

$$T_4 = \left\{ c_1 \begin{pmatrix} machine\_O.D.; \\ r_1 = 0.02"-0.1" \, optimal \, cutting \, depth, \\ r_2 = rough \, cut \, qaulity \, ; \, \cong 0.01" \, over\_\phi\_size \end{pmatrix} \right\};$$

Left side finish-cut tool; optimal cutting depth 0.005" – 0.02".

$$T_5 = \left\{ c_1 \begin{pmatrix} machine\_O.D.; \\ r_1 = 0.005"-0.02" \, optimal \, cutting \, depth, \\ r_2 = finish \, cut \, qaulity \, ; \, 0" \, over\_\phi\_size \end{pmatrix} \right\};$$

¼" x ¾" parting tool.

$$T_6 = \left\{ c_1 \begin{pmatrix} parting\_or\_grooving\_and\_facing\_both\_sides; \\ r_1 = \text{¼" } cutting \, width, \\ r_2 = \leq \text{¾" } cutting \, deep \end{pmatrix} \right\};$$

With this setup, the Feature Extraction module returns the set

$$F = \left\{ (z_1, b_1), (z_2, b_2), (z_3, b_3), (z_4, b_4), (z_5, b_5), (z_6, b_6), (z_{7,8}, b_7) \right\};$$

with $b_{i+1} = b_i + z_{i-1}$; $i = 1, 2, \cdots, 8$; $z_0 = \phi$; $b_1 = b_0$;

For each of the (zi,bi), the Tool Competition and Tool Optimization modules work together to define the optimal set of tool-passes. Each tool-pass is represented by the ordered pair (Tooli, Passi) or (Ti,Pi). The complete set of tool-passes generated by the system for the machining of the cylindrical sleeve is

$$CNC\_Plan = \left\{ \begin{matrix} (T_1, P_1), (T_6, P_2), (T_3, P_{3,4,5,6}), (T_3, P_7), (T_4, P_8), (T_4, P_9), (T_6, P_{10}), (T_2, P_{11}), \\ (T_5, P_{12}), (T_6, P_{13}) \end{matrix} \right\};$$

Where P_1 = drilling;
           P_2 = roughing front face;
           P_{3-6} = rough-boring at 0.05" cutting depth each;
           P_7 = rough-boring at 0.045" cutting depth;
           P_8 = roughing O.D. at 0.1" cutting depth;
           P_9 = roughing O.D. at 0.02" cutting depth;
           P_{10} = finishing front face;
           P_{11} = finishing bore;
           P_{12} = finishing O.D.;
           P_{13} = parting and finishing back face;

The plan can be incorporated into CAD/CAM systems for the CNC programming of the actual machining.

4 CONCLUSION

An automated process planning system can reduce the bottleneck burden of CNC machining. The system presented utilizes knowledge base, blackboard system and machine learning technologies to automatically plan and optimize the CNC programming process. Its flexibility and simplicity allow the inclusion of meta-knowledge and operation specific preferences, as well as dynamic updating. Users can adjust the granularity of the knowledge and information input into the system to satisfy their needs and preferences. It can also be incorporated or integrated into CAD/CAM systems for direct CNC programming. Future expansion of this system will include the development of new machining and automation technologies and relevant knowledge such as high-speed-machining, motion dynamics, geometric error compensation, characteristics of machines, tools and materials, etc. The intelligent process planning system will accommodate and facilitate the implementation and operation of these technologies and processes. It is a true fully automatic planning system for CNC programming and automation-controls within its knowledge domain. It can evolve and grow dynamically along with the acquisition of new knowledge and techniques.

REFERENCES

[1] Chin Sheng Chen, "Developing a feature based knowledge system for CAD/CAM integration", Computers & Industrial Engineering. vol.15, pp.34-40, 1988.
[2] Mantyla M., "Feature-based product modeling for process planning", Organization of Engineering Knowledge for Product Modeling in Computer Integrated Manufacturing, 2nd Toyota Conference. Elsevier, Amsterdam, Netherlands, xii+461, pp.303-324. 1989.
[3] Martin J.M., "A strategy for NC programming", Manufacturing-Engineering. vol.102, no.2, pp.82-84. Feb. 1989.

[4] Pande S.S. and Prabhu B.S., "An expert system for automatic extraction of machining features and tooling selection for Automats", Computer Aided Engineering Journal, vol.7, no.4, pp.99-103, Aug. 1990.

[5] Budde W. and Imbusch K., "EXAPT process planning and NC planning with database-supported management of production data", IFIP Transactions B (Applications in Technology), vol.B-3, pp.119-30, 1992.

[6] Norton N., "Controlling the Shop Floor", Manufacturing-Engineer, vol. 72, no. 6, pp. 272-275. December 1993.

[7] Yeo S.H., "A multipass optimization strategy for CNC lathe operations", International Journal of Production Economics, vol.40, no.2-3, pp.209-218, Aug. 1995.

[8] Fuh J.Y.H., Ji P. and Zhang Y.F., "Future development trends in CAM/CAPP-NC systems", International Journal of Computer Applications in Technology, vol.8, no.3-4, pp.203-210, 1995.

[9] Christensen G.K. and Mogensen O.B., "Backward form feature recognition and removal for an automatic CNC-programming system-BCAM", Advanced CAD/CAM Systems – State-of-the-art and Future Trends in Feature Technology. Chapman & Hall, London, U.K., pp. 205-216, 1995.

[10] Waterman Donald A., "A Guide to Expert Systems", Addision-Wesley Publishing Company, U.S.A., pp.392, 1986.

[11] Waterman Donald A., "A Guide to Expert Systems", Addision-Wesley Publishing Company, U.S.A., pp.388-389, 1986.

[12] Tanimoto Steven L. "The Elements of Artificial Intelligence", Computer Science Press, W.H. Freeman and Company, New York, U.S.A., pp.311-313, 1990.

[13] Narendra K.S. and Thathatchar M.A.L., "Learning Automata – An Introduction", Prentice-Hall International Inc., 1989.

A study and mathematic proof on the tool feed direction for each tool motion with the maximum efficiency in three-axis sculptured surface machining

Z C CHEN
Department of Mechanical and Industrial Engineering, Concordia University, Quebec, Canada
Z DONG and **G W VICKERS**
Department of Mechanical Engineering, University of Victoria, Canada

SYNOPSIS

To generate 3-axis CNC tool paths for sculptured part machining, surface quality and machining efficiency that are determined by the sculptured surface geometry, the cutter, and the tool path interval and direction are the two major concerns. In this work, the relationship between the machining efficiency of a tool motion and the tool feed direction is studied. A machining efficiency measure, the length of the effective cutting edge (ECE), is introduced in 3-axis CNC milling. The ECE length is mathematically proved to be the maximum when the cutter moves along the steepest tangent direction of the sculptured surface from a cutter contact point. This steepest tangent direction thus brings the tool motion the maximum machining efficiency. This study takes a new perspective on how to improve machining efficiency by controlling the tool feed direction instead of adjusting the tool path interval.
Keywords: tool feed direction, tool path generation, CNC programming, sculptured part machining, 3-axis CNC machining

1 INTRODUCTION

Sculptured parts have been widely used in aeronautical, automotive, as well as die and injection moulding industries. The complex shape of these parts imposes technical challenges on their CNC tool path generation, especially on the optimal tool paths that have the maximum machining efficiency and produce adequate surface quality. Adequate surface quality is accomplished when the heights of all cusps on a finished surface stay at the required surface tolerance. The maximum machining efficiency is achieved if all the tool motions are the most efficient. Over the past years, considerable researches have been devoted to find solutions to the challenges [3, 4, 6]. However, a generic method of the optimal tool paths is still beyond our reach. One unresolved problem is if a sculptured part presents dramatic curvature change, some of the tool paths generated by the methods considering surface geometry and tool path interval are inefficient because many tool motions in these tool paths

remove only a small amount of excess material, significantly prolonging machining time [7]. This problem is due to the lack of full understanding on the relationship between tool path geometry (or tool feed direction) and machining efficiency.

Earlier methods generating CNC tool paths automatically were based on simple geometric constraints. The iso-parametric tool path method retrieved from the sculptured surface iso-parametric curves as tool paths [1, 2]. This method is straightforward and easy to implement, but is unable to control the tool path interval across the surface. Recently, research has focused on the surface and cutter geometry to better control surface quality and machining efficiency. Some methods adjust the tool path interval according to the local surface shape to reduce redundant machining. The representative iso-cusped tool path scheme generates tool paths using varying tool path intervals to produce equal-height cusps on the finished surface [4, 5, 6]. Because all cusp heights are equal across the surface, redundant machining is reduced dramatically and machining time is saved considerably. The steepest-directed tree method [7] casts a mesh onto the surface and identifies shape feature at each mesh node. Then the tool path extends from a valley node to a ridge node along a steeper direction. In this method, the CNC machine mills one simple surface shape at a time. This strategy benefits machining efficiency.

These recent developments have advanced the automated tool path generation techniques for 3-axis CNC milling. However, the previous methods cannot ensure high machining efficiency for the entire surface. For instance, when the directions of some iso-cusped tool paths stay away from the tool axis, the tool is able to cut a less amount of material. The iso-cusped tool path generation method has no assurance on favourable tool motion directions, thus causing inefficient tool paths. The mesh density in the steepest-directed tree method is determined subjectively, so adequate surface quality is hard to obtain. In both cases, reduction of machining efficiency may happen.

In this work, the relationship between the machining efficiency of a tool motion and its tool feed direction is studied. A new machining efficiency measure, the length of the effective cutting edge (ECE), is introduced in 3-axis CNC milling. The ECE length is mathematically proved to be the maximum when the cutter moves along the steepest tangent direction of the sculptured surface from a cutter contact (CC) point. So the machining efficiency of the corresponding tool motion reaches the maximum. The study takes a new perspective on how to improve the machining efficiency by changing the tool feed direction instead of adjusting the tool path interval, and lays a basis for a new type of tool path, the steepest-directed tool path. A machining example of a hemi-cylindrical part is presented to verify the new finding.

2 MACHINING EFFICIENCY MEASURE

In order to recognize the key factors influencing the tool-motion machining efficiency, the mechanism of 3-axis CNC machining is examined, and a new machining efficiency measure is defined.

2.1 Generic Machining Model
A sculptured surface machined on a 3-axis vertical CNC milling machine with a torus end-mill is shown in Fig 1 The machine table moves along X- and Y-axes and the cutter moves along Z-axis. The tolerance surface is an offset of the sculptured surface by the specified surface tolerance. An efficient tool path consists of many efficient tool motions (or tool path

segments). In an efficient tool motion the cutter moves from one CC point to another along a favourable tool feed direction and removes a larger amount of material between the tolerance surface and the sculptured surface. To quantify the machining efficiency of a tool motion and to identify the tool feed direction bringing the tool motion the maximum machining efficiency, a generic, geometric model on cutter-surface interaction is introduced for a close examination on tool motion.
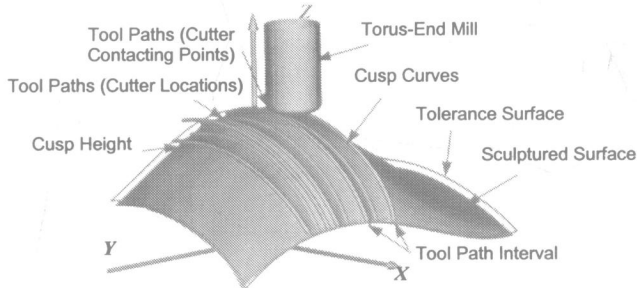
Fig. 1 Sculptured Surface with Adequate Surface Quality

In the generic machining model, as illustrated in Fig. 2, the machining surface shape of a torus end-mill like a fry pan is contacting with a sculptured surface at a cutter contact (CC) point. When the cutter moves from a lower CC point to the next upper CC point along the tool feed direction, an envelope of the torus cutting surface is formed in this tool motion. A planar cutting edge is defined as the profile of the cutter that is projected along the tool feed direction onto an orthogonal projection plane. After aligning the projection plane with the paper in such a way that the tool feed direction points into the plane, the cutting edge, the sculptured surface and the tolerance surface are illustrated in Fig. 3. The envelope of the torus cutting surface is formed by sliding the planar cutting edge along the tool feed direction. In machining, all material that is above the planar cutting edge is removed. *The portion of this cutting edge between the sculptured and tolerance surfaces actually cuts the last layer of excess material, shapes the surface and determines the surface finish.* It is thus called the effective cutting edge (ECE). Fig. 3 indicates that the longer the ECE length, the more the excess material is removed. Since the ECE length can represent the amount of excess material removed in a tool motion, this ECE length is a valid measure on machining efficiency.

2.2 Four Coordinate Frames

To derive the mathematic formulae of the planar cutting edge and the ECE length, four coordinate frames (or systems) are introduced. These include the cutter geometry frame (CGF), the cutter location frame (CLF), the steepest direction frame (SDF), and the cutter motion frame (CMF). Based on the geometric relations between each pair of the successive frames, the expressions of the cutter and the cutting surface normal in the CGF are transformed sequentially [8, 9] through the CLF and the SDF, into the CMF, in which the cutter projection is easily obtained.
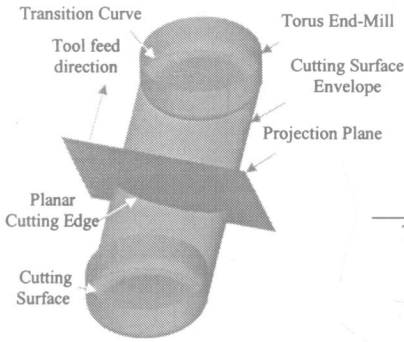
Fig. 2 Generic Machining Model of Sculptured Parts
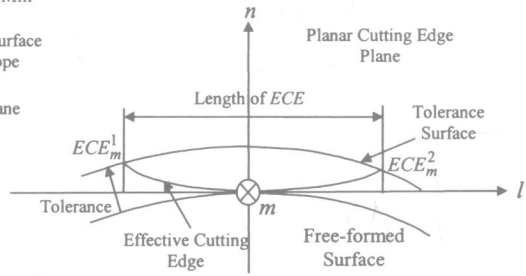
Fig. 3 Effective Cutting Edge

2.2.1 Cutter Geometry Frame (CGF)

For a sculptured surface, S, to be machined on a 3-axis vertical CNC machine with a torus end-mill, the surface, a cutter, and a CC point P_0, is shown in Fig. 4. In the part design coordinate system, the cutter axis is along the direction, $[0,\ 0,\ 1]^T$. The cutter geometry frame (i-j-k) is introduced with its origin located at the cutter tip (centre); its k-axis is aligned with the cutter axis; and its i-axis is on the plane defined by the k-axis and the surface normal n at the CC point, directing off the surface. The expression of the torus cutting surface in the CGF then becomes

$$\begin{bmatrix} i \\ j \\ k \end{bmatrix} = \begin{bmatrix} (R_1 + R_2 \cdot \sin\varphi) \cdot \cos\theta \\ (R_1 + R_2 \cdot \sin\varphi) \cdot \sin\theta \\ R_2 \cdot (1 - \cos\varphi) \end{bmatrix} \tag{1}$$

where $R_1 > 0$, $R_2 > 0$ and $R_1 \square R_2$; $\varphi \in [0,\ \frac{\pi}{2}]$ and $\theta \in [0,\ 2\pi]$. The geometric parameters of the torus end-mill are given in Fig. 5.

In the CGF, the cylindrical cutting surface is represented as

$$\begin{bmatrix} i \\ j \\ k \end{bmatrix} = \begin{bmatrix} (R_1 + R_2) \cdot \cos\theta \\ (R_1 + R_2) \cdot \sin\theta \\ R_2 + h \end{bmatrix} \tag{2}$$

At point P_0, parameter θ equals to π. The point location is also specified by parameter φ, represented as angle φ_0. The CC point coordinates in the CGF is then represented as $[-(R_1 + R_2 \sin\varphi_0),\ 0,\ R_2(1 - \cos\varphi_0)]^T$.
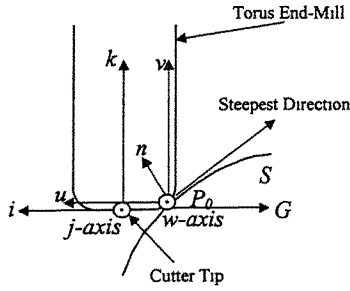
Fig. 4 Cutter Geometry Frame and Cutter Location Frame
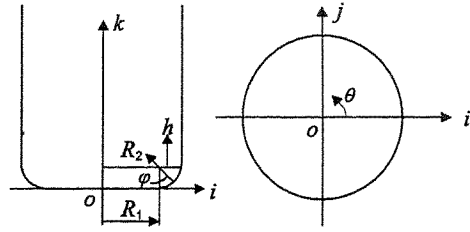
Fig. 5 Torus End-Mill

2.2.2 Cutter Location Frame (CLF)

The cutter location frame (u-w-v) is formed by translating the CGF from the cutter tip (center) to point, P_0. Axes i, j, and k in the CGF become axes u, w, and v in the CLF, as shown in Fig. 4. Here axes v and u are already given in the part design coordinate system. Axis v is vertical, $v = [0, 0, 1]^T$; and the direction of axis u is opposite to the surface gradient G at P_0, which is $[\partial z/\partial x, \ \partial z/\partial y, \ 0]^T$, as shown in Fig. 4 and discussed in [10]. Axis u is then represented as $u = [-\partial z/\partial x, \ -\partial z/\partial y, \ 0]^T$ in the part design coordinate system. Axis w is defined as the cross product of axes v and u:

$$w = v \times u = \left[\frac{\partial z}{\partial y}, \ -\frac{\partial z}{\partial x}, \ 0 \right]^T \tag{3}$$

The previously derived representations of the torus and cylindrical cutting surfaces in the CGF (Eqs. 1 and 2) can now be expressed in the CLF as given in Eqs. 4 and 5.

$$\begin{bmatrix} u \\ w \\ v \end{bmatrix} = \begin{bmatrix} (R_1 + R_2 \cdot \sin \varphi) \cdot \cos \theta + (R_1 + R_2 \cdot \sin \varphi_0) \\ (R_1 + R_2 \cdot \sin \varphi) \cdot \sin \theta \\ R_2 \cdot (\cos \varphi_0 - \cos \varphi) \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} u \\ w \\ v \end{bmatrix} = \begin{bmatrix} (R_1 + R_2) \cdot \cos \theta + (R_1 + R_2 \cdot \sin \varphi_0) \\ (R_1 + R_2) \cdot \sin \theta \\ R_2 \cdot \cos \varphi_0 + h \end{bmatrix} \tag{5}$$

2.2.3 Steepest Direction Frame (SDF)

The surface normal n is known as $[-\partial z/\partial x, -\partial z/\partial y, 1]^T$, and the steepest tangent direction SD of the surface at point P_0 is $\left[\partial z/\partial x, \ \partial z/\partial y, \ (\partial z/\partial x)^2 + (\partial z/\partial y)^2 \right]^T$ (refer [11]). Since the dot

products between any pair of axes $\mathbf{n}$, $\mathbf{SD}$, and $\mathbf{w}$ are null, the axes are perpendicular with each other. These axes are used to build another coordinate frame, the steepest direction frame (n-w-SD), as shown in Fig. 6. In this frame, axes n-, w-, and SD are represented as axes $\mathbf{n}$, $\mathbf{w}$, and $\mathbf{SD}$, respectively. At CC point $P_0(\pi, \varphi_0)$ the angle between axes $\mathbf{n}$ and $\mathbf{v}$ is angle φ_0. The CLF coincides with the SDF when it is rotated along axis w by an angle of $(\pi/2 - \varphi_0)$. The rotation will align axes u and v with axes n and SD, respectively. Any points on the cutting surface can be transformed from the CLF into the SDF by rotating around axis w [9].

$$\begin{bmatrix} n \\ w \\ SD \end{bmatrix} = \begin{bmatrix} \sin\varphi_0 & 0 & \cos\varphi_0 \\ 0 & 1 & 0 \\ -\cos\varphi_0 & 0 & \sin\varphi_0 \end{bmatrix} \cdot \begin{bmatrix} u \\ w \\ v \end{bmatrix} \tag{6}$$

2.2.4 Cutter Motion Frame (CMF)

During machining, the cutter feeds in a direction tangent to the machined surface. The angle between this tool feed direction $\mathbf{m}$ and the steepest tangent direction $\mathbf{SD}$ is angle α, as shown in Fig. 6. Directions $\mathbf{m}$, $\mathbf{n}$, and $\mathbf{l}$, defined by the cross product of directions $\mathbf{m}$ and $\mathbf{n}$, form another coordinate frame, the cutter motion frame (CMF). The orientation of this frame changes to direction $\mathbf{m}$. The difference between the two frames, the SDF and the CMF, is angle α about axis n. With a geometric transformation, the expressions of the introduced torus and cylindrical cutting surfaces of the cutter can be represented in the new CMF, as given in Eqs. 7 and 8.

(a)   (b)
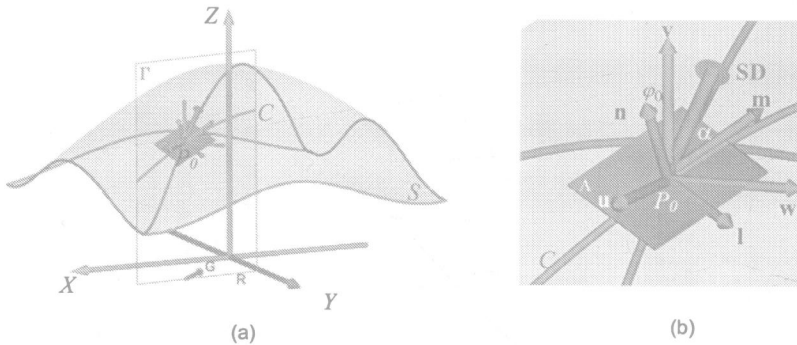
Fig. 6 (a) Steepest Tangent Direction and Cutter Motion Frame; (b) Zoom-in View

The torus cutting surface:

$$\begin{aligned} \begin{bmatrix} n \\ l \\ m \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} n \\ w \\ SD \end{bmatrix} \\ &= \begin{bmatrix} a \cdot R_1 \cdot \cos\theta + a \cdot R_2 \cdot \cos\theta \cdot \sin\varphi - b \cdot R_2 \cdot \cos\varphi + a \cdot R_1 + R_2 \\ R_1 \cdot \sin\theta \cdot \cos\alpha + R_2 \cdot \sin\theta \cdot \sin\varphi \cdot \cos\alpha + b \cdot R_1 \cdot \cos\theta \cdot \sin\alpha + b \cdot R_2 \cdot \cos\theta \cdot \sin\varphi \cdot \sin\alpha + a \cdot R_2 \cdot \cos\varphi \cdot \sin\alpha + b \cdot R_1 \cdot \sin\alpha \\ R_1 \cdot \sin\theta \cdot \sin\alpha + R_2 \cdot \sin\theta \cdot \sin\varphi \cdot \sin\alpha - b \cdot R_1 \cdot \cos\theta \cdot \cos\alpha - b \cdot R_2 \cdot \cos\theta \cdot \sin\varphi \cdot \cos\alpha - a \cdot R_2 \cdot \cos\varphi \cdot \cos\alpha - b \cdot R_1 \cdot \cos\alpha \end{bmatrix} \end{aligned} \tag{7}$$

The cylindrical cutting surface:

$$\begin{bmatrix} n \\ l \\ m \end{bmatrix} = \begin{bmatrix} a\ (R_1+R_2)\ \cos\theta + a\ R_1 + R_2 + b\ h \\ (R_1+R_2)\cdot\sin\theta\cdot\cos\alpha + b\cdot(R_1+R_2)\ \cos\theta\cdot\sin\alpha + b\ R_1\ \sin\alpha - a\ h\ \sin\alpha \\ (R_1+R_2)\cdot\sin\theta\cdot\sin\alpha - b\ (R_1+R_2)\cdot\cos\theta\ \cos\alpha - b\ R_1\cdot\cos\alpha + a\ h\ \cos\alpha \end{bmatrix} \quad (8)$$

where $a = \sin\varphi_0$ and $b = \cos\varphi_0$.

In the CMF, plane $(n\text{-}l)$ is the cutter projection plane that is perpendicular to the tool feed direction m. The geometric representation of the cutter in the CMF can be projected onto plane $n\text{-}l$ by replacing component m in Eqs. 7 and 8 with zero. The projection of the torus cutting surface becomes

$$\begin{bmatrix} n \\ l \\ m \end{bmatrix} = \begin{bmatrix} a\ R_1\ \cos\theta + a\ R_2\ \cos\theta\ \sin\varphi - b\ R_2\ \cos\varphi + a\ R_1 + R_2 \\ R_1\ \sin\theta\ \cos\alpha + R_2\ \sin\theta\ \sin\varphi\ \cos\alpha + b\ R_1\ \cos\theta\ \sin\alpha + b\ R_2\ \cos\theta\ \sin\varphi\ \sin\alpha + a\ R_2\ \cos\varphi\ \sin\alpha + b\ R_1\ \sin\alpha \\ 0 \end{bmatrix} \quad (9)$$

and the projection of the cylindrical cutting surface becomes

$$\begin{bmatrix} n \\ l \\ m \end{bmatrix} = \begin{bmatrix} a\cdot(R_1+R_2)\cdot\cos\theta + a\cdot R_1 + R_2 + b\cdot h \\ (R_1+R_2)\cdot\sin\theta\ \cos\alpha + b\cdot(R_1+R_2)\ \cos\theta\cdot\sin\alpha + b\cdot R_1\cdot\sin\alpha - a\cdot h\cdot\sin\alpha \\ 0 \end{bmatrix} \quad (10)$$

Similarly, the mathematical representations of the surface normal of the torus and cylindrical cutting surfaces can be transformed from the CGF to the CMF, using the same transformation. The unit normal of the torus cutting surface $\mathbf{n}_{torus}$ in the CGF is

$$\mathbf{n}_{torus} = \begin{bmatrix} n_i \\ n_j \\ n_k \end{bmatrix} = \begin{bmatrix} \cos\theta\cdot\sin\varphi \\ \sin\theta\cdot\sin\varphi \\ -\cos\varphi \end{bmatrix} \quad (11)$$

and the unit normal of the cylindrical cutting surface $\mathbf{n}_{cylinder}$ in the CGF is

$$\mathbf{n}_{cylinder} = \begin{bmatrix} n_i \\ n_j \\ n_k \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix} \quad (12)$$

After the coordinate transformations, the expression of the unit normal of torus cutting surface in the CMF becomes

$$\mathbf{n}_{torus} = \begin{bmatrix} n_n \\ n_l \\ n_m \end{bmatrix} = \begin{bmatrix} a\cdot\cos\theta\cdot\sin\varphi - b\cdot\cos\varphi \\ \sin\theta\cdot\sin\varphi\cdot\cos\alpha + b\cdot\cos\theta\cdot\sin\varphi\cdot\sin\alpha + a\cdot\cos\varphi\cdot\sin\alpha \\ \sin\theta\cdot\sin\varphi\cdot\sin\alpha - b\cdot\cos\theta\cdot\sin\varphi\cdot\cos\alpha - a\cdot\cos\varphi\cdot\cos\alpha \end{bmatrix} \quad (13)$$

Similarly, the unit normal of the cylindrical cutting surface in the CMF becomes

$$\mathbf{n}_{cylinder} = \begin{bmatrix} n_n \\ n_l \\ n_m \end{bmatrix} = \begin{bmatrix} a\cdot\cos\theta \\ \sin\theta\cdot\cos\alpha + b\ \cos\theta\cdot\sin\alpha \\ \sin\theta\cdot\sin\alpha - b\cdot\cos\theta\cdot\cos\alpha \end{bmatrix} \quad (14)$$

2.3 Effective Cutting Edge

2.3.1 Planar Cutting Edge

Introduction of the series of frames, ended with the CMF, allows the cutter to be projected along the tool feed direction onto the perpendicular plane n-l. The periphery of this projection is the planar cutting edge, as shown in Fig. 7(a). The original mathematical representations of cutter projection are given in Eqs. 9 and 10. Each point on the planar cutting edge corresponds to a point on the cutter at the CC point. These points are called transition points, and a characteristic of these points is that the cutting-surface normal at these points is perpendicular to the tool feed direction, i.e.

$$\mathbf{m}^{\mathrm{T}} \cdot \mathbf{n} = 0 \qquad (15)$$

where axis $\mathbf{m}$ is the tool feed direction, and axis $\mathbf{n}$ is the cutting-surface normal. The planar cutting edge can thus be obtained by solving Eqs. 9 and 15, or Eqs. 10 and 15.

(a)

(b)
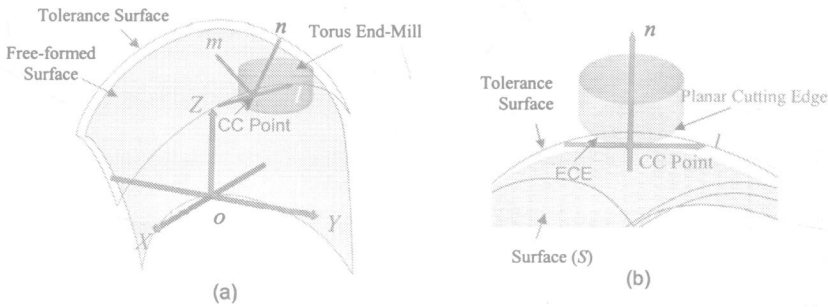
Fig. 7 Diagram of Effective Cutting Edge (a) a Cutter Milling a Surface, (b) Projection of the Cutter on Plane (l-n) along Tool Feed Direction (m)

2.3.2 Effective Cutting Edge (ECE)

If the local shape of the surface is assumed to be uniform around the CC point, the tolerance surface then intersects the planar cutting edge at two points, while component n approximately equals to the surface tolerance, as shown in Figs. 7(b) and 3. The ECE is the portion of the planar cutting edge within the part and tolerance surfaces and between the two intersecting points. The ECE can be obtained using Eqs. 9 (or 10), 15 and 16.

$$a \cdot R_1 \cdot \cos\theta + a \cdot R_2 \cdot \cos\theta \cdot \sin\varphi - b \cdot R_2 \cdot \cos\varphi + a \cdot R_1 + R_2 \le t \quad \text{or}$$
$$a \cdot (R_1 + R_2) \cdot \cos\theta + a \cdot R_1 + R_2 + b \cdot h \le t \qquad (16)$$

where t represents the surface tolerance.

Suppose the two ends of the ECE are $ECE_m^1(\theta_1, \varphi_1)$ and $ECE_m^2(\theta_2, \varphi_2)$, and their coordinates in the CMF are $ECE_m^1[n_1 \quad l_1 \quad m_1]^{\mathrm{T}}$ and $ECE_m^2[n_2 \quad l_2 \quad m_2]^{\mathrm{T}}$ (see Fig. 3). The ECE length, L_{ECE}, is measured by axis l coordinate difference between the two points, and calculated by

$$L_{ECE} = |l_1 - l_2| \qquad (17)$$

2.4 Relationship between Effective Cutting Edge and Tool Feed Direction

During machining, when the tool feed direction axis m changes, the profile of the cutter projection on plane n-l in the CMF and the ECE will change accordingly. The length of ECE is a function of the tool feed direction, which is represented by angle α between axes m and SD, and this length reaches its maximum when the cutter feeds along a certain direction. When the ECE length reaches its maximum, the tool removes the maximum amount of material in the tool motion. Identifying this most productive tool feed direction is critical to tool path planning.

3 MAXIMUM ECE LENGTHS FOR COMMON CUTTERS

In this section, generic ECE length formulae for the commonly used milling cutter, torus end-mill, is derived. The most efficient tool feed direction is identified by maximizing the formulated ECE length. The optimization results show the ECE length of a ball end-mill is independent of the tool feed direction, while the steepest tangent direction of the surface is the most efficient tool feed direction for torus and flat end-mills.

3.1 Maximum ECE Length for a Torus End-Mill

Computing the maximum ECE length for a torus-end mill is very complex. The cutting surface, the offset plane and their intersection are illustrated in Fig. 9. The cutter touches the tangent plane Λ of the surface (not shown) at a CC point P_0. Any Λ-offset plane intersecting the cutting surface generates a closed and convex curve as shown in Fig. 10. When the intersection curve is projected along axis m onto its orthogonal plane (n-l), two points on the intersection curve, whose tangents are parallel with axis m, map the ends of the intersection curve projection. These extremes of the projection are on the ECE. Fig. 9(b) presents a special example of the ECE when axis m is aligned with axis SD.

The ECE length can be derived with the help of intersection curves. A series of planes, parallel to plane Λ, intersect the cutting surface and form a group of closed curves. If these curves are projected onto plane (n-l) along axis m, each curve corresponds to two points on the ECE. In particular, the curve on a Λ-offset plane Γ by the value of the surface tolerance t specifies the two ends of the ECE. The difference of the l-component of these two ends defines the ECE length.

To find the maximum ECE length and the tool feed direction, two steps are needed. First the ECE length ($L_{\alpha=0}$) when axis m is aligned with axis SD is to be derived. Secondly, the ECE length is to be proved to be longer than any other ECE length (L_α) when axis m diverts from axis SD.

3.1.1 ECE length $L_{\alpha=0}$ of the steepest tangent direction

Suppose plane Γ intersects axis n at point P_1 as shown in Fig. 9(a). It also intersects with planes (n-SD) and (n-w) of the SDF at lines SD' and w', respectively. Line SD' is parallel to axis SD, and line w' is parallel to axis w. Similarly, plane Γ intersects planes (n-m) and (n-l) of the CMF at lines m' and l', respectively. Line m' is parallel to axis m, and line l' is parallel to axis l. Point P_1 is the origin for both planes SD'-w' and m'-l'. Fig. 10 illustrates the intersection curve in plane SD'-w' and its projection along line SD' onto axis w'. Two points on the curve, Q_{0L} and Q_{0R}, map the ECE ends, ECE_{SD}^1 and ECE_{SD}^2.
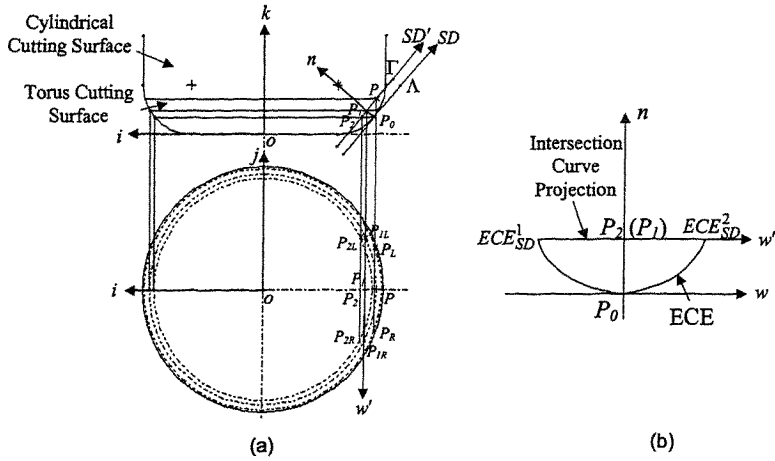
Fig. 9 (a) Intersection between the Cutting Surface and the Λ-offset plane Γ; (b) Zoom-in View of the ECE from Axis SD
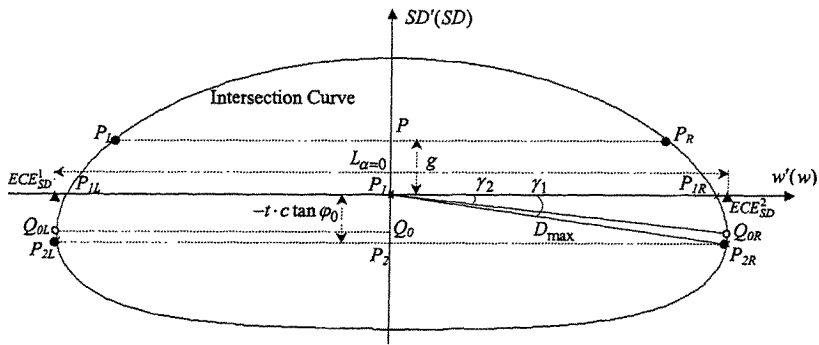
Fig. 10 View along Axis n on the Intersection Curve on Plane Γ

Eq. 6 is the representation of the cutter surface in the SDF. Plane Γ can be represented as $n = t$, and line SD is along $[0 \quad 0 \quad 1]^T$ in the SDF. After substituting these knowns into Eqs. 15 and 16, the two points on the curve, Q_{0L} and Q_{0R}, can be obtained using the following equations.

$$\begin{cases} -b \cdot \cos\theta \cdot \sin\varphi - a \cdot \cos\varphi = 0 \\ a \cdot R_1 \cdot \cos\theta + a \cdot R_2 \cdot \cos\theta \cdot \sin\varphi - b \cdot R_2 \cdot \cos\varphi + a \cdot R_1 + R_2 = t \end{cases} \tag{18}$$

However, finding a closed-form solution to these equations is stunningly complex. One can employ numerical methods to solve these equations. Geometric analysis is used to prove that solution exist for Eq. 27.

The torus cutting surface of the cutter in the CGF has been defined in Eq. 1, and the coordinate of point $P_0(\pi, \varphi_0)$ has been calculated previously in Section 2.2. In Fig. 9, point P_1 is located at $\left[-R_1 - (R_2 - t) \cdot \sin\varphi_0, \ 0, \ R_2 - (R_2 - t) \cdot \cos\varphi_0\right]^\mathrm{T}$ in the CGF, and axis SD' is represented in this frame as well.

$$\begin{bmatrix} i_{SD'} \\ j_{SD'} \\ k_{SD'} \end{bmatrix} = \begin{bmatrix} -R_1 - (R_2 - t) \cdot \sin\varphi_0 - g \cdot \cos\varphi_0 \\ 0 \\ R_2 - (R_2 - t) \cdot \cos\varphi_0 + g \cdot \sin\varphi_0 \end{bmatrix} \tag{19}$$

where line parameter g represents the distance between any point P on the axis and point P_1 (see Fig. 9). Through point P and perpendicular to plane (n-SD), line P_L-P-P_R intersects the curve at two points, P_L and P_R, (see Fig. 9 or 10). These two points are symmetric to axis SD'. When parameter g changes, points P_L and P_R move along the curve. In the SDF, the coordinates of points P_L and P_R can be derived. For instance, component w of point P_R's coordinates in the SDF is in the form of

$$\left(w_{P_R}\right)^2 = 2R_1 \cdot \left\{\sqrt{R_2^2 - \left[(R_2 - t) \cdot \cos\varphi_0 - g \cdot \sin\varphi_0\right]^2} - (R_2 - t) \cdot \sin\varphi_0 - g \cdot \cos\varphi_0\right\} + R_2^2 - (R_2 - t)^2 - g^2 \tag{20}$$

Distance ($D_{P_1 P_R}$) between any point P_R on the curve and origin P_1 can be calculated.

$$\left(D_{P_1 P_R}\right)^2 = 2R_1 \cdot \left\{\sqrt{R_2^2 - \left[(R_2 - t) \cdot \cos\varphi_0 - g \cdot \sin\varphi_0\right]^2} - (R_2 - t) \cdot \sin\varphi_0 - g \cdot \cos\varphi_0\right\} + R_2^2 - (R_2 - t)^2 \tag{21}$$

Distance ($D_{P_1 P_R}$) is a function of parameter g. The maximum distance ($D_{P_1 P_R}$)$_{\max}$ can be found by maximizing this distance with respect the variable g and the value of $g_{\max}$ is

$$g_{\max} = -t \cdot \mathrm{ctan}\varphi_0 \tag{22}$$

The calculated $g_{\max}$ refers to an across point P_2 between axis SD' and a horizontal line passing through point P_0, as shown in Fig. 9(a). Assume a line normal to plane (n-SD) intersects the cutting surface at two symmetric points, P_{2L} and P_{2R} (see Fig. 10). The distance ($D_{\max}$) between point P_{2L} or P_{2R} and point P_1 is the longest among all the distances between any other point on the curve and point P_1. The coordinate of point P_{2R} in the SDF ($n-w-SD$) is $\left[t, \ \frac{1}{\sin\varphi_0}\sqrt{2 \cdot R_1 \cdot t \cdot \sin\varphi_0 + 2 \cdot R_2 \cdot t \cdot \sin^2\varphi_0 - t^2}, \ -t \cdot \mathrm{ctan}\varphi_0\right]$, and the coordinate of point P_{2L} is similar to P_{2R} except a negative component w.

In Fig. 10, the angle between line $P_1 P_{2R}$ and line w' is angle γ_1. The surface normal of the cutter at point P_{2R} is calculated. Its component SD is $b \cdot \sin\varphi_0 - a \cdot \cos\varphi_0 - \dfrac{t}{R_1 + R_2 \cdot \sin\varphi_0}$. This implies that point P_{2R} is not the point on the curve, which maps the end of the ECE, since the surface normal at this point does not meet condition, $\mathbf{m}^\mathrm{T} \cdot \mathbf{n} = 0$. There must exist a point Q_{0R}

between points P_{1R} and P_{2R} and point Q_{0L} between points P_{1L} and P_{2L}, which map the ends of the ECE, ECE_{SD}^1 and ECE_{SD}^2.

Suppose the angle between line P_1Q_{0R} or P_1Q_{0L} and line w' is angle γ_2. When axis $\mathbf{m}$ is in line with axis $\mathbf{SD}$, the projections of the points Q_{0L} and Q_{0R} on axis w' defines the ends of the ECE. Hence, the ECE length $L_{\alpha=0}$ becomes

$$L_{\alpha=0} = Q_{0L}P_1 \cdot \cos\gamma_2 + Q_{0R}P_1 \cdot \cos\gamma_2 \qquad (23)$$

3.1.2 Maximum ECE length
When the tool feed direction $\mathbf{m}$ is not aligned with the steepest tangent direction $\mathbf{SD}$, and angle α between them is not zero, the intersection curve should be projected along direction $\mathbf{m}$ onto plane n-l of the CMF, rather than along direction $\mathbf{SD}$ onto plane n-w of the SDF. The angle between directions $\mathbf{m}$ and $\mathbf{SD}$ is the same as the angle between axes m' and SD', as is the angle between axes w' and l', as shown in Fig. 11. The two ends of the ECE can be identified by projecting the intersection curve along axes m' onto l' (see Fig. 11).

Assume two points Q_L and Q_R on the intersection curve are projected as the two ends of ECE, ECE_m^1 and ECE_m^2, the angle between lines P_1Q_R and P_1Q_{0R} is angle β_1, and the angle between lines P_1Q_L and P_1Q_{0L} is angle β_2, the ECE length L_α is thus formulated as:

$$L_\alpha = Q_R P_1 \cdot \cos(\gamma_2 - (\alpha - \beta_1)) + Q_L P_1 \cdot \cos(\gamma_2 + (\alpha - \beta_2)) \qquad (24)$$

The value of length L_α varies according to angle α. Since the curve is symmetric to axis SD', the ECE lengths remain the same if the two projection directions are also symmetric to axis SD'. Furthermore, when angle α varies from zero to angle γ_1, point Q_R moves from point Q_{0R} to point P_{2R}, and point Q_L from point Q_{0L} to a point on the left-upper part of the curve (see Fig. 11). The projection of the curve onto the positive axis l' extends, while its projection on the other side of the axis shrinks.

In the following, a proof is made to show that length L_α at $\alpha = 0$ is longer than length L_α at $\alpha = (0 \quad \gamma_1]$. Since the ECE length decreases continuously when angle α becomes larger than angle γ_1, there is no need to consider the ECE length for angle α, between angle γ_1 and ninety degree.
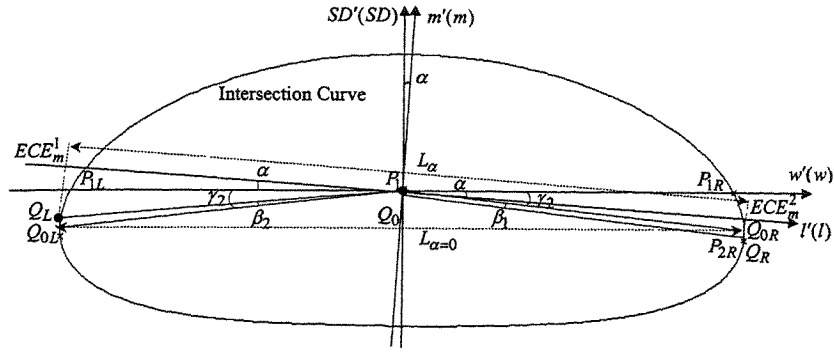
Fig. 11 ECE Length along the Tool Feed Direction, m

From Fig.11, line $Q_{0L}P_1$ is longer than line $Q_L P_1$, say, by length Δ_1; and line $Q_{0R}P_1$ is shorter than line $Q_R P_1$ by length Δ_2. The relation of $\Delta_1 > \Delta_2$, can be deduced from the curve, and line $Q_{0L}P_1$ equals line $Q_{0R}P_1$. The difference between length $L_{\alpha=0}$ and length L_α is defined as

$$L_{\alpha=0} - L_\alpha = Q_{0L}P_1 \cdot \left(2 \cdot \cos \gamma_2 - \cos\left(\gamma_2 - (\alpha - \beta_2)\right) - \cos\left(\gamma_2 + (\alpha - \beta_1)\right)\right)$$
$$+ \Delta_1 \cdot \cos\left(\gamma_2 + (\alpha - \beta_1)\right) - \Delta_2 \cdot \cos\left(\gamma_2 - (\alpha - \beta_2)\right) \tag{25}$$

Since lengthes Δ_1 and Δ_2 are quite small, and $\cos\left(\gamma_2 + (\alpha - \beta_1)\right) < \cos\left(\gamma_2 - (\alpha - \beta_2)\right) < 1$, the term $\Delta_1 \cdot \cos\left(\gamma_2 + (\alpha - \beta_1)\right) - \Delta_2 \cdot \cos\left(\gamma_2 - (\alpha - \beta_2)\right) \to 0$. Eq. 34 can then be simplified as

$$L_{\alpha=0} - L_\alpha = Q_{0L}P_1 \cdot \left(\left(\cos \gamma_2 - \cos\left(\gamma_2 - (\alpha - \beta_2)\right)\right) + \left(\cos \gamma_2 - \cos\left(\gamma_2 + (\alpha - \beta_1)\right)\right)\right) \tag{26}$$

By the safe argument, if $\beta_1 \le \beta_2$ holds, the following result can be deduced.

$$L_{\alpha=0} > L_\alpha \tag{27}$$

As the tool feed direction diverts from the steepest direction, the length of the ECE decreases. The ECE length reaches its maximum at $\alpha = 0$ and its minimum at $\alpha = \frac{\pi}{2}$.

The tool feed direction in 3-axis machining determines the amount of material to be removed in a tool path step (or tool motion). A change of this direction incurs the change of the machining efficiency. The mathematical proof reveals that when a torus end-mill feeds along the steepest tangent direction on the sculptured surface, the ECE length gets the maximum. The maximum amount of material is then removed in the tool motion, and the cutting efficiency reaches its maximum. On the other hand, if the torus (or flat) end-mill cuts along the direction that is perpendicular to the steepest tangent direction on the tangent plane of the surface, the ECE length is the minimum, as is the cutting efficiency.

Thus, the tool feed direction is a control on the tool path machining efficiency. If the tool feed direction of each tool motion in a tool path is always aligned with the steepest tangent direction of the surface, this tool path is called steepest-directed tool path, and it is the most efficient in cutting the local area covered by the tool path. The steepest-directed tool path presents a new tool path generation principle for 3-axis CNC tool path planning.

4 CASE STUDY

To confirm that the steepest direction is the most efficient tool feed direction in 3-axis CNC machining, a hemi-cylindrical part with a radius of 50 *mm* and a length of 100 *mm*, as shown in Fig. 12, is machined using a VM-5 Victor 4-axis Machining Centre. In the test machining, this part represented in a parametric form $S(\omega,v)$ is set up horizontally. The part material is Ren-Shape corrugated fibreboard. The cutter is a 6.35 *mm* flat end-mill, and the specified surface tolerance is 0.2 *mm*. In order to compare the ECE lengths along different tool feed directions, the cutter starts from six points, A to F, and feeds along different directions. These points share the same parameter ω as 80 degree. The different tool feed directions, measured by angle α between the tool feed direction and the steepest direction of the surface, are listed in Table 1.
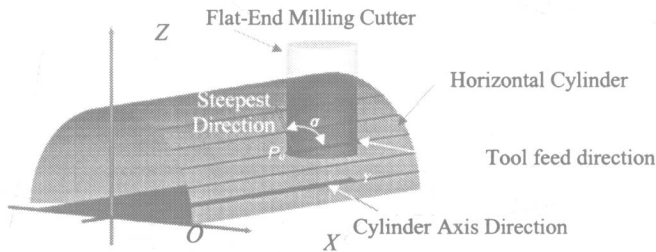
Fig. 12 Horizontal Cylinder and Tool feed directions

Table 1 ECE Length along Different Tool Feed Directions for Different Cutters

Points	Angle α	Measured ECE Length	Theoretical ECE Length	Discrepancy
A	0	7.518 *mm*	7.294 *mm*	3.1 %
B	25.6	6.883 *mm*	6.611 *mm*	4.1 %
C	46.1	5.283 *mm*	5.057 *mm*	4.5 %
D	57	4.165 *mm*	3.978 *mm*	4.7 %
E	68.5	3.026 *mm*	2.860 *mm*	5.8 %
F	74.1	2.515 *mm*	2.363 *mm*	6.4 %

The close-ups of the surface at points A to F are shown in Fig 13. The ECE length at each point is measured manually, and the corresponding theoretical ECE length is calculated using Eq. 27. The discrepancy of the two corresponding ECE length values is less than 6.5 percent, which is at least partially caused by errors of hand measurement. The case study verifies that the ECE length reaches its maximum when the cutter feeds along the steepest direction, as indicated by point A, and decreases considerably when the cutter feeds away from this ideal direction. The steepest direction is proved to be the most efficient tool feed direction.
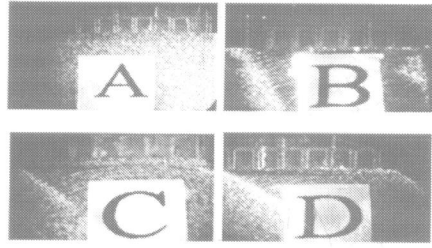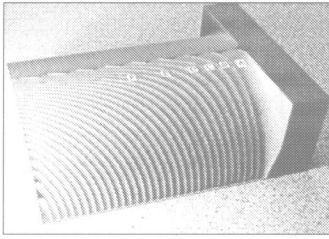
Fig. 13 Machined Hemi-Cylinder Part

5 CONCLUSIONS

Among the sculpture surface geometry, the cutter, the tool path interval and direction, the tool path direction were not taken into consideration in 3-axis tool path planning to increase the machining efficiency. This work initializes research on tool path directions for more efficient tool paths and discloses the relationship between the machining efficiency of a tool motion and its tool feed direction. A new machining efficiency measure, the length of the effective cutting edge (ECE), is introduced in 3-axis CNC milling. The ECE length is mathematically proved to be the maximum when the cutter moves along the steepest tangent direction of the sculptured surface from a cutter contact point. A machining example of a hemi-cylindrical part verifies the new finding.

This study provides better understanding on the relations between tool path geometry and machining efficiency, and introduces a new generic tool path generation principle – the steepest-directed tool path that is the most efficient single tool path. The approach is especially useful to automated CNC tool path generation for complex sculptured surfaces.

ACKNOWLEDGEMENTS

REFERENCE

[1] Broomhead, P., and Edkins, M., 1986, "Generation NC Data at the Machine Tool for the Manufacture of Free-form Surfaces," *International Journal of Production Research*, Vol. 24, No. 1.1-14.
[2] Loney, G.C., and Ozsoy, T.M., 1987, "NC Machining of Free Form Surfaces," *Computer-Aided Design*, 19(2), pp. 85-90, March 1987.
[3] Bobrow, J.E., 1985, "NC Machine Tool Path Generation from CSG Part Representations," *Computer-Aided Design*, 17(2), pp. 69-76.
[4] Huang, Y., and Oliver, J.H., 1994, "Non-constant Parameter NC Tool Path Generation on Sculptured Surfaces," *The International Journal of Advanced Manufacturing Technology*, (1994) 9: 281-290.

[5] Sarma, R., and Dutta, D., 1997, "The Geometry and Generation of NC Tool Paths," *Journal of Mechanical Design*, June 1997, Vol. 119, pp. 253-258.

[6] Suresh, K., and Yang, D.C.H., 1994, "Constant Scallop Height Machining of Free-form Surfaces," *Journal of Engineering for Industry*, May 1994, Vol. 116, pp. 253-259.

[7] Maeng, H., Ly, M., and Vickers, G.W., 1996, "Feature-Based Machining of Curved Surfaces Using the Steepest Directed Tree Approach," *Journal of Manufacturing System*, Vol.15/No.6, pp. 1-13.

[8] Faux, I.D., and Pratt, M.J., 1979, *Computational Geometry for Design and Manufacture*, Ellis Horwood Limited, West Sussex, England

[9] Mortenson, M.E., 1987, *Geometric Modeling*, John Wiley & Sons Inc., New York, 1987.

[10] Marsden J. and Weinstein A., *Calculus III*, Springer-Verlag New York Inc., New York, 1985.

[11] Chen, Z.C., Dong, Z., and Vickers, G.W., 2001, "Steepest-Directed Tool Path in 3-Axis CNC Machining – The Most Efficient Machining Scheme and Its Mathematical Proof," *Proceedings of the 2001 Design Technical Conferences and Computers and Information in Engineering Conference*, ASME, DETC2001/CIE21301.

Parallelly generating NC tool paths for subdivision surfaces

J DAI and **K QIN**
Department of Computer Science and Technology, Tsinghua University, Beijing, China

ABSTRACT

The subdivision surface is the limit of recursively refined polyhedral mesh. It is quite intuitive that the multi-resolution feature can be utilized to simplify generation of NC tool paths for rough machining. In this paper, a new method of parallel NC tool path generation for subdivision surfaces is presented. The basic idea of the method includes two steps: First, extending G-Buffer to a strip buffer (called S-Buffer) by dividing the working area into strips to generate NC tool paths for objects of large size. Second, generating NC tool paths by parallel implementation of S-Buffer based on MPI (Message Passing Interface). The recursion depth of a subdivision surface can be estimated within user-specified error tolerance, then substitute the polyhedral mesh for the limit surface during rough machining. The surface we'll deal with during fine machining is the limit surface of the subdivision. Furthermore, we make use of the locality of S-Buffer and develop a dynamic division and load-balanced strategy to effectively parallelize S-Buffer Method.

Key words subdivision surface, NC tool path, S-Buffer, Parallel computation

1.1 Introduction

Free-form surfaces are standard in CAD/CAM systems, which are widely used in designs of electronic appliances, automobiles, and airplanes. Because control polyhedra of NURBS or B-spline surfaces are restricted to regular meshes, they can represent only limited surfaces of rectangular topology. Subdivision surfaces have emerged recently as an attractive technique in modeling surfaces of arbitrary topology. Currently, there are numerous articles on NC machining of free-form surfaces, but few are dedicated to applications of subdivision surfaces and to how to take advantages of subdivision surfaces in generating NC tool paths.

In practice, a complete NC milling process usually consists of rough machining and fine machining. The main goal of fine machining is for accuracy of the result workpiece, while the most important target of rough machining is to reduce the machining time so as to improve the efficiency. Most of existing NC tool path generation methods derive rough tool paths from fine ones. Some do address the special topic of rough machining and thus develop useful techniques for different applications. But it is hard for them to integrate with the methods used in fine machining.

Motivated by these problems, this paper takes full use of multi-resolution feature of a subdivision surface and the distance bound [2] of the polyhedral mesh to its limit surface, and substitutes the polyhedral mesh within user-specified tolerance for the limit surface during rough machining to make the rough paths simple. A new method called S-Buffer is presented by meliorating G-Buffer method[7] to seamlessly unify the rough and fine path generation, and a parallel method of S-Buffer is implemented using MPI.

2 NC TOOL PATH GERNERATION OF SUBDIVISION SURFACES

2.1 Subdivision surfaces

In CAD/CAM systems, NURBS, B-splines are widely used to free form surface modeling, but they can hardly represent complex shapes of arbitrary topology with just one surface. Subdivision surfaces seem to be more attractive in many fields by generalizing uniform B-spline surfaces to ones of arbitrary topology.

2.1.1 Catmull-Clark surfaces

A subdivision surface is defined as the limit of a sequence of vertices of finer and finer control polyhedron generated in such a way that at each step the old vertices are updated and new vertices are introduced according to subdivision rules.

Assume P_0^n is a vertex with valence N after n times of subdivision. Other $2N$ vertices around P_0^n are labeled as shown in Figure 1. In Catmull-Clark subdivision [1], the new vertices are computed as follows:

Fig. 1 Catmull-Clark subdivision

$$P_{2i}^{n+1} = \tfrac{1}{4}\left(P_0^n + P_{2i-1}^n + P_{2i}^n + P_{2(i\%N)+1}^n\right)$$

$$P_{2i-1}^{n+1} = \tfrac{3}{8}\left(P_0^n + P_{2i-1}^n\right) + \tfrac{1}{16}\left(P_{2i}^n + P_{2(i\%N)+1}^n + P_{2(i-2\%N)+2}^n + P_{2(i-2\%N)+1}^n\right),$$

where $i=1, 2, \ldots, N$, and
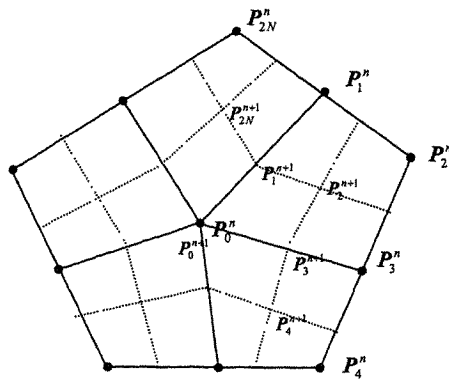
$$P_0^{n+1} = \alpha_N P_0^n + \beta_N \left(\frac{1}{N} \sum_{j=1}^{N} P_{2j-1}^n \right) + \gamma_N \left(\frac{1}{N} \sum_{j=1}^{N} P_{2j}^n \right),$$

where $\alpha_N, \beta_N, \gamma_N \geq 0$, $\alpha_N + \beta_N + \gamma_N = 1$. For example,

$$\alpha_N = 1 - 7/(4N), \quad \beta_N = 3/(2N), \quad \gamma_N = 1/(4N). \tag{1}$$

Let $C_n = \left(P_0^n, P_1^n, \cdots, P_{2N}^n \right)^T$, then $C_{n+1} = AC_n$, where A is the subdivision matrix [2].

2.1.2 Estimating the error between a polyhedral mesh and its limit surface

After certain steps of subdivision, the polyhedral mesh becomes more and more approximate to the smooth limit surface. In order to ensure that the machined surfaces are enough accurate, it is very important to compute the depth of recursion within a user-specified error bound.

Define S_0^n as the set of control vertexes created from certain initial vertex P_0^0 after n steps of recursive subdivision, L(P) as the corresponding point in the limit surface for a control vertex P. Let Circle(P_0^n) = $\left\{ P_0^n, P_1^n, \cdots, P_{2N}^n \right\}$ denote the set of the vertices in the local structure of P_0^n, $C(P_0^n) = \frac{1}{2N} \sum_{j=1}^{2N} P_j^n$, and $abs(F) = \left(|f_{ij}| \right)$ for a matrix $F = \left(f_{ij} \right)$. Then the distance of the control polyhedron to the limit surface after n Catmull-Clark subdivision steps has the following bound[2]:

$$\max_{P \in S_0^n} \left\| P - L(P) \right\|_1 \leq \rho^n K \max_{P \in S^0} \max_{Q \in \text{Circle}(P)} \left\| C(P) - Q \right\|_1, \tag{2}$$

where

$$K = \max_{3 \leq N \leq M} \left\| U \cdot abs(V) diag(0, \left| \frac{\lambda_2}{\lambda_2} \right|^P, \cdots, \left| \frac{\lambda_{2N+1}}{\lambda_2} \right|^P) abs(V^{-1}) \right\|_\infty,$$

$$\rho = \max_{3 \leq N \leq M} \left| \lambda_2 \right|,$$

$$U = diag(1, 0, 0, \cdots, 0),$$

λ_i is the i-th eigenvalue of the subdivision matrix A, and V an invertible matrix whose columns are the corresponding eigenvectors of A.

2.2 Generating tool paths for subdivision surfaces

Tool path generation is the core task of an NC Machining system. A difficult task of the path plan is detection and avoidance of interference and collision. Besides, it's necessary to take path verification and feed rate control into account during path generation.

Precision and efficiency are the two main considerations of NC machining. To reduce machining time without influence on precision, the rough machining should be treated specially, such as simplifying tool paths according to the shape of the original workpiece, slicing multi-layer paths and so on. Meanwhile, these techniques should seamlessly integrate those used in the tool path generation for fine machining.

It is well known that subdivision surfaces are defined as the limit of recursively refined polyhedral meshes. The degree of approximation of a control mesh to the limit surface after any step during the subdivision process can be calculated [2]. So it is natural to make use of this multi-resolution feature to simplify the tool paths of rough machining to improve the efficiency. That is, we choose the control polyhedral meshes after n steps of subdivision as the substitution for the limit surface during the tool path computation of rough machining. The number n is dependent on the precision required for rough cutting and can be easily computed according to Eq. (2).

There are two basic considerations in generating tool paths for subdivision surfaces:
(1) Take the control mesh after enough steps of subdivision as the workpiece during rough machining;
(2) Adopt a unified method for computation of both rough tool paths (from polyhedral meshes) and fine tool paths (from smooth limit surfaces).

2.2.1   S-Buffer

G-Buffer [7] is a very attractive method for NC tool path generation. Based on the Z-Buffer method in computer graphics, it can deal with any surfaces, either analytical, free-form, trimmed surfaces, or polyhedral faces.
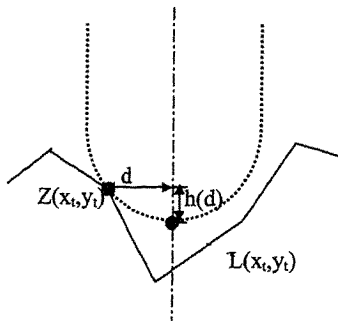
Fig.2 Touch point of tool and wokpiece   Fig.3 Dividing working area into strips
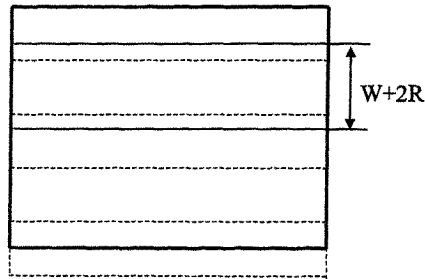
Fig.2 shows how to obtain tool path buffer from a workpiece buffer. Let Z and L be the buffers of the workpiece and the desired tool path, separately, and h(d) be the height of the tool-end at the distance d from the tool axis. When the tool-end touches the workpiece at (x_t, y_t), we have:

$$L(x,y) = Z(x_t,y_t) - h(x - x_t, y - y_t)$$

Therefore, it is easy to get L as follows.

$$L(x,y) = \max_{i,j}(Z(x+i, y+j) - h(i,j)), (i^2 + j^2 < r^2)$$ (3)

where h(i,j) depends on the shape of tool-end and r is the radius.

Now we can obtain different kinds of paths with different scanning strategies.

It is shown in Fig.2 that G-Buffer can not only deal with different shapes of tools, but also create interference-free tool paths. However, the requirement of memory is too large with G-Buffer. If we want to machine a workpiece with 1-meter long and 1-meter wide with the error bound of 0.1 millimeters, and we only record the height field and the normal at each pixel in float format (32 bits), then 1.6G Bytes are needed! This is beyond the maximum size assigned for a single thread in Windows NT/2000.

Based on G-Buffer, we divide the working area into strips and develop a new method, called Strip Buffer (*i.e.*, S-Buffer), as follows:

✦ Dividing the working area

As illustrated in Fig.3, we divide the whole area into many strips with the width of *W*. According to (3), we need to include two adjacent R-wide regions in adjacent strips. So the width of the strip is actually $H = W + 2R$. The value of constant *W* depends on the system configuration. In general, we assign an appropriate value to make full use of the physical memory and to reduce the swap between the memory and the hard disk of the computer system.

✦ Generating NC tool paths with S-Buffer

For each strip, we create two buffers of both the workpiece and tool paths, then compute like G-Buffer to obtain tool paths of the strip.

✦ Simplifying the tool paths

a) Before breaking   b) After breaking
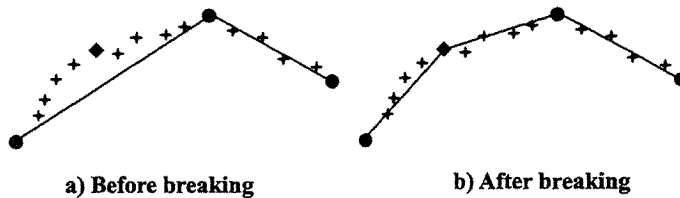
Fig 4 Simplifying the paths

The result paths from the above steps are discretized on pixels. It is indispensable to fit the tool paths with longer line segments within the user-specified tolerance to improve machining efficiency, especially during rough cutting process. A scheme for the line fitting is presented as follows:

- First of all, define the joint points of the line segments of the final paths as "key points", and initially take the start point, end point in the tool path buffer as key points. Connect the adjacent key points to form a poly-line.

- Let S_i denote the set of n_i pixels $(P_0, P_1, \dots, P_{n-1})$ between the i-th pair of key points. Check whether the distance of each point of S_i to the line defined by the key point pair. If the distance is less than the tolerance specified by the user, check next pair of the adjacent key points. Otherwise, break the line into two segments at $P_{(n-1)/2}$, and check these two shorter line segments again (see Fig. 4). The joint point $P_{(n-1)/2}$ is regarded as a new key point added. This "search-and- break" process is recursively done until the distance of each point to the corresponding line defined by the two key points is less than the given tolerance.

❖ Totally combining tool paths into a complete path
  After obtaining all strips' paths, we combine them into a complete NC tool path by zigzag or spiral scanning.

3 PARALLELIZING S-BUFFER

3.1 Parallel computing environment

With the rapid development of network technology and the ever-growing performances of PCs, clusters and other parallel architectures make it possible to constitute high performance computing environments with PCs connected by network. MPI is the standard for message passing among multi-computers and cluster. Using MPI, components of the virtual machine communicate and cooperate with each other by passing messages.

3.2 Program architecture

There are two kinds of patterns in MPI programming: peer-to-peer pattern and Master/Slave pattern. It is obvious that Master/Slave pattern is more suitable for parallelizing S-Buffer. Master process runs on a local host, integrated in CAD/CAM systems, providing a visual interface to users. Slaves distribute on nodes connected to the local host, performing computation assigned by the Master. In this architecture, it is all the Master's duty to distribute tasks, to control slaves and to generate global tool paths after receiving results of all strips.

Parallelizing S-Buffer consists of three main phases:(1) Master calculates the subdivision depth, generates the corresponding polyhedral mesh and sends it to slaves;(2) Slaves compute tool paths of the strips separately and return results to Master;(3) Master generates the global path.

3.3 Data division and parallel granularity

Data division strategy and parallel granularity have great effect on speedup, which is the main index of the performance of a parallel algorithm. As described in 2.2, S-Buffer method divides working area into strips with the width of W, and needs redundant computing of $2R$-wide

adjacent region (for non-margin strips) for each strip. If W is too small, then the redundancy and communication loads will affect the parallel speedup. On the contrary, if we assign a too big value to W, then the degree of parallelism is not high enough, and thus the speedup decreases. Therefore, the value of W is dependent on the total working load, the number of nodes of the virtual machine and their computational capabilities, and the communication capability of the network connecting virtual machine.

As will be discussed in Section 3.4, assigning tasks of stationary size to different slaves is not appropriate in a distributed environment. Although we divide the whole working area into strips of the same size to make the task assignment strategy simple, the number of strips assigned to certain slave should be dependent on its capability. Thus, our data division strategy is adaptive and can contribute to a high speedup

3.4 Performance index and dynamic task assignment

Subtask assignment and load balance strategy is a dominating factor of speedup. In certain circumstances where the performances of nodes and network are stable, we can assign the tasks statically. But in most cases, the virtual machine runs under a dynamic environment. The performances of the nodes and network vary from time to time. So it is necessary to take a dynamic strategy for the task assignment and load balance.

To achieve a high speedup, the load assigned to a node should be decided by its capability. In the parallelization of S-Buffer, there is no need to communicate among slaves. So the computing performance of a node and the communication capability of the network can be comprehensively indicated by the response time, which is the time interval on the master from sending a slave commands to receiving results. Besides, the performance of a node varies dynamically. Therefore, the performance index of a node should reflect this change adaptively. We define weighted-average response time T_i (*i.e.*, the index of performance) as follows:

When receiving a result of a task from the i-th slave, the master refreshes T_i and calculates M_i, which is the number of strips that should be assigned to the slave this time. Let N_i be the total number of strips that have been dealt by slave i, n_i the number of strips in this task, and t_i the response time of this task. Then:

$$T_i = \frac{T_i \times N_i}{N_i + n_i} + \frac{t_i}{n_i}, \text{ and } N_i = N_i + n_i \quad (\text{initially } T_i = N_i = 0);$$

$$M_i = \begin{cases} \left\lfloor \dfrac{\sum T_k}{N} \bigg/ T_i \right\rfloor, & N \neq 0 \\ 1, & N = 0 \end{cases}$$

where N is the number of the nodes whose $N_i > 0$.

Our task assignment strategy is also preemptive. Let C be the number of strips that have never been dealt with, E_u the time that has elapsed since the u-th strip was assigned, and a constant k ($0<k<1$) "preemptive coefficient". Define "dealing strips" as strips that certain

slaves are dealing with. When $M_i > C$, the master assigns all the C strips to slave i, then traverses all the "dealing strips". For each dealing strip (marked here by strip u), with which some slave (say, slave j) is dealing, if $T_i < k * (T_j - E_u)$, then the master deprives slave j of strip u and assign it to slave i, until the number of strips that have been assigned to slave i is equal to M_i.

4   EXAMPLES

a) Initial mesh   b) Divided mesh   c) Realistic image

d) Tool paths of 7 layers   e) Single path
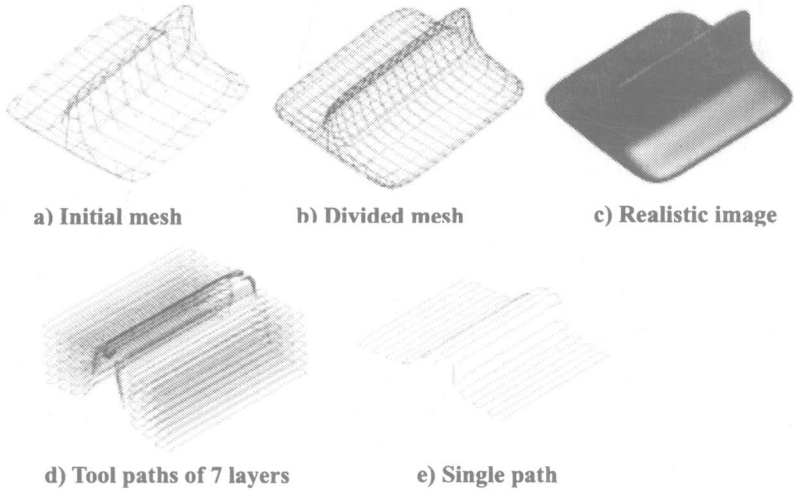
Fig. 5 Machining a T-shaped object

a) Initial mesh   b) Divided mesh   c) Realistic image

d) Multilayer paths   e) Single path
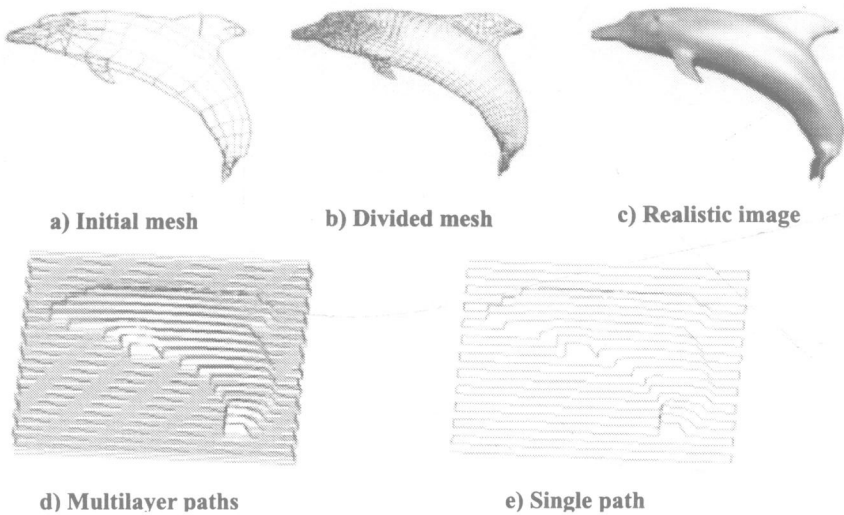
Fig. 6   Machining a dolphin

The S-Buffer method has been implemented on a local 10M ethernet network of seven PCs. The software platform consists of Windows 2000 and the corresponding WMPI v1.5.Figures 5 and 6 illustrate the rough tool paths of a T-shaped object and a dolphin model, respectively.

We have tested the parallelized S-Buffer Method with data as follows: The working area is 0.952-meter long and 1.99-meter wide, using a spherical milling tool with the radius of 8 millimeters, and the size of a pixel is 0.5 millimeters. Table 1 demonstrates the time costs for different H, and Fig. 7 illustrates the speedups for different numbers of nodes. It is shown that the value of H affects time costs greatly, because the ratio of redundant calculation is determined by the value of H, but the speedup is not sensitive to the number of nodes, because the computation of S-Buffer is a time-consuming process (more than 400 seconds) and thus the cost of communication is just a little in comparison with the computing time. Therefore, the scalability of S-Buffer is very satisfying.

Table 1 Time costs (second)with different H

	1	2	3	4	5	6	7
H=100	671.35	336.11	224.48	168.64	134.95	112.56	96.56
H=200	414.73	207.06	138.32	103.95	83.25	69.42	59.68
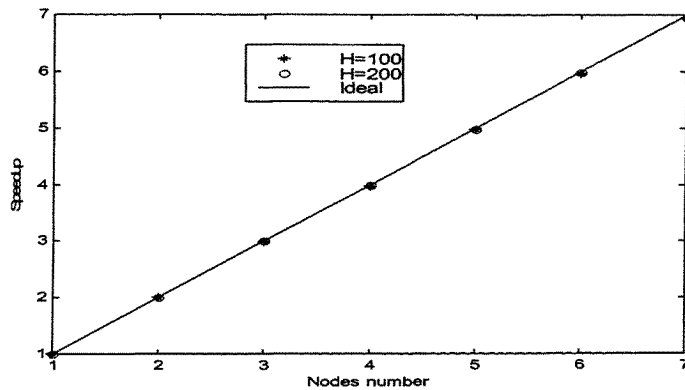
Fig 7 Scalability of the virtual machine

5   CONCLUSIONS

Subdivision surfaces are widely adopted in computer aided geometric design and modeling. We can substitute the polyhedral meshes of subdivision for the smooth limit surfaces during tool path generation for rough NC machining. The S-Buffer method presented can combine the rough and the fine machining seamlessly and generate NC tool paths for workpieces of

large size on a single PC. It is shown that the parallel S-Buffer method is very efficient, scalable and robust.

ACKNOWLEDGEMENT

REFERENCES

1   **E Catmull** and **J Clark.** Recursively Generated B-spline Surfaces on Arbitrary Topological Meshes. *Computer-Aided Design,* 10:350-355, 1978

2   **Huawei Wang, Kaihuai Qin.** Estimate of the Subdivision Depth of Catmull-Clark Surfaces within User-specified Error Bounds to the Limit Surfaces

3   **J Stam.** Exact Evaluation of Catmull-Clark Subdivision Surfaces at Arbitrary Parameter Values. *Computer Graphics (Proceedings of SIGGRAPH'98).* 1998

4   **U Reif.** A Unified Approach to Subdivision Algorithms near Extraordinary Vertices. *Computer Aided Geometric Design,* 12:153-174,1995

5   **Qin Kaihuai, Wang Huawei.** Eigenanalysis and Continuity of Non-Uniform Doo-SabinSurfaces, Proceedings of Pacific Graphics'99, SNU, Korea, Oct. 1999, p.179-196.

6   **Qin Kaihuai, Wang Huawei.** Continuity of Non-Uniform Recursive SubdivisionSurfaces, Science in China (Series E), Vol.43, No.5, 2000, p.461-472.

7   **Takafumi Saito, Tokiichiro Takahashi.** NC Machining with G-buffer Method, Computer Graphics, Volume 25, Number4, July 1991

The geometrical theory of machining free form surface by cylindrical cutter in five-axis NC machine tools

L X CAO, H J WU, and J LIU
School of Mechanical Engineering, Dalian University of Technology, China

SYNOPSIS

In this paper, on the basis of the differential geometry, the geometrical theory of machining free form surface by cylindrical cutter in 5-axis NC machine tool is investigated. The second order osculating condition between the cutter and the theoretical surface is analyzed. By means of the condition of positive semidefinite of a quadratic form, the location of the tool-axis can be acquired. The osculating condition between the cutter and the theoretical surface is equivalent to the osculating condition between the tool-axis and the offset surface of the theoretical surface and has been demonstrated. By means of the proposed machining method, the cusp height can be limited within the third order infinitesimal. The machining precision and the efficiency can be improved, and the potential ability of the 5-axis NC machine tools could be developed.

1 INSTRUCTION

In the last decade, the machining technology of free form surface by means of 5-axis NC machine tools was developed very quickly, and has been the popular research projects in the world. It has become an important method for machining complex free form surface and has been applied to the automotive, aerospace, shipbuilding, turbine industries, etc. In general, a

---

free form surface is machined with a ball-ended tool. Its main advantages include: the spherical surface can be adapted to the normal of the theoretical surface automatically, the number of the axis of the machine tools can be reduced, the programming will be simple, and if the radius of the cutter is smaller than the minimum curvature radius of the machined surface, then the interference can be avoided. Its main disadvantages are that: the machining precision and the efficiency are lower. For these reasons, many types of cutter are developed, such as flat-end cutter, filleted endmill cutter, disc cutter, drum-head cutter, cylindrical cutter, conical mill and torus cutter, for improving machining precision and efficiency. The flat-end cutter, disc cutter and torus cutter are theoretically suitable for machining the large, even and opened free form surface, but the drum-head cutter, cylindrical cutter and the conical mill are suitable for machining complex, small and medium dimensional surface[1]. Since the cylindrical cutter is simple in its construction, fast in material-removal rates, melioration in surface finish and can be used for machining tunnels[2], such as the milling of the integral turbo-wheel of turbo-compressor[1]. Hence, the cylindrical cutter has been used more popularly than others. The interference of non-ball end cutter is more complex than that of ball-end cutter. So, the researches are concentrated on the gouge-free tool-path generation, the analysis of cutting error, local and global tool accessibility, etc. In 1992, Wang[3] present the curvature-catering method for machining sculptured surface with a disc cutter. This method can ensure each strip-shaped envelope formed by tool nose trace and the theoretical surface to have the same derivatives up to the third order, thus the steps required can be greatly reduced. In 1993, Fritz[4] analyzed the characteristics of ruled surface, and proposed a method to milling the ruled surface by cylindrical cutter. He pointed that this kind of surface was composed of hyperbola points and if the tool-axis is located at the principal direction of the minus principal curvature, the interference can be avoided. In 1995, Liu[1] investigated the algorithms for the tool-path generation of 5-axis machining the sculptured surfaces using cylindrical cutter, and presented a single point offset algorithm and a double point offset algorithm for calculation of cutter location. In 1997, by using filleted endmill and considering both local and global surface shapes, Lee[5] presented a methodology and algorithms of admissible tool orientation control for gouging avoidance in 5-axis machining. Unlike the traditional graphical verification and user-interactive correction of tool-path generation, Lee's methodology can be used to generate cutter path for 5-axis machining automatically. In the same year, Elber[6] presented a 5-axis side milling scheme for freeform surface based on automatic piecewise ruled surface approximation. In 1998, Lee[7] developed an ellipse-offset method for 5-axis rough cutting of ruled surface pockets with cylindrical cutter. In 2000, Monies[8] investigated a 5-axis side milling method for ruled surfaces using a conical milling cutter. In 2001, Ni[9] present a methodology and algorithm of optimal torus tool orientation control for 5-axis NC machinging.

Even though the 5-axis NC machine tools have been used for machining free form surface more frequently, and many achievements have been achieved. The lack of understanding of the complex tool movements and geometric relationships makes 5-axis machining a difficult task to perform. Just for the machining methods, the ability of the 5-axis NC machine tools has not been developed fully. In this paper, we proposed a new method for machining the free form surface. By means of adjusting the position of the axis of the cylindrical cutter in every stroke, let the cutter surface and the theoretical surface reach osculation or near osculation. Thus the machining efficiency can be greatly increased.

2 FUNDAMENTAL PRINCIPLE OF THE SECOND ORDER OSCULATING MACHINING METHOD

As shown in Figure 1, point P_0 is the tangent point of the cylindrical cutter to the machined surface Σ, i.e. cutter-contact (CC) point. P_0—$e_1e_2e_3$ is defined to be the Frenet frame, e_1, e_2 are the unit vectors of the surface Σ along the principal directions at the point P_0, respectively, and e_3 is the unit normal vector of the surface Σ at point P_0, and define e_3 is positive when it points at out-space.. The point C is the intersection point of e_3 and the tool-axis, its radius vector can be written as: $r_c = r_0 + Re_3$, where R denotes the radius of the cutter. r_0=r (u_0,v_0) is the radius vector of the point P_0. r=r (u,v) is the radius vector of the point P of the machined surface Σ. t is the unit vector of the tool-axis. its vector equation can be written as

$$t = e_1 \cos\alpha + e_2 \sin\alpha \qquad (1)$$

where α denotes the angle between the unit vector t and the unit vector e_1.
The minimum distance between the point P and the tool-axis can be written as

$$L = \sqrt{(r_0 + Re_3 + (\Delta r \cdot t)\, t - r)^2} = \sqrt{R^2 - 2R\Delta r \cdot e_3 + (\Delta r \times t)^2} \qquad (2)$$

where, $\Delta r = r - r_0$
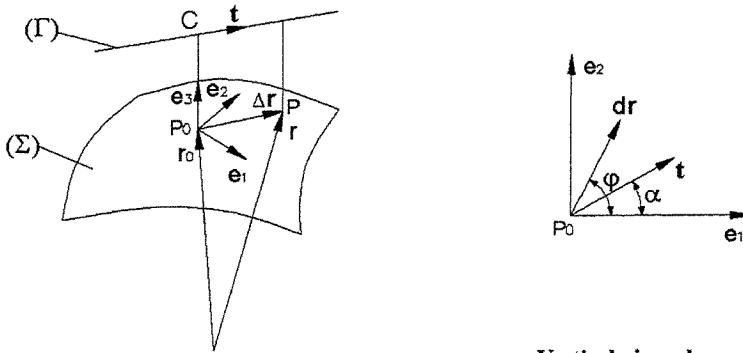
Vertical view along e_3

Figure 1. Second order osculating method for 5-axis machining with cylindrical cutter

In order to discuss the osculation between the cutter and the machined surface Σ, we will focus our attention on the neighboring region of the point P_0 in the following discussion. Since the $\Delta r \cdot e_3$ and $(\Delta r \times t)^2$ of the Eq. (2) are second order quantity, if we neglect the fourth order quantity and use the approximation equation of the extraction of root, then we have

$$L = R - \Delta r \cdot e_3 + \frac{(\Delta r \times t)^2}{2R} \qquad (3)$$

the normal distance (cutting remaining error) between the point P and the cylindrical surface of the cutter can be written as

$$\delta = L - R = -\Delta\mathbf{r} \cdot \mathbf{e}_3 + \frac{(\Delta\mathbf{r} \times \mathbf{t})^2}{2R} \qquad (4)$$

expanding $\Delta\mathbf{r}$ in Talor series at the point P_0, we have

$$\Delta\mathbf{r} = \mathrm{d}\mathbf{r} + \frac{\mathrm{d}^2\mathbf{r}}{2} + \varepsilon \qquad (5)$$

where ε denotes the remainder, namely, third order remainder and above.
By using of differential equations of the frame[10], we have

$$\mathrm{d}\mathbf{r} = \mathbf{e}_1\sigma_1 + \mathbf{e}_2\sigma_2 = (\mathbf{e}_1\cos\varphi + \mathbf{e}_2\sin\varphi)\sigma \qquad (6)$$

$$\mathrm{d}^2\mathbf{r} = -\mathbf{e}_1(\omega_3 + \mathrm{d}\varphi)\sin\varphi\sigma + \mathbf{e}_2(\omega_3 + \mathrm{d}\varphi)\cos\varphi\sigma + \mathbf{e}_3(\omega_1\sin\varphi - \omega_2\cos\varphi)\sigma \qquad (7)$$

where $\qquad \sigma^2 = \sigma_1^2 + \sigma_2^2 \qquad (8)$

$$\omega_1 = k_2\sigma_2 \qquad \omega_2 = -k_1\sigma_1 \qquad \omega_3 = k_{g1}\sigma_1 + k_{g2}\sigma_2 \qquad (9)$$

σ_1, σ_2 denote the infinitesimal arc-lengths along $\mathbf{e}_1$, $\mathbf{e}_2$ at the point P_0 of the machined surface Σ respectively

σ denotes the infinitesimal arc-length along $\mathrm{d}\mathbf{r}$ at the point P_0 of the machined surface Σ

φ denotes the angle between $\mathrm{d}\mathbf{r}$ and $\mathbf{e}_1$

$\omega_1, \omega_2, \omega_3$ denote the infinitesimal rotating quantities about $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ respectively.

k_1, k_2 denote the principal curvatures of the point P_0 of the machined surface Σ, respectively

k_{g1}, k_{g2} denote the geodesic curvatures along the curvature lines, respectively

Substituting Eqs. (6), (7) and (9) into Eq. (5), we have

$$\Delta\mathbf{r} \cdot \mathbf{e}_3 = \frac{\sigma^2}{2}\left(k_1\cos^2\varphi + k_2\sin^2\varphi\right) + \varepsilon \cdot \mathbf{e}_3 \qquad (10)$$

Since $\mathrm{d}\mathbf{r} \cdot \mathrm{d}^2\mathbf{r} = 0$, we have

$$(\Delta\mathbf{r} \times \mathbf{t})^2 = |\Delta\mathbf{r}|^2 \sin^2(\varphi - \alpha) = \sigma^2\sin^2(\varphi - \alpha) + \varepsilon_4 \qquad (11)$$

where ε_4 denotes fourth order remainder and above
Substituting Eqs. (10), (11) into Eq. (4), we have

$$\delta = \frac{\sigma^2}{2}\left[\cos^2\varphi\left(k_1\sin^2\alpha - k_1\right) - 2\sin\varphi\cos\varphi k_1\sin\alpha\cos\alpha + \sin^2\varphi\left(k_1\cos^2\alpha - k_2\right)\right] + \varepsilon \cdot \mathbf{e}_3 \qquad (12)$$

where $k_t = \dfrac{1}{R}$

In order to discuss the second order osculation between the cutter and the machined surface, the third and higher order remainder are neglected. Then the second order remaining error can be written as

$$\delta = \frac{\sigma^2}{2}\left[\cos^2\varphi\left(k_t\sin^2\alpha - k_1\right) - 2\sin\varphi\cos\varphi k_t\sin\alpha\cos\alpha + \sin^2\varphi\left(k_t\cos^2\alpha - k_2\right)\right] \qquad (13)$$

From the Eq. (13), we know that the second order remaining error is only related to φ and α when σ, k_1, k_2 and k_t are given. The changing of the angle φ means the changing of the position of a given point among the neighboring region, and the changing of the angle α means the changing of the orientation of the tool-axis. In order to avoid the interference between the cutter and the machined surface, the second order remaining error must be non-negative, i.e. $\delta \geq 0$. In the following, we will discuss the conditions about $\delta \geq 0$. The concrete method is adjusting the angle φ firstly to make the second order remaining error be a minimum extreme value, then adjusting the angle α to make the second order remaining error equals zero, i.e. the cutter is osculating with the machined surface Σ. According to the above discussion, the Eq. (13) can be written as a quadratic form

$$\delta = \begin{bmatrix}\cos\varphi \\ \sin\varphi\end{bmatrix}^T \begin{bmatrix} k_t\sin^2\alpha - k_1 & -k_t\sin\alpha\cos\alpha \\ -k_t\sin\alpha\cos\alpha & k_t\cos^2\alpha - k_2 \end{bmatrix} \begin{bmatrix}\cos\varphi \\ \sin\varphi\end{bmatrix} \qquad (14)$$

Clearly, the condition $\delta \geq 0$ is equivalent to the quadratic form be positive semidefinite. From reference [11], we know that this condition is equivalent to the following three condition equations

$$k_t\sin^2\alpha - k_1 > 0 \qquad (15)$$

$$k_t\cos^2\alpha - k_2 > 0 \qquad (16)$$

$$\begin{vmatrix} k_t\sin^2\alpha - k_1 & -k_t\sin\alpha\cos\alpha \\ -k_t\sin\alpha\cos\alpha & k_t\cos^2\alpha - k_2 \end{vmatrix} = 0 \qquad (17)$$

Eqs. (15) and (16) are the constrained conditions of the radius of the cutter. From engineering view point, the radius of the cutter must be smaller than the positive principal curvature radius of the machined surface Σ, i.e., $k_t > k_1$, $k_t > k_2$. Thus, Eqs. (15)、(16) are satisfied. From Eq. (17), we have

$$\cos^2\alpha = \frac{k_1 k_2 - k_2 k_t}{k_1 k_t - k_2 k_t} \qquad (18)$$

Substituting Eq. （18） into Eq. （13） and let it equals to zero, then we have

$$tg\varphi = \frac{k_t \sin\alpha \cos\alpha}{k_t \cos^2\alpha - k_2} \qquad (19)$$

The orientation of the tool-axis can be determined by the angle α of Eq. （18）, and the osculating direction between the cutter and the machined surface Σ can be determined by the angle φ of Eq. (19). This means that the position of cutter can be determined if the point P_0, k_1, k_2, e_3 and R are given. And the surface Σ can be machined at the status of the second order osculation, i.e. line contact machining.

3 FUNDAMENTAL GEOMETRICAL CHARACTERISTICS OF THE SECOND ORDER OSCULATING MACHINING METHOD

From the above, we know that the minimum distance between the tool-axis Γ and the machined surface Σ is equals to the radius R of the cylindrical cutter. Assume Σ^* be the offset surface of the surface Σ with the offset distance R. This implies that the tool-axis must be a tangent line of the offset surface Σ^*. For this reason, we can discuss the second order osculating machining method from the contacting order between the tool-axis and the offset surface Σ^*. From the knowledge of differential geometry, we know that the surface Σ and Σ^* have the same normal vector and the same principal directions at the corresponding point. And the principal curvatures have the following relations:

$$k_1^* = \frac{k_1}{1 - k_1 R} = \frac{k_1 k_t}{k_t - k_1} \qquad (20)$$

$$k_2^* = \frac{k_2}{1 - k_2 R} = \frac{k_2 k_t}{k_t - k_2} \qquad (21)$$

If the asymptotic directions exist on the offset surface Σ^*, then from the Euler's formula, we have

$$tg^2\varphi_j^* = -\frac{k_1^*}{k_2^*} \qquad (22)$$

where φ_j^* denotes the angle between the asymptotic directions and the principal direction e_1 of the offset surface Σ^*
Substituting Eqs. （20） and （21） into Eq. （22）, we have

$$tg^2\varphi_j^* = \frac{k_1 k_2 - k_1 k_t}{k_2 k_t - k_1 k_2} \qquad (23)$$

Changing Eq. （18）, we also have

$$tg^2\alpha = \frac{k_1 k_2 - k_1 k_t}{k_2 k_t - k_1 k_2} \qquad (24)$$

Comparing Eq.(23) and Eq.(24), we can achieve an important conclusion: the angle between the asymptotic direction of the offset surface Σ^* and the principal direction e_1 is equals to the angle between the tool-axis and the principal direction e_1. In other words, if the tool-axis and the offset surface Σ^* keeps the second order contact, i.e. the tool-axis is located at the asymptotic direction of the offset surface Σ^*, then the cylindrical cutter keeps the second order osculation with the machined surface Σ; conversely, if the cylindrical cutter keeps the second order osculation with the machined surface Σ, then the tool-axis must be located at the asymptotic direction of the offset surface Σ^*.

Combining Eq. (19) and Eq. (24), we have

$$tg\,\varphi\,tg\,\alpha = -\frac{k_1}{k_2} = tg^2\varphi_J \qquad (25)$$

where φ_J denotes the angle between the asymptotic directions and the principal direction e_1 of the machined surface Σ

Eq. (25) gives a very simple relationship about φ, α and φ_J. If we know the two parameters of these parameters, then we can have the other parameter.

In the following, let's discuss the conditions of the second order osculating machining method which can make the cylindrical cutter keep line contact with the machined surface. From Eq. (24), we know that the right term of the Eq. (24) must be non-negative. Thus we have

$$k_1 k_2 (k_t - k_1)(k_t - k_2) \leq 0 \qquad (26)$$

From Eqs. (15) and (16), we know: $(k_t - k_1)(k_t - k_2)$ is positive, so we have $k_1 k_2 \leq 0$. This implies that the second order osculating machining method is suitable for surface, which is composed of hyperbola points or parabola points. But for surface that composed of ellipse points, this method is not suitable. In order to decrease the machining error, the tool-axis should be placed at directions where the absolute value of normal curvature is minimum.

4 CUSP HEIGHT OF THE SECOND ORDER OSCULATING MACHINING METHOD

In the second section, we discussed the second order osculating condition between the cylindrical cutter and the machined surface. By means of this condition, the orientation of the tool-axis can be determined. When the point P_0 is moving along the parametric curve of the machined surface, then a strip region can be formed by the envelope of the cutter on the workpiece. Since the cutter and the machined surface keeps the second order osculation at every tool-path, the width of the strip region is increased greatly, i.e. the cutter keeps line contact with the machined surface Σ. In the following, we will discuss how to determine the width of the strip region and the effective length of the cutter, to make the cusp height

between the two strips is less than given tolerance. These can be analyzed by means of the third order remaining error. From the above discussion, we know that the third order remaining error between the cutter and the machined surface can be written as

$$\varepsilon_3 = \varepsilon \cdot e_3 = -\frac{1}{6} d^3 r \cdot e_3 \qquad (27)$$

From Eqs. (7), (8) and (9), we have

$$d^2 r \cdot e_3 = (k_1 \cos^2 \varphi + k_2 \sin^2 \varphi) \sigma^2 \qquad (28)$$

Differentiating Eq. (28), we have

$$d^3 r \cdot e_3 + d^2 r \cdot de_3 =$$
$$\sigma^3 [k_{11} \cos^3 \varphi + k_{12} \cos^2 \varphi \sin\varphi + k_{21} \cos\varphi \sin^2 \varphi + k_{22} \sin^3 \varphi + 2(k_2 - k_1) \sin\varphi \cos\varphi d\varphi] \qquad (29)$$

If we discuss the third order remaining error at the normal section of the point P_0 of the machined surface Σ, we have

$$d^2 r \cdot de_3 = 0 \qquad (30)$$

Substituting Eq. (7) into Eq. (30). By using differential equations of the frame [13], we have

$$d\varphi = -\sigma (k_{g1} \cos\varphi + k_{g2} \sin\varphi) \qquad (31)$$

By using the structure equation of the theory of surface[11], we know

$$k_{12} = k_{g1}(k_1 - k_2) \qquad (32)$$

$$k_{21} = k_{g2}(k_1 - k_2) \qquad (33)$$

Substituting Eqs.(29), (30), (31),(32) and (33)into Eq. (27). By using differential equations of the frame [11], we have

$$\varepsilon_3 = \frac{\sigma^3}{6} [k_{11} \cos^3 \varphi + 3k_{12} \cos^2 \varphi \sin\varphi + 3k_{21} \cos\varphi \sin^2 \varphi + k_{22} \sin^3 \varphi] \qquad (34)$$

where $\quad k_{11} = d k_1 \sigma_1 \quad k_{12} = d k_1 \sigma_2 \quad k_{21} = d k_2 \sigma_1 \quad k_{22} = d k_2 \sigma_2$

$\quad k_{11}$、k_{12} denote the partial derivative of k_1 with respect to e_1 and e_2 respectively

$\quad k_{21}$、k_{22} denote the partial derivative of k_2 with respect to e_1 and e_2 respectively

From the Eq (34), we know that the third order remaining error is in direct proportion to σ^3. When the σ is given, then the third order remaining error can be achieved. In the same way, if the tolerance is given, then we can also achieve the σ. The effective radius of the cutter

corresponding to σ can be written as

$$s = 2\sigma \cos(\varphi - \alpha)\tag{35}$$

5 CONCLUSION

- By using the cylindrical cutter, the surface composed of hyperbola points or parabola points can be machined by the second order osculating machining method. In the machining process, the tool-axis should be adjusted correspondingly. In addition, the important parameter α can be determined by the principal curvatures and the radius of the cutter.
- If the tool-axis is located at the asymptotic direction of the offset surface with the offset distance equals the radius of the cutter, then the cylindrical cutter keeps the second order osculation with the machined surface; conversely, if the cylindrical cutter keeps the second order osculation with the machined surface Σ, then the tool-axis must be located at the asymptotic direction of the offset surface Σ^*.
- By using the cylindrical cutter, the cusp height of the second order osculating machining method is far smaller than the cusp height of using ball-end cutter at the same steps. The cusp height can be evaluated by the third order infinitesimal. This is very helpful for developing machining precision and machining efficiency of the 5-axis NC machine tools.

REFERENCES

1 **Liu, X. W.** (1995) Five-axis NC cylindrical milling of sculptured surfaces. Computer-Aided Design, Vol.27, No.12, pp.887-894.
2 **Jiao J. B., Yu H.** (2001) Study on the tolerance in 4-coordinate side milling of ruled surface. Chinese Journal of Mechanical Engineering, Vol.37, No.4, pp.44-47.
3 **Wang X. C., Li Y. et al.** (1992) A New approach for machining sculptured surfaces by curvature catering method. Chinese Journal of Mechanical Engineering, 5(3): pp.221-230.
4 **Fritz R.** (1993) Collision-free five-axis milling of twisted surfaces, Annals of the CIRP, Vol.42, pp.457-461.
5 **Lee, Y. S.** (1997) Admissible tool orientation control of gouging avoidance for 5-axis complex surface machining. Computer-Aided Design.Vol.29, No.7, pp.507-521.
6 **Elber, G., Fish, R.** (1997) 5-Axis freeform surface milling using piecewise ruled surface approximation. Journal of Manufacturing Science and Engineering. Vol.119, pp.383-387.
7 **Lee, Y. S., Bahattin Koc.** (1998) Ellipse-offset approach and inclined zig-zag method for multi-axis roughing of ruled surface pockets. Computer-Aided Design.Vol.30, No.12, pp.957-971.
8 **Monies F., et al.** (2000) Improved positioning of a conical mill for machining ruled surfaces: application to turbine blades, Proc. Instn. Mech. Engrs., Vol.214,pp.625-634.
9 **Ni Y. R., Ma D. Z. et al.** (2001) Optimal orientation control for torus tool 5-axis sculptured surface NC machining. Chinese Journal of Mechanical Engineering, Vol.37, No.2, pp.87-91.
10 **Sasaki, S.** (1956) Differential Geometry (in Japanese) .Kyoritsu Press,Tokyo.
11 **Tu B. X.** (1987) Higher Algebra, ShangHai Science and Technology Press.

Modelling cutter swept angle at cornering cut

H S CHOY and **K W CHAN**
Department of Mechanical Engineering, The University of Hong Kong, Hong Kong

ABSTRACT

When milling concave corners, cutter load increases momentarily and fluctuates severely due to concentration and uneven distribution of material stock. This abrupt change of cutter load produces undesirable machining results such as wavy machined surface and cutter breakage. An important factor for studying cutter load in 2.5D pocket milling is the instantaneous Radial Depth of Cut (RDC). However, previous work on RDC under different corner-cutting conditions is lacking.

In this paper, we overview typical work done by other researchers on cornering cut, followed by presenting our study on RDC for different corner shapes. In our work, we express RDC mathematically in terms of the instantaneous cutter engage angle which is defined as Cutter Swept Angle (CSA). An analytical approach for modeling CSA is explained. Finally, examples are shown to demonstrate that the proposed CSA modeling method can give an accurate prediction of cutter load pattern at cornering cut.

1 INTRODUCTION

Pocket milling is a metal removal operation commonly used for creating depressions in machined parts. Generating tool path for milling a pocket begins by slicing the pocket into a number of horizontal layers. The layer gap width represents the incremental depth of cut along the spindle or cutter axis. The pocket boundary of each layer is revealed by evaluating the curves formed between the intersection of the dissecting plane and the pocket's wall faces. Using the determined pocket boundary of each layer, different tool path patterns can then be employed to remove the stock material within the pocket region.
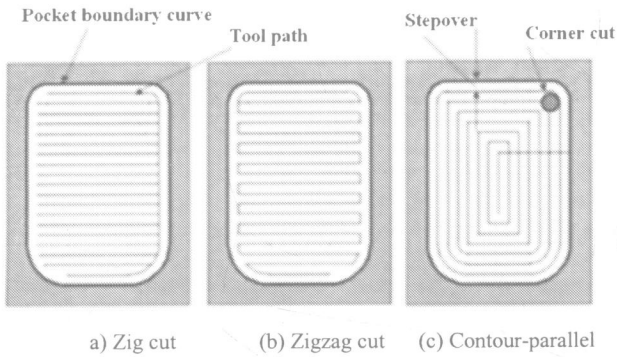
a) Zig cut   (b) Zigzag cut   (c) Contour-parallel

Figure 1. Three commonly used tool path patterns

Popular pocket milling tool path patterns provided by contemporary CAM systems are zig, zig-zag and contour-parallel offset (CPO) as illustrated in Figure 1. Among these three patterns, CPO tool path is most widely adopted because it produces lesser idle tool path portions and can maintain a consistent use of down-cut (or up-cut) milling method. However, CPO tool path inherently produces many convex and concave corners.
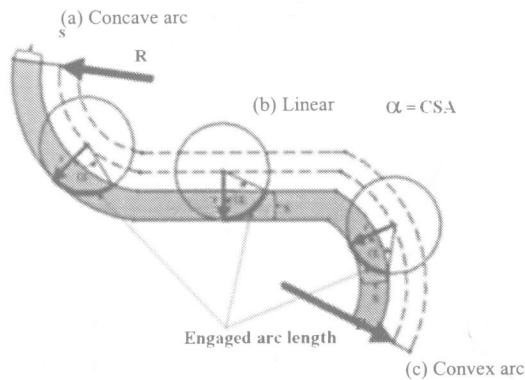
Figure 2. Cutting path classification

By referring to Figure 2, it can be easily observed that when a cutter moves from a linear tool path segment into a convex corner, the cutter engage arc length effectively decreases. Since cutter load or chip load is directly related with cutter engage arc length, this means that cutter load will drop when performing convex cornering cut. This also implies that convex corners create less problem to machining since the change of chip load involved is of a decreasing nature. On the contrary, when a cutter moves from a linear tool path segment into a concave corner, the cutter engage arc length rises quickly, achieving a maximum value at the middle part of the corner, and then decreases as the cutter leaves the corner. Such a sharp rise and fall of cutter engage arc length at concave corners will lead to undesirable results such as machine vibration, chatter marks on machined surface and even tool breakage when the cutter load is excessive. To avoid these adverse consequences, machining practitioners usually retreat to using a lower cutting feedrate and or depth of cut which represents a reduction of machining

efficiency. Maintaining a high machining efficiency, however, is of paramount importance for increasing business competitiveness amidst today's stringent market conditions. Motivated by these background reasons, we therefore conducted an in-depth investigation on concave cornering cut and reported our findings in this paper.

The remainder of this paper is organized as follow. Section 2 overviews previous work on cornering cut. Section 3 describes CSA for different corner cases in detail. The formation of different CSA zones is explained with illustrations and the determination of the CSA values in different zones is expressed mathematically. Examples are given and compared for different approaches in sections 4. Section 5 concludes our work.

2 REVIEW OF PREVIOUS WORK

In earlier work, Kline et al. [1] used a mechanistic model to estimate chip load and found that there was a dramatic change in cutting force at cornering cut. Iwabe et al. [2] established a geometry analysis of the interaction between an end mill cutter and an inside corner. Tlusty et al. [3] proved that the change in radial depth of cut at corner had an adverse effect on machining stability such as causing high frequency chatters.

Two major approaches can be identified for tackling cornering cut problem, namely the adaptive control and the modification of tool path geometry. The former approach focuses on controlling the cutting performance by adjusting the cutting parameters instanteously during cutting. For instance, Tarng et al. [4] attempted to maintain a constant metal removal rate in pocket milling by adjusting feedrate adaptively. Spence et al. [5] scheduled the feedrate automatically so as to satisfy force, torque, and dimensional error constraints. Fussell et al. [6] used a feed rate planning system to select feed rate for 3-axis sculpture milling by integrating a Z-buffer geometric model with a discrete mechanistic model of the cutter. A pre-requisite for applying this approach is a sophisticated NC machine that posses NC program look-head function for advanced calculation of cutting conditions and a rapid acceleration/deceleration control mechanism to response to the frequent and quick change of feedrate.

The latter approach aims to adjust the chip load by using special tool path trajectory. For example, Tsai et al. [7] modified CPO tool path segments in order to reduce corner cutting problems such as chatter vibration and excessive machining errors. Stori et al. [8] presented a constant cutter engagement tool path for reducing cutting force variation at corner cutting. However, their spiraling tool path can only be applied to a limited set of corner shapes. Hinduja et al. [9] applied 2-D union operation to combine the area swept by the bottom face of an end mill and the remaining stock area left over by the previous machining path for finding RDC variation. More recently, they [10] also reported the application of RDC variation to selecting optimum cutter diameter for pocket machining. However, their approximate approach for finding the instantaneous RDC cannot handle more complicated corner shapes that were defined by Choy et al. [11].

The static cutter engage angle along line and arc segments was formally defined by Kramer [12]. However, the continuous dynamic change of cutter engage angle at concave corners has not been addressed in his work. Instead of simply using the static cutter engage angle or RDC, we consider that the instantaneous cutter swept angle CSA is a more appropriate parameter to describe the corner cutting condition because CSA can be used to describe the different stages

of cutter engage angle change during a cornering cut. We are also not aware of any previous research work done on CSA for complicated concave corners that are formed by different boundary geometries. Our work therefore focuses on modeling CSA at cornering cut for complicated corner shapes.

3 ANALYTICAL REPRESENTATION OF CSA

We focus our consideration on pocket milling tool paths that are formed by line, concave and convex arc segments. Referring to Figure 2, the governing equations of CSA, α, for these three basic tool path geometries are:

Linear	$\cos(\alpha) = 1 - s/r$	(1)
Concave arc	$\cos(\alpha) = 1 - s/r - \Gamma$	(2)
Convex arc	$\cos(\alpha) = 1 - s/r + \Gamma$	(3)

where $\Gamma = s(r - 0.5s)/(Rr)$, and s, r and R are the radial depth of cut, cutter radius and circular cutting path radius respectively. It can be observed that $\Gamma \geq 0$ since $2r \geq s$, and α decreases as s decreases.

Moreover, different cutting modes can be discerned by the following rules:
If CSA = 180°, the cutter is performing a slot cutting operation,
If 90 < CSA < 180°, the cutter is subject to both up-cutting force on one side and down-cutting force on another side,
If CSA ≤ 90°, the cutter is performing up-cutting when the cutter rotation vector is opposite to the feeding direction vector of the workpiece or down-cutting when the cutter rotation vector is in line with the feeding direction vector of the workpiece.

An arc segment can be classified as either clockwise (CW) or counterclockwise (CCW) depending on its traversal direction with respect to the starting point S and ending point E (Figure 3). Based on the possible connection of the tool path entity and their traversal order of the three basic types of tool path geometry (Figure 3), nine different corner types can be formed as shown in Figure 4.
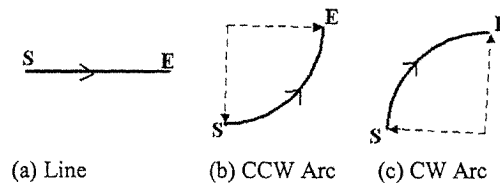
(a) Line   (b) CCW Arc   (c) CW Arc
Figure 3. Tool path geometries

GEOMETRY TYPE	LINE	CCW ARC	CW ARC
LINE			
CCW ARC			
CW ARC			

Figure 4. Different corner types

As mentioned in section 1 and judged from equation 3, convex corners do not introduce drastic rise of chip load problem. Hence only concave corners are considered. Besides corner shape, the geometry of the previous and current tool paths also affects the corner cutting conditions significantly because the remaining stock material is governed by the geometry of the previous tool path. We therefore studied the following three tool path cases for deriving the expression of CSA in concave corner region.

3.1 Classification of corner cases

Based on the geometric configuration of the previous and current tool centre paths, three different corner cases can be identified:

(i) Case 1   (ii) Case 2   (iii) Case 3
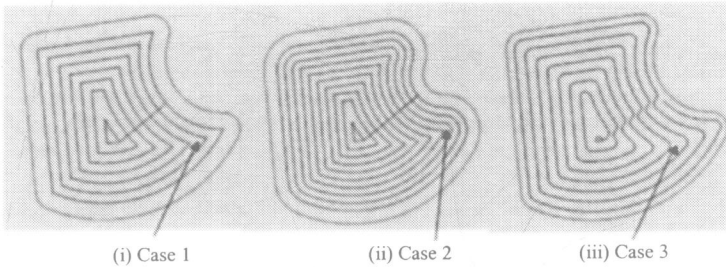
Figure 5. Possible corner cases

Case 1: Both previous and current tool centre paths have no joining fillet at corner (Figure 5 (i)),

Case 2: Previous tool centre path has no joining fillet but current tool centre path has joining fillet at corner (Figure 5 (ii)), and

Case 3: Both previous and current tool centre paths have joining fillets at corner (Figure 5 (iii)).

(i) Case 1   (ii) Case 2   (iii) Case 3

Figure 6. Nine possible corner shapes for the three cases

Case 1: This case occurs when the cutter radius is smaller than or equal to the corner radius of the pocket. Nine possible corner shapes for this case are shown in Figure 6 (i). For explanation purpose, the case shown in Figure 6 (i) (a) is used. From the enlarged view shown in 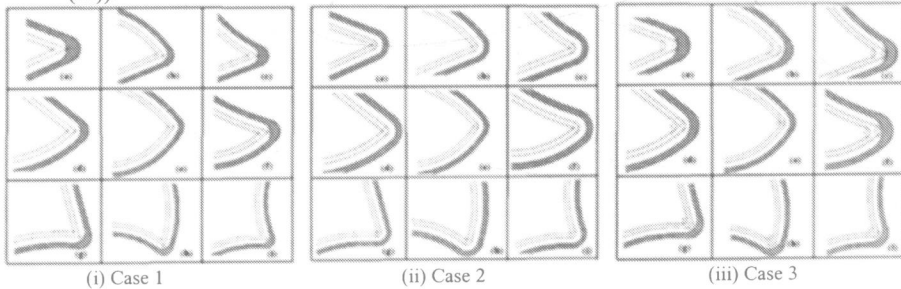Figure 7, it can be seen that the previous tool centre path is indicated by segments AB and BC while the current tool centre path is shown by segments DE and EF. Segments GH, HI and IJ represent the remaining stock boundary swept by the cutter when the cutter centre moves along the previous tool center path. Two intermediate points of interest, D_1 and E_1, are introduced in the figure. Point D_1 lies on DE and represents the particular cutter centre position when the cutter boundary intersects point H which is created by a perpendicular projection from point B to GH. Similarly, point E_1 lies on EF such that the cutter boundary intersects point I which is formed by projecting a line perpendicularly from point B to IJ.

Case 2: This case occurs when the cutter radius is much smaller than the corner radius of the pocket. Nine possible corner shapes for this case are displayed in Figure 6 (ii). The case shown in Figure 6 (ii) (a) is enlarged in Figure 8 for the following explanation. Segments AB and BC represent the previous tool centre path while segments DE, EF and FG display the current tool centre path. Segments HI, IJ and JK represent the remaining stock boundary swept by the cutter when the cutter centre moves along the previous tool center path. Point D_1 lies on DE and represents the particular cutter center position when the cutter boundary intersects point I. Point I is made by the perpendicular projection from point B to HI. Similarly, point E_1 lies on EF such that the cutter boundary intersects point J which is made by a line normally projected from point B to JK.

Case 3: This case occurs when both the previous and current tool centre paths have a specified fillet radius. Nine possible corner shapes for this case are depicted in Figure 6 (iii). Figure 9 shows the enlarged view of the first case shown in Figure 6 (iii) (a) for explanation purpose. Segments AB, BC and CD show the previous tool centre paths while segments EF, FG and GH illustrate the current tool centre paths. Segments IJ, JK and KL represent the remaining stock boundary swept by the cutter when the cutter centre moves along the previous tool center path. Point E_1 lies on EF and indicates the particular cutter centre position when the cutter boundary intersects point J. Point J is made by the perpendicular projection from point O to IJ. Similarly, point G_1 lies on GH and represents the cutter centre position when the cutter boundary intersects point K which is created by projecting a perpendicular line from point O to KL.

3.2 Defining different CSA zones
According to the identified corner cases, the cutting stages for each corner case were scrutinized and different zones of CSA for the three exemplary corner cases were defined below.
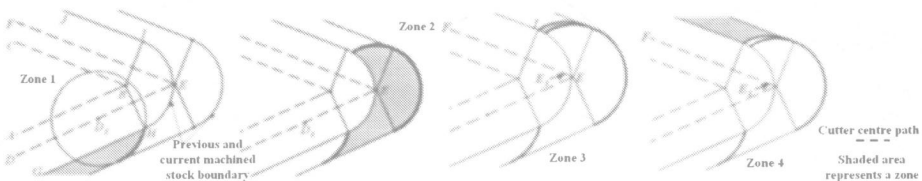

Figure 7. Different CSA zones for case 1

Case 1: Four CSA zones as shown by the shaded regions in Figure 7 can be defined and summarized in Table 1.

Table 1. Scrutiny of CSA zones for case 1

Zone	Description	Value of CSA	Determination of CSA value
1	Corresponds to the region machined when the cutter centre moves from points D to D_1	Constant	If DD_1 is a line, determine by equation (1), else if DD_1 is a CCW arc, determine by equation (2), else DD_1 is a CW arc, determine by equation (3).
2	Corresponds to the region machined when the cutter centre moves from points D_1 to E.	Varying	By the special procedure described in section 3.3.1
3	Corresponds to the region machined when the cutter centre moves from points E to E_1	Varying	By the special procedure described in section 3.3.1
4	Corresponds to the region machined when the cutter centre moves from points E_1 to F.	Constant	If E_1F is a line, determine by equation (1), else if E_1F is a CCW arc, determine by equation (2), else E_1F is a CW arc, determine by equation (3).

Zone 1   Zone 2   Zone 3   Zone 4   Zone 5

Figure 8. Different CSA zones for case 2

Case 2: Five CSA zones as shown by the shaded regions in Figure 8 can be defined and summarized in Table 2.

Table 2. Scrutiny of CSA zones for case 2

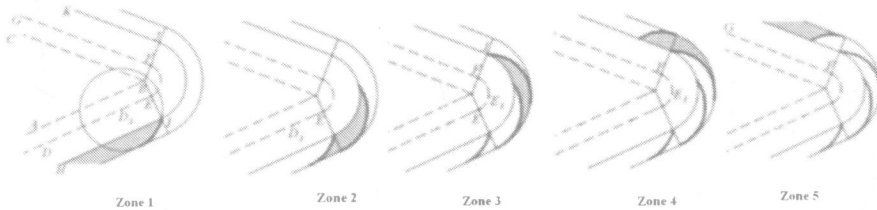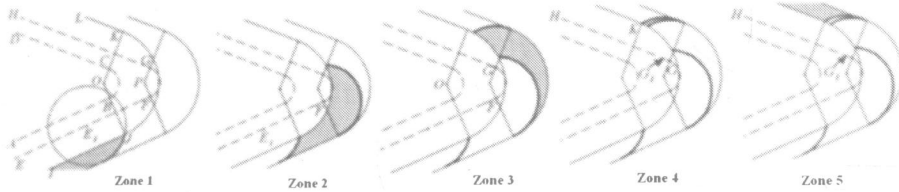| Zone	Description	Value of CSA	Determination of CSA value
1	Corresponds to the region machined when the cutter centre moves from points D to D_1	constant	If DD_1 is a line, determine by equation (1), else if DD_1 is a CCW arc, determine by equation (2), else DD_1 is a CW arc, determine by equation (3).
2	Corresponds to the region machined when the cutter centre moves from points D_1 to E.	varying	By the special procedure described in section 3.3.1
3	Corresponds to the region machined when the cutter centre moves from points E to E_1	constant	By equation (2) because the cutter is basically cutting along a CCW arc.
4	Corresponds to the region machined when the cutter centre moves from points E_1 to F.	varying	If JK is a line, determine by the special procedure described in section 3.3.2. Otherwise, determine by the special procedure described in section 3.3.1.
5	Corresponds to the region machined when the cutter centre moves from points F to G.	constant	If FG is a line, determine by equation (1), else if FG is a CCW arc, determine by equation (2), else FG is a CW arc, determine by equation (3).

Figure 9. Different CSA zones for case 3

Case 3: Five CSA zones as shown by the shaded regions in Figure 9 can be defined and summarized in Table 3.

Table 3. Scrutiny of CSA zones for case 3

Zone	Description	Value of CSA	Determination of CSA value
1	Corresponds to the region machined when the cutter centre moves from points E to E_1	constant	If EE_1 is a line, determine by equation (1), else if EE_1 is a CCW arc, determine by equation (2), else EE_1 is a CW arc, determine by equation (3).
2	Corresponds to the region machined when the cutter centre moves from points E_1 to F.	varying	By the special procedure described in section 3.3.1
3	Corresponds to the region machined when the cutter centre moves from points F to G.	varying	By the special procedure described in section 3.3.1
4	Corresponds to the region machined when the cutter centre moves from points G to G_1.	varying	By the special procedure described in section 3.3.1
5	Corresponds to the region machined when the cutter centre moves from points G_1 to H.	constant	If G_1H is a line, determine by equation (1), else if G_1H is a CCW arc, determine by equation (2), else G_1H is a CW arc, determine by equation (3).

3.3 Special CSA calculation procedures

Determination of the CSA values for some portions of the corner tool path segments mentioned above requires more detailed analysis.

Let point I_1 be the intersection point made between the cutter boundary and the currently machined stock boundary as shown in Figure 10, and point I_2 be the intersection point made between the cutter boundary and the previously machined stock boundary. It can be observed that if points I_1 and I_2 are found, CSA can be determined simply by using cosine rule.
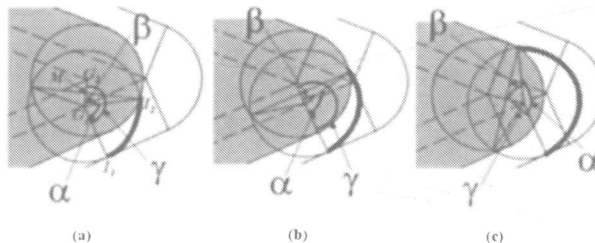

Figure 10. CSA for two circle case

Point I_1 can be determined easily by projecting a line from the current cutter center location G_1, normal to the current tool centre path segment and intersecting the currently machined stock boundary.

However, since point I_2 can be formed by two possible cases: 1) intersection between two circles (i.e. current cutter boundary intersects an arc segment in previously machined stock boundary) or 2) intersection between a line and a circle (i.e. current cutter boundary intersects a line segment in previously machined stock boundary), its determination requires the following two different sets of procedures.

3.3.1 Intersection of two circles case

Consider two circles shown in Figures 10 and 11 with centers G_1 and G_2 and radii r_1 and r_2 respectively. Let M be the intersection point made between the line G_1G_2 and the common chord of the two circles.

Let γ be the included angle between line G_1I_1 and line G_1G_2, and β be the included angle between line G_1I_2 and line G_1G_2. The CSA, represented by α, can be determined by

Case A (Figure 10 (a)): If G_1G_2 has not been parallel to G_1I_1 when the cutter entering the corner, $\alpha = \gamma - \beta$.

Case B (Figure 10 (b)): If G_1G_2 is parallel to G_1I_1 when the cutter entering the corner, $\gamma = 180$ degrees and $\alpha = \gamma - \beta$.

Case C (Figure 10 (c)): If G_1G_2 has once been parallel to G_1I_1 when the cutter entering the corner, $\alpha = 360° - (\gamma + \beta)$.

γ is determined by the equation:
$\cos(\gamma) = <G_1I_1 \cdot G_1G_2> / (|G_1I_1| |G_1G_2|)$ where $<>$ and $||$ are dot product and magnitude operator respectively. G_1I_1 is the line vector from point G_1 to point I_1. Similarly, G_1G_2 is the line vector from point G_1 to point G_2.

β can be determined by the following rule
If $r_1 = r_2$ (this means the two circles have equal diameters)
then $\cos(\beta) = |MG_1| / r_1$ (or $\cos(\beta) = |MG_2| / r_2$)
where point M lies on the mid-point of G_1G_2.
else if $r_1 \neq r_2$ (this means the two circles have unequal diameters as shown in Figure 11)
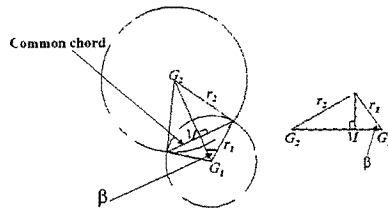then $\cos(\beta) = \{(G_1G_2)^2 + (r_1)^2 - (r_2)^2\} / 2(G_1G_2)(r_1)$

Figure 11. Circles with different radii
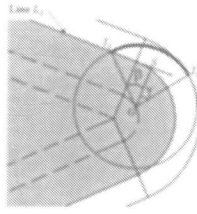
Figure 12. CSA for a line and a circle case

3.3.2 Intersection of a line and a circle case

Consider a circle of radius r with center G (x_c, y_c) and a line L_2 with equation a.x + b.y + c = 0 shown in Figure 12. Let point S be the closest distance point between circle center G and line L_2, β be included angle between line GI_2 and line GS, γ be the included angle between line GS and GI_1. Hence, it can be observed that $\alpha = \gamma + \beta$.

β can be determined by cos $(\beta) = |GS| / r$ where $\|$ is the magnitude operator and $|GS|$ = the absolute value of $((a.x_c + b.y_c + c)/\sqrt{a^2 + b^2})$, a, b and c are the coefficients of line L_2. The calculation of distance GS is obtained by considering the normal distance of the point G from the line L_2.

γ can be determined by tan $(\gamma) = (m_2 - m_1)/(1 + m_1.m_2)$ where m_1 is the slope of the line I_1G and m_2 is the slope of a line normal to line L_2.

If the calculated α in sections 3.3.1. or 3.3.2. is greater than 180 degrees, α will be automatically set to 180 degrees since this is not the maximum possible CSA, indicating that the cutter is performing a slot cutting operation.

3.4 Plotting the change of CSA

The calculated CSA values for the exemplary corners shown in Figures 7, 8 and 9 are plotted graphically in Figures 13 (a), (b) and (c) respectively. The corner angle and radial depth of cut for Figures 7, 8 and 9 are the same. The corner angle and cutter diameter are 46 degrees and 10mm respectively. The radial depth of cut and fillet radius used are both 2mm.

4 DISCUSSION OF RESULTS

Figure 13. Examples for modeling the cutter swept angle for different corner cases

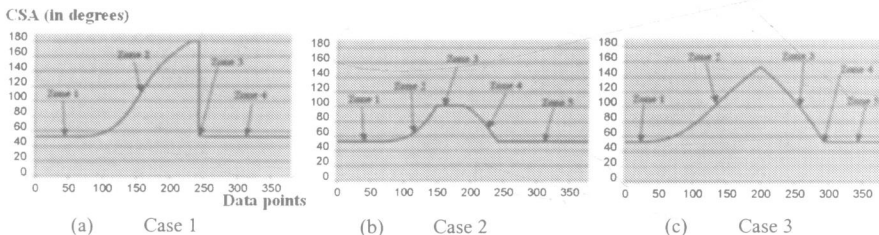It can be seen from Figure 13 (a) that as the cutter moves from the linear tool path into the corner, the CSA rises very rapidly, reaching a maximum of 180° at the middle part of the corner, and then falls off vertically as the cutter leaves the corner. This vividly shows that without round fillets added to the tool centre paths, the chip load at cornering cut will increase momentarily as indicated by the sharp "spike" shape of the CSA plot. This will produce adverse machining results and even tool breakage especially when performing high speed machining.

Figure 13 (b) represents the case that a circular fillet has been introduced to the current tool centre path. Since there is no fillet in the previous tool centre path, the stock material left behind for the current filleted tool path to remove is effectively less. Hence, the rise of CSA in zone 2 is less rapid. Most importantly, there is a stable zone 3 in which the maximum CSA value is significantly reduced to about 100°. The decrease of the CSA value in zone 4 is also more gentle. Comparing with the plot in Figure 13 (a), it can be seen that the middle part shape of the CSA plot in Figure 13 (b) has been stretched horizontally. This indicates a beneficial effect as it implies that the chip load during corner cutting can be suppressed and maintained more steady in this case.

It can be observed from Figure 13 (c) that CSA rises steadily to a maximum of about 160° and then slides down gently without exhibiting a constant CSA zone. The overall CSA plot is stretched horizontally. In comparison, the maximum CSA obtained in Figure 13 (c) is greater than that in Figure 13 (b). It is because, after joining a fillet in the previous tool center path, the stock material left behind for the current filleted tool path to remove becomes greater.

The plotted results for these three exemplary cases cannot fully represent the other cases because of the following four factors. Firstly, different tool path geometries can alter the shape of the initial and final zones of the figures. Secondly, different tool path stepover distance can shift the CSA curve upwards or downwards. Thirdly, different specified fillet radius at the tool centre path corner can narrow or widen the middle part of the CSA curve. Fourthly, different corner angles can cause different extent of material accumulation. For example, a corner of acute angle has more corner material than that of an obtuse angle.

5 CONCLUSIONS

The issue of drastic change of cutter resistance when milling concave corners is addressed. The instantaneous cutter swept angle CSA is considered as a suitable parameter for studying chip load. A classification based on a combination of different corner boundary geometries and tool centre path geometries is established. Based on this classification, a detailed analysis of CSA at different intermediate stages of corner cutting is conducted. The mathematical equations for evaluating CSA in different cutting zones are also deduced.

Graphical plotting of the CSA values calculated by the derived equations can be used to explain the severe fluctuation of cutting forces in cornering cut. Cutting force measuring experiments will be conducted to verify the correlation between the CSA and the actual cutting resistance incurred. The established CSA modeling scheme will be an important tool for supporting investigation of the problems of cornering cut.

REFERENCES

1 **Kline, W. A., Devor, R. E. and Lindberg, J.** (1982) Prediction of cutting forces in end milling with application to cornering cuts. International Journal of Machine Tool Design and Research, 22, pp. 7-22.

2 **Iwabe, H., Fujii, Y., Saito, K. and Kishinami, T.** (1989) Study on corner cut by end mill – analysis of cutting mechanism and new cutting method at inside corner. In Japanese, Journal of Japan Society of Precision Engineering, 55(5), pp. 841 – 846.

3 **Tlusty, J., Smith, S. and Zamudio, C.** (1990) New NC routines for quality in milling. Annals of the CIRP, 39(1), pp. 517 – 521.

4 **Tarng, Y. S. and Shyur, Y. Y.** (1993) Identification of radial depth of cut in numerical control pocketing routines. International journal of Machine Tools & Manufacture, 33(1), pp. 1 – 11.

5 **Spence, A. D. and Altintas, Y.** (1994) A solid modeller based milling process simulation and planning system. Journal of Engineering for Industry, 116, pp. 61-69.

6 **Fussell, B. K., Jerard, R. B. and Hemmett, J. G.** (2001) Robust feedrate selection for 3-Axis NC machining using discrete models. Journal of Manufacturing Science and Engineering, Transactions of the ASME, 123, pp. 214 – 224.

7 **Tsai, M. D., Takata, S., Inui, M., Kimura, F. and Sata, T.** (1991) Operation planning based on cutting process models. Annals of CIRP, 40(1), pp. 95 – 98.

8 **Stori, J.A. and Wright, P.K.** (2000) Constant engagement tool-path generation for convex geometries. Journal of Manufacturing Systems, 19(3), pp. 172-184.

9 **Hinduja, S., Ma, Y.S. and Barrow, G.** (1995) Determination of radial width of cut and cutting modes in milling. International Journal of Machine Tools & Manufacture, 35(1), pp. 689 – 699.

10 **Hinduja, S., Roaydi, A., Philimis, P. and Barrow, G.** (2001) Determination of optimum cutter diameter for machining 2-1/2 D pockets. International Journal of Machine Tools & Manufacture, 41, pp. 687 – 702.

11 **Choy, H. S. and Chan, K. W.** (2002) Enhanced strategy for milling corners. accepted for publication in Journal of Engineering Manufacture.

12 **Kramer, T. R.** (1992) Pocket milling with tool engagement detection. Journal of Manufacturing Systems, 11(2), pp. 114 – 123.

Analytical approach for selection of optimal feedrate in efficient machining of complex surfaces

G VIKRAM, P HARSHA, and N R BABU
Department of Mechanical Engineering, Indian Institute of Technology, Chennai, India

ABSTRACT

The present work deals with an analytical approach that can suggest feedrate variations over the contour of complex surfaces while finish machining with ball nose end mill. In selecting the feedrate, this approach considers the material removal rate (MRR) and tool life, apart from the consistent dimensional accuracy over the contour. The effectiveness of the proposed approach is demonstrated by comparing its results with those obtained with the approach proposed by Ip [1] and other criteria. Finally, the suitability of the proposed approach for efficient machining of complex surfaces is highlighted.

1 INTRODUCTION

The process of generation of complex yet precise contours using multi axis computer numerically controlled (CNC) machines is an important topic of study. The complex surfaces can be produced from any given shape by forming the part shape with the removal of excess stock in a series of passes of cutter, i.e. rough machining. Once rough machining forms the part shape, the final shape is produced by means of a finishing pass. The accuracy of the finished surface and the efficiency of the process depend on the inherent characteristics of CNC machine and the process parameters chosen.

Machining of complex surfaces using CAD/CAM systems requires CAD system to discretize the contour to be machined, which converts complex curves into simple linear or curved segments. The path to be followed by the cutter is given as an input to the CNC part program. Since it is a discretized version of the actual path to be followed, it introduces a certain error, known as discretization error. The nature of movement of tool in discrete steps from one point to another, i.e. linear or circular interpolation, introduces command errors. Incapability

of CNC system to instantaneously provide feedback regarding discrepancies in position and velocity introduces tracking errors [2,3,4]. The aforementioned errors in the system manifest themselves on the manufactured part as contour, orientation and feedrate errors [3].

Contour error may be defined as the distance between the curve and its linear approximation in the region under consideration [3,5,6]. Orientation error is the deviation of the tool axis from the surface normal [7,8,9]. Feedrate error is defined as the deviation of the actual feedrate from desired feedrate in machining the contour of a surface [9]. Among these errors, contour and orientation errors are a result of discretisation and command errors in the system, which are specific to the process of discretization adopted and interpolation capability of the controller respectively. Orientation errors are applicable only to multi axis CNC machining. The feedrate error arises due to the feedrate variation over different steps on the contour and also due to the incapability of the CNC system to respond to the variation in feedrate. This error is dependent on the process parameters chosen in directing the cutter over the contour. Both contour and feedrate errors result in dimensional inaccuracies of the machined part and hence have to be minimized. As mentioned above, the contour errors are specific to the process and the controller in consideration and hence are fixed. Different process parameters such as depth of cut (d.o.c), width of cut (w.o.c), cutting speed and feedrate affect the tracking errors, which manifest as feedrate errors. However, the feedrate errors due to tracking errors are generally small in comparison to feedrate errors arising due to discretization of feedrate over the surface to be machined.

Another important aspect of machining a complex surface on a CNC machine is efficiency and productivity, which is dependent on the choice of process parameters like d.o.c and w.o.c, cutting speed, feedrate. In case of finish machining with ball nose end mill, d.o.c and w.o.c are relatively small and the spindle speed is normally kept constant. Though the spindle speed is kept constant, the cutting velocity varies due to the variation of effective diameter as the point of contact of the ball nosed end mill with the surface changes continuously while finishing the complex contours. This results in variable MRR that can cause variations in cutting forces and influence the life of cutting tool. Ip [1] suggested a fuzzy based approach for selecting optimal feedrate variation in order to optimize MRR, while maintaining the cutting force constant throughout the machining process. No doubt this approach can ensure constant cutting force while machining the surface with variable feedrates. But it was noticed that this feedrate variation does not guarantee the optimal use of tool in certain segments of the contour. Moreover, this approach did not consider the errors that normally occur due to the variation in feedrate. These two aspects such as optimal utilization of tool and the errors due to feedrate variation are of considerable importance in order to achieve high efficiency, accurate complex surface machining with CNC machine tools.

In view of the above, the present work proposes a new approach that can suggest the feedrate variation to be adopted over the complex contours, in order to minimize the feedrate error and to maximize MRR and utilization of cutting tool.

2 THEORY

In complex surface machining on 3-axis CNC machine tools using a ball nosed end mill, the contour is discretized into a series of linear or circular segments. Consider a curve C shown in Figure1. Let the position of the tip of the tool be represented by the position curve $r(\xi)$, the speed of the tip of the tool along the tangent of C be represented by the feed curve $f(\xi)$. The

scalar ξ is a common parameter for position and feed curve. As shown in Figure 2, e_{ci} represents the contour error in i^{th} interval, which is the maximum distance between $r(\xi)$ and its linear approximation. e_{fi} represents the feedrate error in the ith interval, which is the maximum deviation of discretized feedrate from $f(\xi)$. The upper bounds on these errors in terms of the interval size are given by [3]

$$e_{ci} \quad 1/8 \, \Delta\xi_i^2 \sup \quad r''(\xi_i) \tag{1}$$

$$e_{fi} \quad 1/8 \, \Delta\xi_i^2 \sup \quad f''(\xi_i) \tag{2}$$

where sup represents the suprimum of the value.

In this analysis, the discretization of the curve was chosen based on a constant contour error. The given complex surface is divided into linear segments in such a way that the contour error is always less than the given limit e_{clim}. From the equation (1), the length of the segment for this discretization is given as

$$\Delta\xi_i = \sqrt{(8 \, e_{clim} / \sup \quad r''(\xi_i) \quad)} \tag{3}$$

While machining the discretized segment, using a ball nosed end mill, the point of contact with the surface does not change in that particular segment. However, as shown in figure 3, the point of contact changes from one discretized segment to another, which in turn changes the cutting velocity due to the change in effective diameter. The effective cutting velocity is given as [1]

$$V_e = \pi D_e N \tag{4}$$

where the effective diameter is a function of the local surface gradient and is given by

$$D_e = D \sin\theta \tag{5}$$

The change in cutting velocity over the contour affects the extent of tool life. According to the Taylor tool life equation, the tool life is given in terms of cutting velocity (V_e), feedrate (f) and axial (h) and radial depth of cut, (w) as [10]

$$\text{Tool life} = \frac{K}{V_e^{1/a} \, f^{1/b} \, h^{1/c} \, w^{1/d}} \tag{6}$$

The exponents a, b, c and d depend on the tool and work material. Tool utilization, which is defined as the fraction of tool life spent to machine a given component, is of more significance than the tool life itself. Tool utilization can aid in determining the number of components that can be produced with the tool life of the cutter. Smaller the tool utilisation in producing a part, greater will be the efficiency of cutting process. For a segment with a constant gradient, the tool utilization is given in terms of fraction of tool life spent for machining a discretized segment of length $| r(\xi_i) - r(\xi_{i-1}) |$

$$\text{Fraction of tool life spent} = \frac{| r(\xi_i) - r(\xi_{i-1}) | \, V_e^{1/a} \, f^{1/(b-1)} \, h^{1/c} \, w^{1/d}}{K \, N \, m} \tag{7}$$

where m represents the no of cutting teeth.

From the equation (7), it can be seen that the fraction of tool life spent is also a function of d.o.c, w.o.c, spindle speed and feedrate. The d.o.c, w.o.c and spindle speeds are usually kept constant during the machining. However, the feedrate can be changed in order to minimize the fraction of tool life spent. But, the feedrate variation also results in variation of MRR and the machining time. Both MRR and the machining time for machining a discretized segment of length $|r(\xi_i) - r(\xi_{i-1})|$ are given by

$$MRR = m\ h\ w\ f\ N \qquad (8)$$

$$\text{Machining time} = \frac{r(\xi_i) - r(\xi_{i-1})}{m\ f\ N} \qquad (9)$$

From the above discussion, it is evident that feedrate is an important parameter which can be varied over the contour to be machined in order to obtain optimal MRR and tool utilisation for high efficiency and high productivity. However, the variation of feedrate will lead to feedrate error, which can affect the dimensional accuracy over the machined surface [4]. In order to maintain a good dimensional accuracy over the contour and achieve optimal MRR and tool utilisation, an attempt is made to optimise feedrate variation by means of formulating a multi-objective function that considers all three objectives with certain weightage to each of them.

3 MULTI-OBJECTIVE FUNCTION

In the proposed approach, the multi-objective function consists of three terms involving the feedrate error arising out of feedrate variation, the fraction of tool life spent and the machining time.

Let the feedrate error produced in the ith interval due to the linear discretization of the feedrate be represented by $E(f_i)$. Then from equation (2)

$$E(f_i) = 1/8\ \Delta\xi_i^2\ \sup\ f''(\xi_i)$$

where $\Delta\xi_i$ is chosen based on the constant contour error. If e_{clim} is the permissible contour error, then $\Delta\xi_i$ is given by

$$\Delta\xi_i = \sqrt{(8\ e_{clim}\ /\ r''(\xi_i))}$$

The interval size is small enough that $f''(\xi_i)$ either increases or decreases monotonically. If $f''(\xi_i)$ is monotonically increasing, $E(f_i)$ will reduce to

$$E(f_i) = 1/8\ \Delta\xi_i^2\ f''(\xi_i) \qquad (10)$$

If $f''(\xi_i)$ is monotonically decreasing, $E(f_i)$ will reduce to

$$E(f_i) = 1/8\ \Delta\xi_i^2\ f''(\xi_{i-1}) \qquad (11)$$

Let the fraction of tool life spent in machining the ith discretized segment be represented by $T(f_i)$, which can be evaluated from the equation (7) by replacing the feedrate, f, with the discretized value of feedrate in the ith interval, f_i. Similarly, $M(f_i)$, the term used to represent

the time required to machine the ith discretized segment can be determined from the equation (9) by replacing f with f_i.

The order of magnitude and physical dimensions of each of the terms, which are used to represent the various objectives, are different. Hence, normalizing is adopted to remove the discrepancies caused by terms involving different physical quantities. One of the ways of normalizing is to choose the normalizing factors, which are the reciprocal of the reasonable user expected values for each of the terms. This not only eliminates the discrepancies due to different units but also non-dimensionalises the terms, thus bringing all the three terms to the same order. By using these normalizing factors, the objective function, which is used to determine the optimal value of feedrate in the ith interval, is formulated as

$$S(f_i) = N_1 E(f_i) + N_2 T(f_i) + N_3 M(f_i) \qquad (12)$$

Here N_1, N_2, N_3 are the normalizing factors for feedrate error, tool utilization and machining time respectively. Using this objective function, the optimal feedrate for every segment is evaluated by minimizing the objective function for each discretized segment under consideration. The set of all feedrates for all the segments together gives the feedrate variation over the contour under consideration.

4 ESTIMATION OF FEEDRATE VARIATION

The multi-objective function is a second order differential equation in terms of the feedrate. Substituting the expressions for $E(f_i)$, $T(f_i)$ and $M(f_i)$, in the equation (12), it can be written as

$$S(f_i) = C_1 f_i'' + C_2 f_i^{(1/b-1)} + C_3 / f_i \qquad (13)$$

where $C_1 = N_1 \Delta\xi_i / 8$

$$C_2 = \frac{N_2 \, |r(\xi_i) - r(\xi_{i-1})| \, (\pi \, D \, Sin\theta)^{1/a} \, h^{1/c} \, w^{1/d}}{K \, m \, N^{(1-1/a)}}$$

and

$$C_3 = \frac{N_3 \, |r(\xi_i)-r(\xi_{i-1})|}{m \, N}$$

The derivatives in the objective function are substituted with numerical derivatives using Newton's backward differencing, thus reducing the objective function into an iterative equation. Substituting

$$(f_i - f_{i-1}) / (\Delta\xi_i) \quad \text{for } f' \quad \text{and}$$

$$\{(f_i - f_{i-1})/\Delta\xi_i - (f_{i-1}-f_{i-2})/\Delta\xi_{i-1} \}/ \, \Delta\xi_i \quad \text{for } f''$$

in the equation (13), it reduces to

$$S(f_i) = K_1 f_i + K_2 f_i^{(1/b-1)} + K_3 / f_i + K_4 \qquad (14)$$

where K_1, K_2 and K_3, which are functions of f_{i-1}, f_{i-2}, are constants. By differentiating the objective function with respect to f_i and equating it to zero, the optimal value of the feedrate in the discretized segment under consideration can be evaluated.

$$S'(f_i) = K_1 + K_2 (1/b-1) f_i^{(1/b-2)} - K_3 / f_i^2 = 0 \qquad (15)$$

All the constants in the above equation are positive and the function is strictly increasing. Hence, $S'(f_i) = 0$ can have at the maximum only one root. Also

$$S''(f_i) = K_2 (1/b-1)(1/b-2) f_i^{(1/b-3)} + 2 K_3 / f_i^3 > 0 \text{ for all } f_i$$

Thus, the corresponding root of $S'(f_i) = 0$ will be minimum of the function $S(f_i)$ if it exists. Therefore, analytical methods of finding the minimum can be used to determine the optimal feedrate, as there exists only one minimum i.e. the global minimum. Solving the equation (15) for feedrate f_i using Newton-Raphson iteration, the optimal feedrate is determined for every discretized segment.

5 RESULTS AND DISCUSSION

In order to validate the effectiveness of the proposed approach, the analysis is carried out by considering the finish machining of mild steel hemisphere with a diameter of 50mm, using a HSS ball nosed end mill of diameter 10mm. The depth and width of cut are considered as 1mm and are maintained constant throughout the machining process. The spindle speed is selected as 636 rpm based on the work and tool material combination. The discretization of hemispherical contour was carried out considering a constant contour error of $1\mu m$. For HSS ball nosed end mill, the constants used in Taylor's tool life equation are a = 0.09, b = 0.17, c = 0.25, d = 0.25 and k = 70×10^{12} [11]. The normalizing factors, which are used to remove the discrepancies in the physical dimensions, are selected as $N_1 = 127.2$, $N_2 = 1800$, $N_3 = 180$.

In order to demonstrate the effectiveness of the proposed approach, the following criteria were chosen for estimating the feedrate variation.
Criterion 1: Based on the proposed approach that considers multiple objectives of minimizing feedrate error, machining time and maximizing the tool utilization.
Criterion 2: Based on the consideration of minimizing machining time and maximizing tool utilisation.
Criterion 3: Based on constant feedrate for machining the entire contour.
Criterion 4: Based on the approach proposed by Ip [1].

The feedrate variation obtained by using the criteria 1 and 2 are shown in Figures 4 and 5 respectively. Using the feedrate variation estimated from different criteria, the feedrate error, tool life spent and machining time for machining different discretized segments, each having different surface gradients, along the contour of the hemisphere, are evaluated. These results are presented in Tables 1, 2, 3 and 4. The feedrate errors determined analytically using the criteria 1 and 2 are shown in Figures 6 and 7 respectively. In case of the criteria 3, there is no feedrate error with a constant feedrate.

From the above results, it can be observed that there is a considerable difference in the feedrate error, tool utilization and machining time obtained using different criteria. Analysis of the results presented in Tables 1, 2, 3 and 4 shows that the criteria 1 and 2 give good tool utilisation, but at the expense of higher machining time. Table 5 presents a comparative analysis of different parameters estimated employing different criteria. It is clear that the feedrate generated with criteria 1 is found to take more machining time but with less fraction of tool life spent. In contrast to this, the constant feedrate criterion takes less machining time but the fraction of tool life spent is an order of magnitude more than that observed with criteria 1 and 2. This particular aspect is of considerable significance when we attempt to

optimize the machining performance in terms of machining time and optimal utilization of cutting tools.

In order to assess the suitability of criteria 1 and 2, the feedrate variation, which is of practical significance in implementing it during path control, is considered. In any case, one is ultimately interested in the total magnitude of error resulted due to feedrate variation. From the feedrate variation and feedrate error presented in the earlier figures (Fig.4, 5, 6, 7), it can be noticed that the feedrate variation is very high with criteria 2. At the same time, the error due to feedrate variation obtained using criteria 2 is also found to be more compared to the results obtained by using criteria 1.

By comparing the results obtained with the present approach (criterion 1) with that obtained using Ip's approach [1] (criterion 4), it is evident that the present approach ensures very good utilization of tool and dimensional accuracy. The tool life spent using the feedrate variation proposed by Ip [1] is nearly two orders of magnitude greater than that estimated with the present approach. Also the feedrate errors are found to be relatively high with Ip's approach. Though the machining time estimated using Ip's approach is smaller, the advantage is lost as the tool change time increases by 50 times due to very high fraction of tool life spent in machining the contour.   Thus, the above analysis clearly indicates the effectiveness of the proposed approach in generating suitable feedrates over different segments of the contour while maintaining the good dimensional accuracy in terms of controlled feedrate errors and permitting the optimal utilization of tool and MRR.

6 SUMMARY

The proposed analytical approach considers the variation in the surface gradient for estimating the feedrate over different segments of the complex contour with a view to maximize the tool utilization, MRR and minimize the error due to feedrate variation. For this purpose, it makes use of a multi objective optimization approach. The effectiveness of the proposed approach for selecting the feedrate for efficient machining of complex contour was illustrated in comparison with the other criteria, including the approach proposed by Ip [1]. The merits of this approach were especially seen in minimizing the errors due to feedrate variation and maximizing the tool utilization substantially, thus enhancing the efficiency of machining with a ball end mill.

As the approach considered finish machining of contours, the influence of cutting forces on the tool was neglected. However, this needs to be addressed when the same approach is to be applied for generating the feedrate for semi-finishing of contours, where cutting forces are significantly large.

7 REFERENCES

1. **Ip, W. L. R.,** (1998) A Fuzzy Basis Material Removal Optimization Strategy for Sculptured Surface Machining using Ball-Nosed Cutters. In *International Journal of Production Research*, Vol. 36, No.9, pp.2553-2571.
2. **Koren, Y.,** (1983) *Computer Control of Manufacturing Systems*, McGraw-Hill, NewYork.
3. **Radha Sarma,** and **Aarthi Rao,**(2000) Discretizors and Interpolators for Five-Axis CNC Machines. *Journal of Manufacturing Science and Engineering* Vol. 122, pp.191-197.

4. **Kurt Zierhut,** (2000) How to Best Make Use of the High Speed Machining Capabilities of Your Control. In Proceedings of SME Conference on High Speed Machining, May 4-5, Cincinnati, Ohio, U.S.A, paper no. MR00-132.

5. **Filip, D. J., Magedson, R.,** and **Markot, R.,** (1986) Surface Algorithms using Bounds on Derivatives. In *Computer Aided Design*, Vol. 3, pp.295-311.

6. **Wang, F. C.,** and **Yang, D. C. H.,** (1992) Nearly Arc-length Parameterized Quintic Spline Interpolation for Precision Machining. In *Computer Aided Design*, Vol. 25, pp.281-288.

7. **Koren, Y.,** and **Lin, R. S.,** 1995 Five-Axis Surface Interpolators. In *Annals of CIRP*, Vol. 44, No.1, pp.379-382.

8. **Lo, C-C.,** (1992) Cross-Coupling Control of Multi-Axis Manufacturing Systems. *Ph.D. Dissertation,* Department of MEAM, The University of Michigan, Ann Arbor, MJ.

9. **Lo, C-C.,** (1997) Feedback Interpolators for CNC Machine Tools. In *ASME Journal Manufacturing Science Engineering*, Vol. 119, pp.587-592.

10. **Taylor, F. W.,** (1907) On the Art of Cutting Metals. In *Transactions of ASME*, Vol. 28, pp.31-279.

11. **Metcut Research Associates** 1980, *Machining Data Handbook, 3rd ed.* (Cincinnati, OH: Metcut).
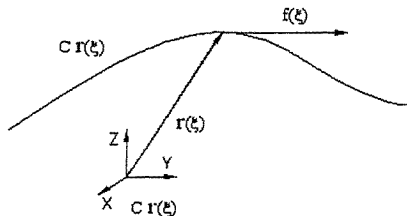
Figure 1: Representation of position and feed along a complex curve

Figure 2: Contour error due to linear discretization of the curve

Figure 3 Influence of surface gradient on effective cutter diameter

Figure 4: Feedrate variation estimated by using Criterion 1

Figure 5: Feedrate variation estimated by using Criterion 2

Figure 6: Feedrate error estimated by using Criterion 1

Figure 7: Feedrate error produced by using Criterion 2

Table-1 Various parameters estimated at different surface gradients using Criterion 1

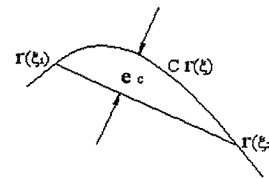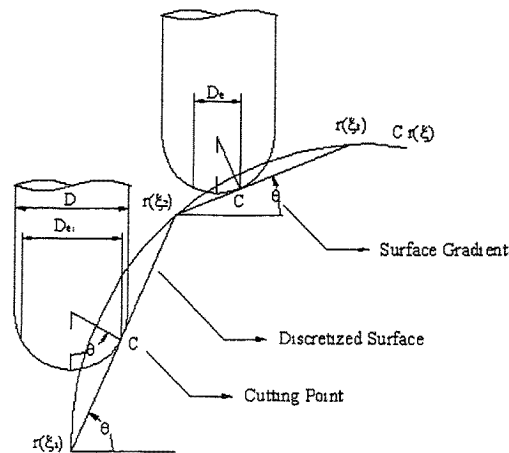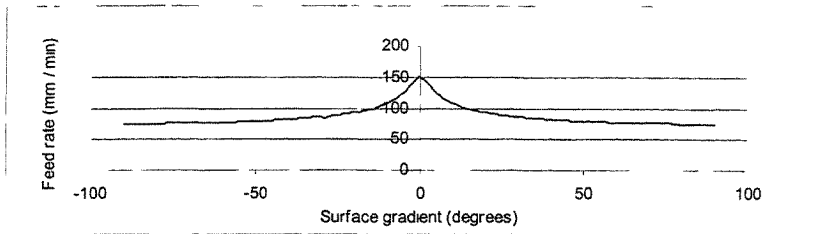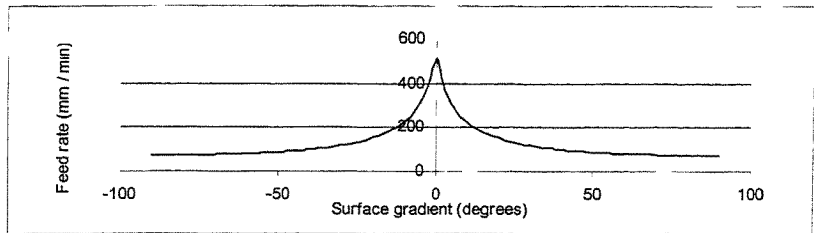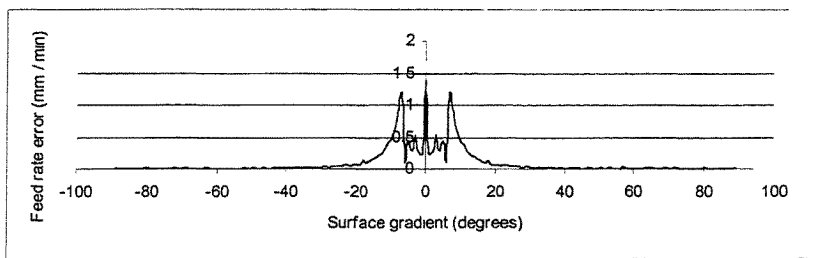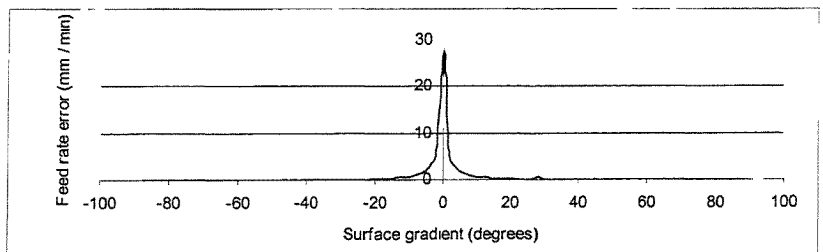| Surface gradient (Degrees)	Feedrate (mm/min)	Feedrate error (mm/min)	Fraction of Tool life spent	Machining time (min)
0	152.64	1.4030	0	2.85e-3
15	98.516	0.1530	2.8e-9	4 43e-3
30	86.114	0.0238	1.65e-6	5.05e-3
45	80.644	0.0110	5.62e-5	5.4e-3
60	77.592	0.0079	4.43e-4	5.61e-3
75	75.874	0.0013	1.36e-3	5.74e-3
90	75.302	-	1.88e-3	5.78e-3

Table-2 Various parameters estimated at different surface gradients using Criterion 2

Surface gradient (Degrees)	Feedrate (mm/min)	Feedrate error (mm/min)	Fraction of Tool life spent	Machining time (min)
0	513.82	27.750	0	8.38e-4
15	180.62	0.3738	2.3e-9	2.32e-3
30	118.29	0.1520	7.65e-6	3.83e-3
45	95.4	0.0303	1.52e-4	4.65e-3
60	80.772	0.0134	5.5e-4	5.35e-3
75	76.32	0.0110	1.35e-3	5.7e-3
90	74.412	-	1.8e-3	5.8e-3

Table-3 Various parameters estimated at different surface gradients using Criterion 3

Surface gradient (Degrees)	Feedrate (mm/min)	Feedrate error (mm/min)	Fraction of Tool life spent	Machining time (min)
0	120	0	0	2.85e-3
15	120	0	1.12e-8	4.43e-3
30	120	0	1.16e-5	5.05e-3
45	120	0	4.76e-4	5.4e-3
60	120	0	4.17e-3	5.61e-3
75	120	0	0.0132	5.74e-3
90	120	-	0.0185	5.78e-3

Table-4 Various parameters estimated at different surface gradients using Criterion 4

Surface gradient (Degrees)	Feedrate (mm/min)	Feedrate error (mm/min)	Fraction of Tool life spent	Machining time (min)
0	84.588	1.815	0	2.85e-3
15	92.22	0.2385	7.85e-10	4.43e-3
30	102.07	1.11	1.885e-6	5.05e-3
45	120.84	0.0397	2e-4	5.4e-3
60	139.28	0	3.88e-3	5.61e-3
75	157.72	0.039	0.024	5.74e-3
90	167.28	-	0.047	5.78e-3

Table-5 Comparison of different parameters estimated employing different criteria

Criteria	Total Tool life spent	Total Machining time (min)
Criterion1	0.01706	0.928
Criterion 2	0.01838	0 771
Criterion 3	0.4296	0.654
Criterion 4	1.12	0.436

A machining feature extraction approach for casting and forging parts

B F WANG, Y F ZHANG, and J Y H FUH
Department of Mechanical Engineering, National University of Singapore, Singapore

ABSTRACT

An approach to extract machining features based on machined faces (M-faces) for casting and forging parts is presented. There are four phases in the recognition process: 1) identification of M-faces; 2) decomposition of the removed simple volumes into delta simple volumes (DSV) with M-faces. 3) gluing these DSVs into sets of maximal simple volumes (MSV) by M-faces. 4) mapping these MSVs into feasible machining process types. This strategy is process oriented and feature independent. It recognizes features that can be produced by common machining operations in a uniform way and produces alternative sets of machining features.

NOMENCLATURE

Raw part model (RPT)	a 3D stock specified by the user
Final part model (FPT)	a 3D part to be produced from RPT
Total removal volume (TRV)	the Boolean difference of the RPT and FPT, which may consist of one or more disconnected bodies or lumps
Machined face (M-face)	the face on the FPT at which material is removed by machining processes from the RPT
Machined face group (M-group)	the collection of M-faces, which can be produced by a single machining process
Simple volume (SV)	a volume produces an empty set partitioned by its M-faces
Delta simple volume (DSV)	the partitioned delta simple volume
Maximal simple volume (MSV)	the glued maximal simple volume
Color attribute face(C-face)	the face on the DSV produced by partitioning

1 INTRODUCTION

Mechanical parts are typically manufactured using multiple manufacturing processes that include primary and secondary processes. Primary processes, such as casting and forging, realize the primary shape of the part, while secondary processes such as machining generate more detailed shape of the part. The machining operations are carried out on the component

where critical functional requirements like fits and assembly arise, and hence the material removal volume generally falls below 10% of the total stock volume. This is a common phenomenon in automobile and machine tool manufacturing industries. This paper focuses on such types of part models and provides a unique approach to extract their complex features.

Unlike most of the previous research, our approach is based on the general techniques for dealing with intersecting features. The features handled in this paper are restricted to machining features only, which can be considered as the portion of a part having some manufacturing significance and can be created by common machining operations, such as drilling, boring, reaming, milling, shaping, planing, broaching, etc. The initial input to the system is FPT, RPT and the reference position of FPT with respect to RPT in the workspace. The TRV is obtained by conducting a Boolean decomposition process. Faces on the TRV are segregated into M-faces and non-M-faces. And the interacting removal volumes are partitioned into delta simple volumes with their M-faces. A generate-and-test strategy is then used to glue the delta simple volumes into accessible and maximal simple volume features. Computational geometry techniques are used to produce the sets of maximal simple volume machining features. Because some sets of machining features may not be accessible, the features' accessibility and their interactions with others are analyzed and tested in a verification phase. The validity tests ensure that each set of the proposed features is accessible, does not intrude into the desired part, and satisfies other machinability conditions. The process continues until it produces a complete decomposition of the volumes into the maximal simple volume features. The maximal simple volume features are then concatenated into M-groups according to the M-faces. The geometrical and topological details from the M-faces of each M-groups and machining attributes like tolerance and surface roughness are also mapped into all the feasible operation types, based on a machining process library.

The system has been implemented using Parasolid libraries [1] and in the Visual C++ environment, and on the basis of the work done in [2]. Based on the tests, the developed algorithm is able to recognize most of the common machining operations performed on the casting and forging components.

2 RELATED WORKS

There is much literature in the area of feature extraction. The following review focuses on those closely related to the machining feature extraction based on volume decomposition problem dealt in this research.

Woo [3] developed a decreasing convex hull algorithm, which decomposes the workpiece either as a series of additive or subtractive solids. Kim [4] used an alternative sum of volumes with partitioning (ASVP) derived from some edges to solve non-convergence problem of convex decomposition. Currently the method can be applied to the polyhedrons because of the complexity of convex hull composition for curved objects. But the conversion from decomposed convex components to machining features is not sure. Wang and Chang [5] developed a backward-growing feature recognizer capable of decomposing intersecting features into elementary machined shapes. Tseng and Joshi [6] combined volume decomposition and volume reconstruction to provide multiple interpretations for interacting machining features based on predefined features. Karinthi and Nau [7] proposed a feature extraction approach based on algebra of features to solve feature interaction problem. Dong

and Vijayan [8] developed a BS-CE approach to minimize the total number of setups, and to remove the machining volume as much as possible in a setup. However, it's hard to get the LBS and NCE (NP), and also some un-machined features may be produced. Shen and Shah[9,10] proposed a process-oriented classification of machining features by using a half-space partitioning method, which recognizes features that can be produced by common machining operations. Bidarra et al. [11] introduced a cellular representation for feature models that can manage the data effectively to overcome the limitation in the feature modeling system. Shirur et al. [12] also used the process-oriented classification for mapping machining volumes to machining operations by encoding geometric shapes into algebraic expression. The process-oriented approach has an advantage over the feature-based approach in that an undefined feature can still be handled as long as it can be produced by a process.

Most of the above approaches match the decomposed volume features into "standard" machining features defined by using feature-based or hint-based method, which nevertheless will cause problems in robustness. First, the number of representative templates may not be enough in number. Second, the processes of characterizing the templates are almost all heuristic and may not be valid.

The approach introduced in this paper employs the process-oriented approach. It first obtains the machining removal volumes by conducting a Boolean difference operation between the RPT and the FPT. It is noted that the procedure up to this stage is similar to the volume decomposition approach [6,9,12,13]. But from the removal volumes, the reported approaches basically map the volumes to processes, and the raw stock used is always a rectangular block. When such an approach is applied to casting and forging stock, problems will arise. For example, the shape of the removal volumes is generally not regular but rather complex in nature. Hence, it is not always possible to map the volumes into processes. To overcome this problem, in our approach, the M-faces are identified from the removal volumes and grouped into clusters (M-groups), which can be produced by a single process type. The objects handled are faces, not volumes. In such a way, casting and forging components can be handled effectively. The partitioning procedure seems to be the same with refs [6,8,9,10], but [6] partitioning with extended faces, [8] with the maximal face and [9,10] with half-space. Those are computationally intensive. Here, we use M-faces. Each disconnected body obtained by Boolean difference operations between the RPT and the FPT of casting and forging parts is always local, and the number of M-faces on it is usually small. Therefore the computational complexity is not so intensive. The gluing procedure is also based on M-faces relationship. So the glued volumes can be directly mapped into machining process.

3 THE PROPOSED FEATURE EXTRACTION APPROACH

The inputs to the proposed feature extraction approach are the FPT, RPT, and the reference position and orientation in which the two models overlap. The feature extraction process proceeds with the following steps, as illustrated in Fig.1:

(1) generating TRV and sets of disconnected body V_i;
(2) identifying M-faces on V_i;
(3) partitioning V_i into sets of DSV_{ij} with its M-faces;
(4) gluing each sets of DSV_{ij} into sets of MSV_{ijn} based on its M-faces;
(5) classifying MSV_{ijn} into M-groups; and
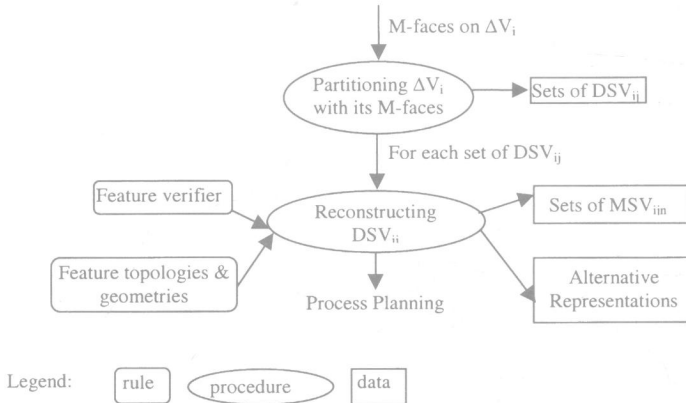
(6) mapping M-groups to operation types.

Fig.1 Feature extraction process

3.1 Generating TRV

The FPT is translated to the reference position and the orientation of the RPT. An example is shown in Fig.2. This is followed by the Boolean difference of FPT and RPT, resulting in the TRV, which may be one body or a set of disconnected bodies. The TRV can be expressed as

$$TRV = \{\Delta V_1; \Delta V_2; \Delta V_3; \Delta V_4; ...; \Delta V_n\}$$
(1)

where n is the number of disconnected bodies (ΔV_i). Fig.3 shows the volume obtained by Boolean difference of FPT and RPT as positioned and oriented in Fig.2.

(a) RPT   (b) FPT   c) Position & orientation of
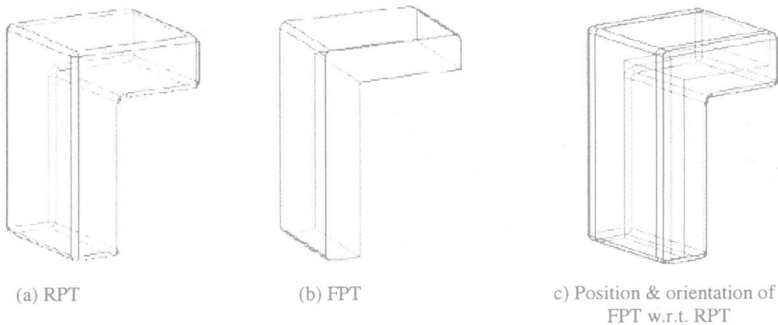   FPT w.r.t. RPT

Fig.2 Boolean difference of FPT and RPT

3.2 Identifying M-faces

The faces on the machining volume are primarily categorized into two types, namely, M-faces which come from the FPT, and non machined faces (NM-faces) which come from the RPT. In the TRV, each ΔV_i possesses at least one M-face. By performing topological intersections

and Vijayan [8] developed a BS-CE approach to minimize the total number of setups, and to remove the machining volume as much as possible in a setup. However, it's hard to get the LBS and NCE (NP), and also some un-machined features may be produced. Shen and Shah[9,10] proposed a process-oriented classification of machining features by using a half-space partitioning method, which recognizes features that can be produced by common machining operations. Bidarra et al. [11] introduced a cellular representation for feature models that can manage the data effectively to overcome the limitation in the feature modeling system. Shirur et al. [12] also used the process-oriented classification for mapping machining volumes to machining operations by encoding geometric shapes into algebraic expression. The process-oriented approach has an advantage over the feature-based approach in that an undefined feature can still be handled as long as it can be produced by a process.

Most of the above approaches match the decomposed volume features into "standard" machining features defined by using feature-based or hint-based method, which nevertheless will cause problems in robustness. First, the number of representative templates may not be enough in number. Second, the processes of characterizing the templates are almost all heuristic and may not be valid.

The approach introduced in this paper employs the process-oriented approach. It first obtains the machining removal volumes by conducting a Boolean difference operation between the RPT and the FPT. It is noted that the procedure up to this stage is similar to the volume decomposition approach [6,9,12,13]. But from the removal volumes, the reported approaches basically map the volumes to processes, and the raw stock used is always a rectangular block. When such an approach is applied to casting and forging stock, problems will arise. For example, the shape of the removal volumes is generally not regular but rather complex in nature. Hence, it is not always possible to map the volumes into processes. To overcome this problem, in our approach, the M-faces are identified from the removal volumes and grouped into clusters (M-groups), which can be produced by a single process type. The objects handled are faces, not volumes. In such a way, casting and forging components can be handled effectively. The partitioning procedure seems to be the same with refs [6,8,9,10], but [6] partitioning with extended faces, [8] with the maximal face and [9,10] with half-space. Those are computationally intensive. Here, we use M-faces. Each disconnected body obtained by Boolean difference operations between the RPT and the FPT of casting and forging parts is always local, and the number of M-faces on it is usually small. Therefore the computational complexity is not so intensive. The gluing procedure is also based on M-faces relationship. So the glued volumes can be directly mapped into machining process.

3 THE PROPOSED FEATURE EXTRACTION APPROACH

The inputs to the proposed feature extraction approach are the FPT, RPT, and the reference position and orientation in which the two models overlap. The feature extraction process proceeds with the following steps, as illustrated in Fig.1:

    (1) generating TRV and sets of disconnected body V_i;
    (2) identifying M-faces on V_i;
    (3) partitioning V_i into sets of DSV_{ij} with its M-faces;
    (4) gluing each sets of DSV_{ij} into sets of MSV_{ijn} based on its M-faces;
    (5) classifying MSV_{ijn} into M-groups; and

$$TRV = \sum_{i=1}^{n} \Delta V_i = \sum_{i=1}^{n} \sum_{j=1}^{m} DSV_{ij} \qquad (4)$$

m, n has the same meaning as in Equation (3). This is an exhaustive partitioning. Each set of DSV_{ij} will be reconstructed into MSV_{ijn} in the following step. The computational complexity is: CC ~ $N*2^m$, where N is disconnected bodies, m is number of M-faces on each disconnected bodies. Fig.4 shows the sets of DSV obtained by partitioning the ∆V in Fig.3. There are nine DSVs after partitioning. The number of M-faces is four.

3.4 Gluing DSV$_{ij}$ by M-faces

This stage reconnects and glues the DSV_{ij} produced in the partitioning process above. The aim is to get the sets of MSV that can be machined in a set-up operation. A set of MSV can be taken as a different set of operation or a different sequence required for machining the intersecting volume. Some sets may result in better manufacturing practice from the engineering point of view and some may not be lucrative in terms of economic aspects, which are more significant during the process selection.

3.4.1 Gluing procedure

Fig.5 below shows the gluing algorithm. There are several stages in the gluing procedure:
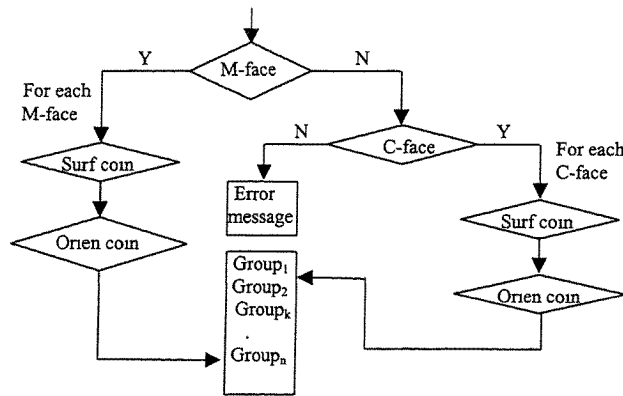

Fig.5 Gluing algorithm

(1) Find M-faces on ΔV_i, and get their surfaces and orientations.

(2) Find M-faces on each partitioned DSV_{ij}, and compare their surfaces and orientations with the ones on the ΔV_i got from step 1 respectively.

(3) If more than one M-faces found, then judge if there are more than one M-faces or C-faces found in the previous $DSV_{i(j-1)}$. If true, then copy the previous $DSV_{i(j-1)}$'entities which are classified, and put this entity to the corresponding groups separately according to M-faces on it. If false, no need to copy the previous $DSV_{i(j-1)}$'entities which are classified, just put this entity to the corresponding groups separately

according to M-faces on it. If no M-faces found, go to step 4. Else, change to the next delta simple machining volume $DSV_{i(j+1)}$, and go to step 2.

(4) Find C-faces, which are produced by partitioning, get their surfaces & orientations, and compare their surfaces and orientations with the ones on the ΔV_i respectively.

(5) If more than one C-faces found, then judge if there are more than one M-faces or C-faces found in the previous $DSV_{i(j-1)}$. If true, then copy the previous DSV 'entities which are classified, and then put this entity to the corresponding groups separately according to C-faces on it. If false, no need to copy the previous $DSV_{i(j-1)}$ 'entities which are classified, just put this entity to the corresponding groups separately according to C-faces on it. If no C-faces found, report error message. Else, change to the next delta simple machining volume $DSV_{i(j+1)}$, and go to step 2.

(6) Unite the entities in each group G_{ik} and get the MSV sets for ΔV_i, then go to ΔV_{i+1}.

This algorithm glues the DSV_{ij} based on the M-faces on V_i. According to our algorithm, more than one set of MSV for each V can be produced. The number is determined by the number of M-faces or C-faces on each DSV. Then a best set can be obtained through the tool access test and optimization. By far, we can express the V_i and TRV as following:

$$\Delta V_i = \bigcup_{s=1}^{l} \sum_{i=1}^{k} MSV_j \qquad (5)$$

$$TRV = \sum_{i=1}^{n} \Delta V_i = \sum_{i=1}^{n} \bigcup_{s=1}^{l} \sum_{j=1}^{k} MSV_{ij} \qquad (6)$$

where l is the number of MSV sets, k is the number of MSV, n is the number of V_i

In the nine DSVs in Fig.4, two have three C-faces, one has two M-faces, and the others have one M-face each, so they can be glued into eighteen sets of MSV. And there are six distinct sets as shown in Fig.6. Next, we will test the tool accessibility and other manufacturing constraints to check if all of them are machining feasible and try to find an optimal one.

3.4.2 Testing

The sets of MSV produced above may be candidates for possible machining feature sets. In general, many combinations exist and the feature gluing can, if desired, find all valid combinations that satisfy the gluing rules. To help determine whether the candidate volumes can be matched to at least one machining process we classify each volume based on tool accessibility and relative degrees of freedom of each volume with respect to the part. The procedures described above are "process-oriented" rather than "feature-oriented", making them independent of feature types or interaction types to be recognized. We evaluate each candidate volume set for its machining characteristics, rather than predefining a closed set of features and hints for recognizing it. Some set may be tool inaccessible, so the validity test is required to check its accessibility. With this procedure, we can get the feasible sets and choose the optimal one by optimizing the machining time. The optimization takes into account two conditions, namely, minimal machining area and minimal setups.
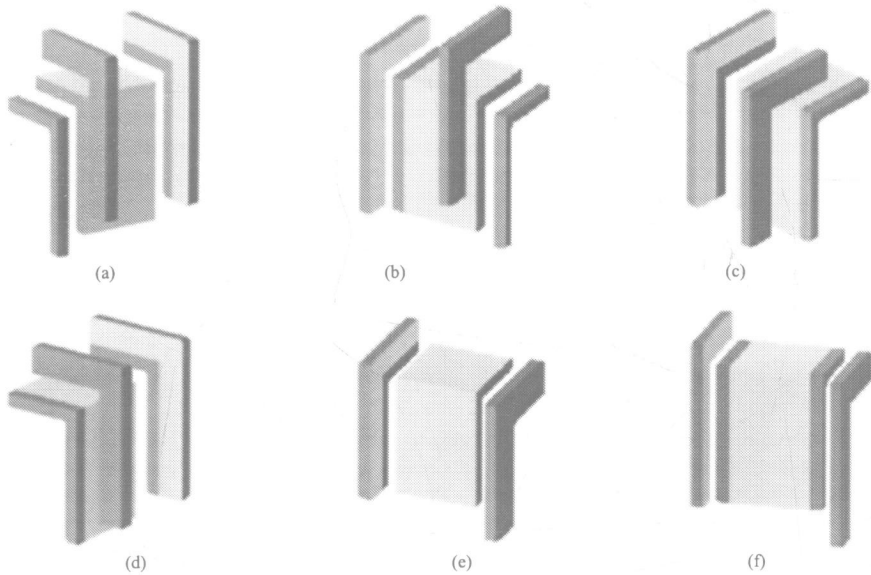
Fig.6 Sets of MSV

The test shows the eighteen sets of MSV are all machining feasible, but the best one is the (e) in Fig.6. This set has the minimal machining area and the minimal setups.

3.5 Mapping sets of MSV$_{ijn}$

In this procedure, M-faces are concatenated into M-groups. Since each MSV is obtained based on M-faces, it is easy to map them into M-groups. Operation selection from M-groups is a process of identifying what manufacturing activity can be used to transform raw material into a finished part. The selection task is done in two levels. The first selects the rough machining operation and the next searches for the successive and final operations. This is based on the selection knowledge stored in the process library, which can be categorized as geometric and engineering constraints:

1. Geometric constraints
   - characteristic surfaces and their inter-relations,
   - tool approach directions (TADs) of M-groups.
2. Engineering constraints
   - dimension and tolerance,
   - material hardness,
   - surface roughness.

4 CASE STUDIES

The current implementation of the system uses Parasolid Library [1] as the geometry kernels, runs on Windows XP platform and in the Visual C++ environment. Parts are created on

Unigraphics 17.0 platform, and transformed to *.xmt_txt* files which can be recognized by Parasolid. We have tested the system with a large variety of parts. The above case is one example tested by the programs developed.

5 CONCLUSIONS

This paper reports an approach to extract machining features for casting and forging parts. The approach is process-oriented and feature independent. Each disconnected body obtained by Boolean difference operation between the RPT and the FPT of casting and forging parts is always local, and the number of M-faces on it is usually small. Therefore the computational complexity is not so intensive. Since feature mapping and the gluing procure are based on the machined faces, the glued volumes can be directly mapped into machining process and the shape of the removed volume does not play any significant role. These characteristics make the system suitable to handle the complex casting and forging parts.

REFERENCES

(1) Parasolid Solid Modeler, Version 10.1. *Programming Reference*, Unigraphics Solutions Inc., Maryland Height, 1999.
(2) S.B.Kailash, Y.F.Zhang, J.Y.H.Fuh. A volume decomposition approach to machining feature extraction of casting and forging components. *Computer-Aided Design* 2001; 33:605-617.
(3) Woo TC. Feature extraction by volume decomposition. In: *Proceedings of the Conference on CAD/CAM* 1982; p.39-45.
(4) Kim YS. Recognition of form features using convex decomposition. *Computer-Aided Design* 1992; 24(9):461-76.
(5) Wang M-T, Chang T-C. Feature recognition for automated process planning. In: *Proc. Manuf. Int. 90*, Part-2: Adv. Manuf. Syst., ASME, New York, 1990, p.49-54.
(6) Tseng YJ, Joshi S. Recognizing multiple interpretations of intersecting machining features. *Computer-Aided Design* 1994; 26(9):667-88.
(7) Karinthi R, Nau D. An algebraic approach to feature interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1992; 14(4):469-84.
(8) Jian Dong, Sreedharan Vijayan. Manufacturing feature determination and extraction Part II: a heuristic approach. *Computer-Aided Design* 1997; Vol.: 29, No.7, pp.475-484.
(9) Shen Y, Shah JJ. Recognition of machining features based on HSPCE decomposition, feature composition, and process oriented classification. *Transactions of the ASME, Journal of Mechanical Design* 1998; 120:668-77.
(10) Yan Shen, JJ. Shah. Feature recognition by volume decomposition using half-space partitioning. *Advances in Design Automation* 1994; DE-vol. 69-1, Volume 1, ASME.
(11) Bidarra R, de Kraker K, Broonsvoort W. Representation and management of feature information in a cellular model. *Computer-Aided Design* 1998; 30(4):301-13.
(12) Shirur A, Shah JJ, Hirode K. Machining algebra for mapping volumes to machining volumes for developing extensible generative CAPP. *Journal of Manufacturing Systems*, 1998; 17(3):167-82.
(13) Sakurai H, Chin C. Definition and recognition of volume features for process planning. In: Shah, et al., editors. *Advances in feature based manufacturing*, Amsterdam: Elsevier 1994.

Intelligent Systems

An application of expert system in manufacturing – a case study

H K WONG
School of International Business, University of South Australia, Australia

SYNOPSIS

This paper presents an application of an expert system for a computer chipset assembly process in a manufacturing domain. The system focuses on assisting maintenance personnel in troubleshooting and diagnosing product defects. The case study was developed using the student version of VP-Expert as the development tool. This paper discusses the need for expert system for a ball-attach process, its applicability, and the process of knowledge acquisition, coupled with knowledge representation using a production rules system. Implementation of the expert system to assist product defect diagnosis can be expected to simplify maintenance jobs and help to increase overall productivity of the ball-attach process with reduced machine down time.

1 INTRODUCTION TO EXPERT SYSTEM AND VP-EXPERT

Expert systems have gained a great amount of attention in recent decades. Varied applications, efficiencies, and ease of understanding of the tool have lead to a phenomenal growth in this field of study. Briefly explained, expert systems is a branch of computer science concerned with the design and implementation of programs that are capable of emulating human thinking skills such as problem solving, visual perception and understanding of language. An expert system can simulate the performance of human experts in their respective domains, which involves both a set of knowledge and the manipulation of the knowledge (Lee, 1992). The tool often operates as an interactive system that responds to questions, asked by users through "expert" recommendations, guidance and decision making processes.

VP-Expert is a low cost software PC based tool used to develop expert systems. It has been applied in different fields to solve engineering, business and manufacturing problems. It is a particular tool that combines ease of use, economy, and the ability to use spreadsheets and database files. The modular nature of VP-Expert can facilitate restructuring, maintenance and building up of the knowledge base through knowledge acquisition. These are the main reasons why VP-Expert was selected.

2 AN EXPERT SYSTEM MODEL

Expert systems can be applied in a wide variety of fields. Most applications of expert systems will fall into one of the following categories; interpreting, identifying, predicting, diagnosing, designing, planning, monitoring, debugging and testing, instructing and training and controlling (Adedeji, 1992).

As VP-Expert is a low cost software PC-based computer tool, it is found to have an increasing number of applications in the manufacturing environment. Application areas include machine breakdown and fault diagnosis, system configuration, vision, data interpretation and process control, to name only a few.

The expert system model that is the subject of this paper is product defect diagnosis for a computer chipset assembly. The system focuses mainly on assisting maintenance personnel in troubleshooting and diagnosing products defects encountered in the ball attach process. Basically, the expert system will lead users through a series of observations on product defects encountered at the ball attach process of the chipset assembly. A hypothesis on the product defects are drawn based on the observations, coupled with an "expert" recommended action to be taken.

3 PROCESS OVERVIEW

The product selected is known as ball grid array (BGA) computer chipset. First, a copper frame is mounted onto the core of computer chipset, which is an organic compound. Second, with the transition of the product, a new process, known as ball-attach, is introduced. The entire assembly flow for the BGA product consists of five sub-processes as shown below:

I. **Die-attach:** This is the first process of assembly, in which the incoming substrate with core material and build-in copper traces for connectivity purposes is introduced. The functional die will be attached onto the individual unit of the substrate.

II. **Wire-bond:** In this process, gold wire with 99.999% purity will be bonded onto the functional die to provide the connectivity to the substrate.

III. **Mould:** During the process, the appropriate amount of organic mould compound will be moulded onto the substrate to protect the functional die and gold wire from physical damage.

IV. **Ball attach:** In this process, eutectic solder balls (63%Sn / 37%Pb) will be attached onto the bottom side of the substrate to provide further connection in which the chipset unit will later be attached onto the printed circuit board in the computer motherboard assembly.

V. **Laser mark / Singulation:** The substrate, which consists of five fully assembled chipset units, will be laser marked with product logo, singulated into individual units, and ready for final inspection and packing. The graphical presentation of the process flow is as shown in Figure 1.
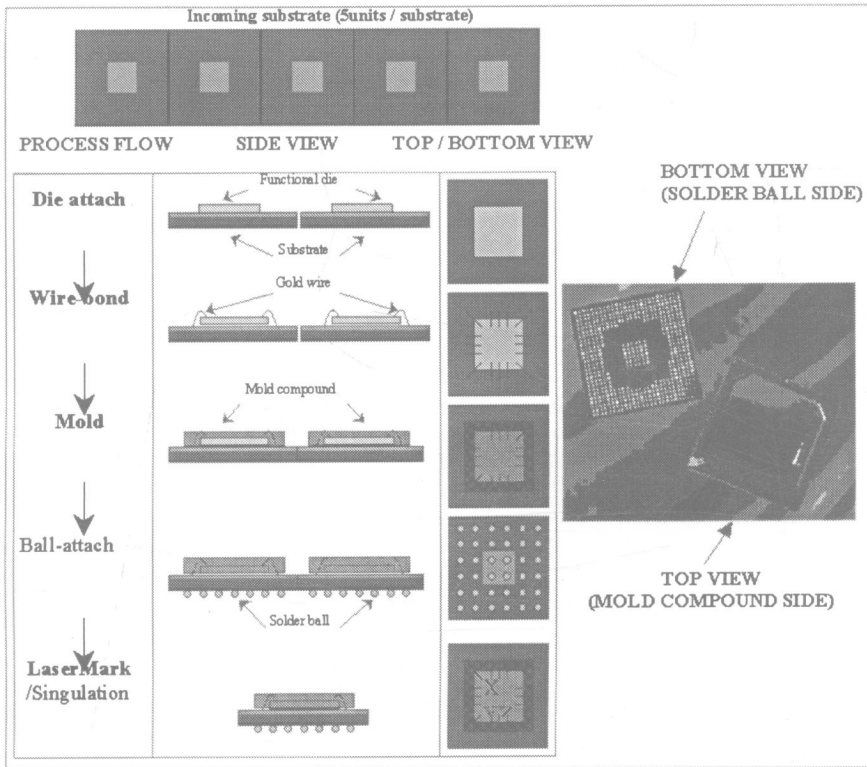
Figure 1. Assembly flow for computer chipset products (modified original from Carlo et al., 1998)

4 THE NEED FOR EXPERT SYSTEM

The ball attach process operates in a very tight process window and is sensitive to external factors, such as incoming material, upstream processes and environmental particle count. A slight drift of such factors can badly impact product quality. In addition, the process is the bottleneck operation of the entire assembly line due to limited capacity of the equipment. It is currently operating at the maximum capacity with minimum down time allowed. Any drift in the process requires a quick fix to contain the problems and reduce machine idle time, in order to reduce loss of any production volume. Because the high complexity of the equipment had greatly complicated product defects diagnosis, high-skill maintenance personnel are required at all time to handle the challenging task.

However, with the recent introduction of the process to the assembly line, only a few maintenance personnel had acquired the required knowledge. The situation was further complicated by the existing job rotation system to increase job diversity of the maintenance personnel. This was intended to maintain their interest in the job by having a change in the working environment. Moreover, the cost of maintenance will rise, as there are few skilled

technical people to maintain constant service for the machines. Such adverse conditions raise the possibility for an expert system to improve maintenance performance for the ball-attach process. The expert system can be set up through a presentation of menu-driven instruction guides to assist the maintenance personnel in performing complex repairs job (Adedeji, 1992 & Carlo et al., 1998).

5 APPLICABILITY OF AN EXPERT SYSTEM

The concerns discussed above appeared to fit the model for an expert system to improve process diagnosis and maintenance performance. Some of the reasons why an expert system is applicable for such situation are listed below:

- *Human expert is not available at all time:* Chipset manufacturing operates 24-hours per day non-stop. With only two maintenance personnel with required expertise, it is impossible to cover all of the three work shifts per day to assist in troubleshooting. With the use of an expert system, the knowledge can be evenly distributed to all working shifts.

- *The diagnosis process has a narrowly defined domain:* There is a very clear-cut boundary between each process in the chipset assembly, as each process is unique. For the application of this paper, the domain selected for the expert system model will be the ball-attach process only.

- *The diagnosis process is highly complex but requires a quick fix:* The ball-attach equipment consists of complex mechanisms, which require high-skilled maintenance personnel to perform the troubleshooting job. However, due to the limited capacity of the equipment as mentioned earlier, the troubleshooting process requires a quick fix to reduce machine idle time so that the entire production will not be impacted. It was found that a non-expert often took a few hours to solve a problem, which is unacceptable. However, an expert system with a complex knowledge base that can represent a human expert is capable of handling such situations at a reasonable speed.

- *It serves as a "corporate knowledge base":* There is another assembly site producing the same chipset product. Due to the high cost and time involved to get the maintenance personnel fully trained, it is recommended to set-up the expert system which can be applicable to the other assembly site without lots of training effort required.

6 KNOWLEDGE ACQUISITION

6.1 Knowledge Gathering

There are three techniques used in knowledge gathering as listed below:

- *Relevant documents:* There are two types of documents used. The first document is the failure mode and effect analysis (FMEA) documentation for the ball-attach process (Adedeji, 1992 & Carlo et al., 1998). This documentation was based on the data

collected from the maintenance records, including the type of defect modes, the occurrence frequency, and the total machine idle time affected, coupled with the feedback from the maintenance personnel. The FMEA documentation contains all of the common product defect modes encountered in the past, with a list of associated causes of failures. All of this information has been used to develop the basic knowledge base for the expert system. The second type of document used is the operational manual for the ball-attach equipment (Carlo et al., 1998 & Vanguard 5200/5500 Operational Manual, 1986). The guidance available in the manual has helped in developing the preliminary solution for each associated cause of failures.

• **Past experiences:** The preliminary solutions for the failure modes have been further improved through personnel work experience as a process engineer under the chipset-manufacturing environment. There were a lot of troubleshooting skills gained through process and equipment training sessions and hands-on experiences.

• **Interview with human expert:** This is the last technique used for knowledge acquisition. Since the human experts are not available in Australia, an interview session was conducted through e-mail. It consists of a list of questionnaires to obtain the recommended solutions and actions to be taken for some given failure modes.

7 KNOWLEDGE REPRESENTATION

7.1 The Process and the Equipment

The knowledge gathered on the product defects was grouped into seven different categories. They are as shown in the list below:
  I.   *Missing solder ball(s) from the original ball pad*
  II. *Solder ball offset from the original ball pad*
  III. *Physical damage to the product*
  IV. *Foreign material found on the product*
  V.   *Flux related problems*
  VI. *Smashed solder balls found on the product*
  VII. *Black solder ball(s) found on the product*

These product defects can be either induced from the environment, the process or mechanisms involved during the process. As for the process, it can be further divided into five individual steps as shown below:
  I.   *On-load*
  II.   *Fluxing*
  III. *Ball placement*
  IV.   *Vision inspection*
  V.   *Off-load*

The flow of the steps corresponds to their functions. Mechanisms involved in the equipment correspond with the technical terms used. Each mechanism serves as a very important element in the ball-attach process and contributes to the product defects listed above in one way or another.

7.2 Knowledge Base

With the seven product defects being identified earlier, the knowledge base of the expert system was built through a series of observations for a specific product defect mode observed. This was then tabulated using the logic of a decision table as shown below.

Table 1. Sample of decision table on product defect diagnosis for the ball-attach process

1st observation	Damage condition	Occurrence place	Area in on loader	Side mould flash	Area in off loader	Defective magazine	Crack condition	Hypothesis	Recommendation
Physical damage-1C	Major, thru naked eyes –5A	On loader - ball attach machine-6A	Fluxer-7A	Yes-8A	-			Upstream process drifted	Inform mould engineer
Physical damage	Major, thru naked eyes	On loader - ball attach machine	Fluxer	No-8B	-			Improper walking beam finger gripping	Fine-tune the walking beam finger walking beam finger gripping
Physical damage	Major, thru naked eyes	On loader - ball attach machine	Vision inspection -7C	-			-	Push arm over travelled	Adjust push arm
Physical damage	Major, thru naked eyes	Reflow oven-6B	-	-			-	Substrate jammed between the conveyor link	Replace the broken link
Physical damage	Minor, hairline crack-5B	-	-	-			Mould gate-11A	Squeegee pressure too high	Reduce squeegee pressure
Physical damage	Minor, hairline crack	-	-	-			On front and end of substrate-11B	Push arm over travelled	Adjust push arm
Physical damage	Dented on solder ball surface	-	-	-			-	Unloader stop gate sensor faulty	Re-teach the sensor

The knowledge is then converted into a set of production rules, using the IF-THEN commands as shown in the format below:

    IF   Condition 1 AND
          Condition 2 AND
          Condition 3 AND

          .
          .

          THEN Hypothesis
          Recommendation

For instance, the observation of physical damage of the product may be as described in the first row and column of table 1 leading to the hypothesis is " upstream process drifted" (row 1 - column 9) with the recommendation to inform the upstream process engineer. The knowledge base can be presented using the production rules shown below:

    IF   Damage condition = major, detected through naked eyes AND
          Occurrence place = On-loader – ball attach machine AND
          Area in on-loader = Fluxer AND
          Side mould flash = Yes
    THEN Hypothesis = Upstream process drifted
          Recommendation = Inform mould engineer

I002/025/2002 © With Authors 2002

7.3 Input and Output Screen

The expert system developed is intended to be simple and user friendly for easier acceptance by maintenance personnel with minimal training required. It consists of a simple input and output screen, using simple English. During the demonstration session, the sequences as shown below will be followed:

I.   A question is asked by the system to determine the defect mode observed.
II. The user is asked to select the relevant number, which correlates to the defect mode observed on the product.
III. Then, a series of observations on the selected defect mode will be provided by the system to get a more detailed description of the defect mode encountered.
IV. The user needs to select the appropriate numbers again, which corresponds to a particular condition for each observation as the inputs to the consultation.
V.   Based on the inputs by the user, a hypothesis and a recommendation will be derived and shown as the output to assist the maintenance personnel on proper actions to be taken.

Due to the limitation of characters that shows as the output of the system by using the student version of the VP-Expert package, a condition number is assigned for each condition. A sample of input and output screen during the demonstration session is shown below.
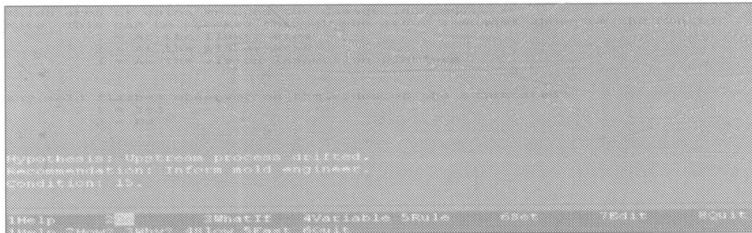

Figure 2. A sample of input and output screen during the consultation session by using the student version of VP Expert system (source from program)

8 IMPLEMENTATION STRATEGY

An expert system is normally implemented on the production floor using personal computers. A pilot running strategy will be employed. This will help ensures that the entire system is running smooth before full proliferation to other assembly sites, which are running on same product family.

The expert system for product defect diagnosis allows access by the maintenance personnel at any time, so the immediate consultation can be made without further delay. The system can lead the maintenance personnel through a sequence of observations on problems encountered as explained earlier. It was expected to provide consistent guidance for maintenance personnel to perform their jobs with high efficiency and effectiveness that can thereby bring up the competency of the new trained maintenance personnel in the short period.

9 USEFULNESS

The expert system assists product defect diagnosis for the ball-attach process and is expected to deliver the following common benefits (Lee, 1992 & Adedeji et al., 1993):

- To evenly distribute expert-level knowledge throughout all working shifts, and also cross-site to other assembly plant in the future
- To reduce the MTTR (mean time to repair – the average time between the moment when the machine is down and the time taken to start the repair job) as consultation on the required action taken can be obtained directly from the system
- To facilitate real-time expert level repair decision by regular maintenance personnel

10 CONCLUSION

This paper covered an introduction to an expert system. The selected expert system shell is VP-EXPERT. The case study of the expert system focussed on the ball-attach process of a computer chipset assembly. The constrained capacity and highly complex configuration of the ball-attach equipment, coupled with the limited maintenance personnel and the required skill required an expert system to assist in product defects diagnosis. In other words, the system also narrowly defined domain, the ability of human expert to assist in system development and its major role to serve, as a "corporate knowledge base" are to reduce training effort cross-site, which had further justified for the applicability of the expert system. This expert system is currently used in single process only. Once the system is widely acceptable to the maintenance personnel, it will be proliferated to other non-bottle neck processes to improve the overall product defect diagnosis of the entire assembly processes.

To conclude, the case study showed the practicality of an expert system used in product defects diagnosis for ball-attach process under a manufacturing environment. In a sense, benefits gained from the case study of the expert system are great and useful. It enables the expert-level knowledge to be evenly distributed, machine down times to be reduced and overall productivity of the process to be increased, just to name some of it. Hence, the application of the Expert System should not merely confine to bottleneck processes. Rather, it should serve as a valuable tool to be proliferation to other non-bottle neck processes and the entire assembly line for improving the overall product defect diagnosis process and the productivity of the entire assembly process.

REFERENCES:

1. **"Vanguard 5200/5500 Operational Manual"**, (1986) Arizona.
2. **Adedeji, B. B., Bob, L. F. and Peters, J. W.** (1993) A Case Study of Economic Justification of Expert Systems in Maintenance Operations, *The Engineering Economist*, VOL.38, No.2, pp. 46-58.
3. **Adedeji, B. B.** (1992) Expert systems: Applications in Engineering and Manufacturing, Prentice Hall.
4. **Carlo, V., Tan, T. L., Wong, A.** (1998) Emmanuel Villanueva "Operational Specification for ball attach process – Attachment H: AT Virtual Factory FMEA", High Technology.
5. **Lee, H. S. L.** (1992) Applications of expert systems in manufacturing: A case study, *Computer in Industry*, Elsevier, pp193-198.

Agent-based control of a flexible assembly cell

C K FAN and T N WONG
Department of Industrial Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong

Abstract

This paper describes the development of an agent-based infrastructure for the control of a flexible assembly cell (FAC). The proposed agent-based FAC control system comprises a collection of agents implemented in a distributed control network. The approach of the agent design is based on the object modeling technique (OMT). According to the proposed control architecture, a standard agent template for the manufacturing cell is proposed to implement the agent-based system.

1. Introduction

Advanced manufacturing systems are mostly complex systems, in which a control system is commonly established to be responsible for the planning, control and monitoring functions. In particular, the manufacturing control system has to cope with product and process changes. Traditionally, manufacturing systems are commonly established with centralized control software. In hierarchical control structure, control decisions and decomposition of the tasks are distributed downward to the lower hierarchy and the information of individual components is sent upward to components in the higher hierarchy. If one of the components in the same hierarchy is added or removed, it is difficult to modify the tight working relationships among components. Hence, it is not flexible enough to cope with the dynamic changes of manufacturing processes. Thus, it is in general difficult to develop and implement the software of the hierarchical control architecture for complex manufacturing systems.

This paper describes the development of an agent-based infrastructure for the control of a flexible assembly cell (FAC). The proposed agent-based FAC control system comprises a collection of agents implemented in a distributed control network. The approach of the agent design is based on the object modeling technique (OMT). According to the proposed control architecture, a standard agent template for the manufacturing cell is proposed to implement such an agent-based system.

2. Agent-Based FMS Control Structure

For FMS design, researchers have recognised the importance to establish an intelligent and distributed control framework in the development of the control system. Recently, researchers have identified the potential of distributed AI (DAI) in solving various complex manufacturing system problems with multi-agent systems (MAS) (1) (2) (3). In the last decade, increased research resources and attention have been given to DAI theories and

applications. An MAS is a DAI system with a collection of autonomous agents. The agents in the MAS co-operate with each other to reach common objectives, while simultaneously each agent pursues its individual objectives. In particular, the distributive control structure is commonly established with the agent-based approach (4) (5) (6) (7) (8) (9).

Agent models are built and implemented in order to attain the flexibility, reusability and fast response to both internal and external uncertainties in the shop floor environment (10). Agent-based FMS control systems are mostly established in the hybrid structure. The agent-based structure is used to enhance the inter-communication, negotiation and cooperation. This is a compromise between the hierarchical and heterarchical control structures. The main purpose of the agents is to obtain global objective from local agent solutions (11). The use of agents enriches the reliability and flexibility of manufacturing planning and scheduling functions in the dynamic manufacturing system (12). The agent structure provides reconfigurability to changes. The agents also provide the fault tolerance capability to the manufacturing cell. This capability is achieved by the re-allocation of the resources.

Contemporary agent-based systems are usually established with the object-oriented approach. The object-oriented modeling methods are appropriate modeling and implementation approaches to represent the flexibility, real time response of the complex agent-based manufacturing control systems (13) (14) (15). The object-oriented method is able to reduce the design complexity of the FMS control software and increase the reusability of the software. Figure 1 depicts a flow chart for the object-oriented design and modeling procedure for the design and analysis of the FMS control system.
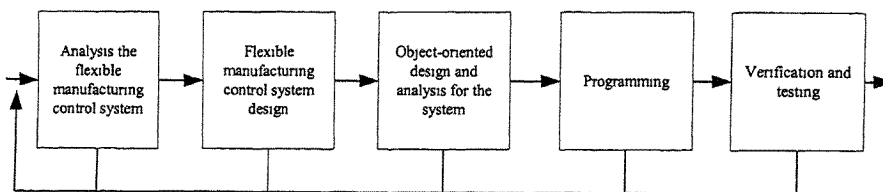

Figure 1 The Object-oriented model design procedure

3. Configuration of the FAC

Figure 2 depicts the configuration of the FAC. The FAC comprises two Adept SCARA-typed assembly robots and a conveyor loop with 4 conveyors. The two robots are controlled by a single Adept programmable controller. Assembly operations and material handling tasks can be handled by these two robots. Stoppers are located at the front of the two robots (the area is called the robot processing area) so that pallets are clamped for robot processing. The robot controller is equipped with two serial communication ports for communication.

There are four conveyors in the FAC. They form a loop connection for pallets transportation. Sensors, pneumatic cylinders, signal indicators are controlled by a programmable logic controller (PLC). The PLC is employed as the central processing unit for the motion control of the conveyor system. Ladder diagram is used to prepare the control program for the control logics of the transportation system. The PLC is equipped with a serial communication port for communication.
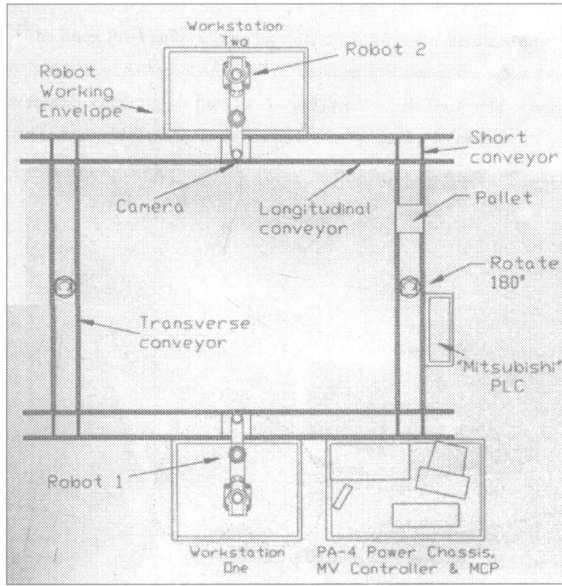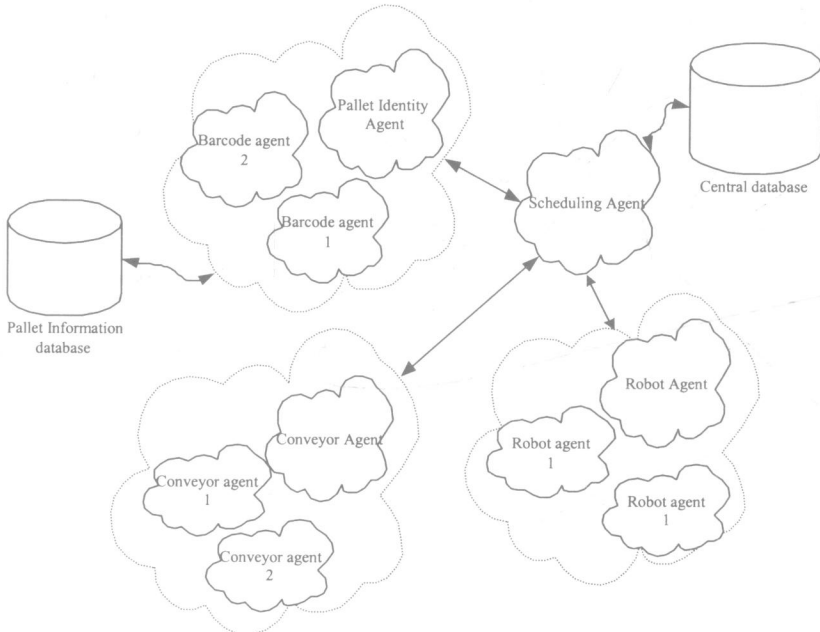
Figure 2. Configuration of the FAC

Figure 3. Some agents in the example FAC

4. Agent in the FAC control system

The FAC control system comprises a number of agents responsible for different functions. Figure 3 depicts some of the agents in the control system. In this example, four agents (scheduling agent, pallet identity agent, conveyor agent and robot agent) are used to form the core working elements in the system. The scheduling agent is responsible for coordinating the tasks between the pallet identity agent, the barcode reading agent and the conveyor control agent. This agent links up with the central database to retrieve and save the overall system information. It is embedded with a specific scheduling algorithm to cater for the scheduled instruction in the FAC. The pallet identity agent is responsible for the provision of the local pallet information and pallet position information from the sensor signal interpreter agent. The information is sent to the scheduling agent, barcode reading agent and conveyor agent. The agent has its own pallet information database. The conveyor control agent is responsible for the provision of the local solution from the two sub-level conveyor agents. The information is required by the scheduling agent, the pallet identity agent and the barcode reading agent. The agent has its barcode information database. The robot control agent is responsible for the provision of the local robot manipulation. The agent also assigns the required control program to the robot to meet the order requirements.

5. Object-Oriented Analysis and Design Methodology

The object-oriented approach is adopted in the development of the agent-based FAC control system. Basically, agents in the system are modelled by specific object relationship, class attribute and behaviour of the physical components or operating functions in the FAC. With the help of the dynamic model and the functional model, a proposed framework of the agent system is built. The agent relationships give the indication for the construction of the object model. The object model provides the program's definition on the objects or classes required for the agents. A group of objects or classes may represent an individual agent but the object model provides the basic object units definition for each element inside the cell. The object model in Figure 4 provides more practical information for the connection between the devices and the logical or controlling units inside the flexible manufacturing cell. The model shows the set of objects exist in the system, together with their corresponding relationship, attributes and operations. As object classes are the essential element in the OMT, this can show the static view of the overall agent system and show relationship of dependency between classes. The structure of the object model shows the preliminary infrastructure of agents. The necessary agents and objects are clearly shown during the system development stage. Additional agents can be built at any time as objects in the system program with proper message interface.

6. Implementation of the Agent System

In the development of this agent-based control system, both JAVA and Visual basic are used. They are object-oriented programming languages. Object-oriented programming centers on identifying and working with abstraction. Abstraction allows a concept or idea to be expressed and then repeatedly used without knowing the full details inside the object. The agent template is the generic agent program structure for the agents throughout this FMS control system. The object or class creation for the agents are similar in object-oriented programming languages.

Corresponding to the implementation results of the agent infrastructure of the project, a formalized agent structure in the program code is summarized. In this FMS control system,

each agent comprises 6 important layers. They are communication layer, data layer, preceptor layer, effector layer, agent's logic layer and error handling layer. Communication layer deals with the TCP/IP communication task with other agents. Preceptor and effector layers are responsible for the control of the real time shop floor information. The Agent's logic layer is the "brain" of the agents. The degree of the autonomous of the agents depends on this layer design structure. Intelligent object can be added inside this layer to increase the flexibility of the agents. Besides, this layer handles rapid responses to the dynamic changes to the shop floor environment of the agent. Error handling such as machines breakdown or other undesired events are handled by the error handling layer of the general agent structure. Java programming language format is used to illustrate the mapping of individual agent structure to the object-oriented programming structure. Additional layers are reserved for other purposes in the future development of the agent.
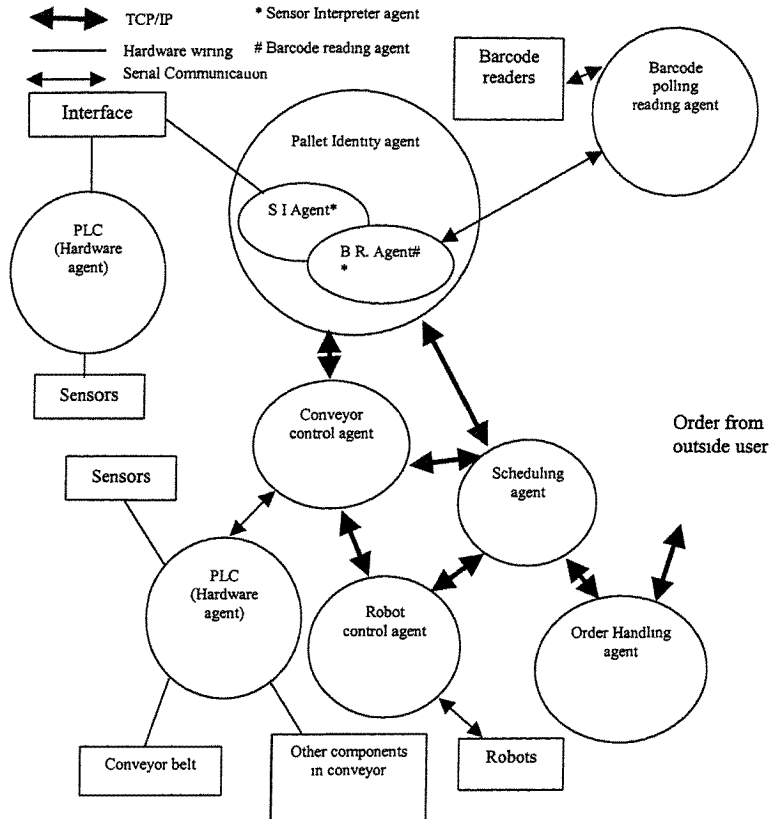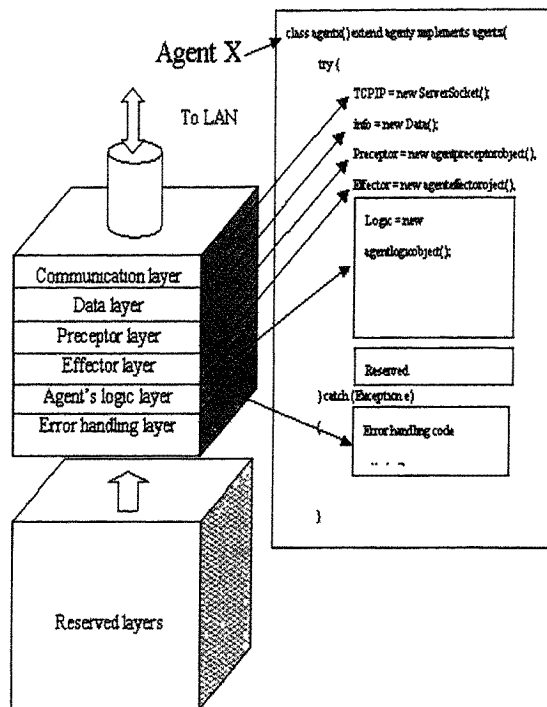

Figure 4. Agents Relationships

Figure 5 Generic Agent Template to object-oriented language (Java) structure

Figure 5 shows an example on the mapping of the agent design structure to the object oriented (Java) programming structure. For a new agent, a name should be given for identification. In the example, the name of the new agent is "agentx". "agentx" can be inherited from an existing agent, "agenty". That is, agenty is the parent of the new agent. All the characteristics of the parent agent can be re-used by this new agent. Additional common structure of the agent's methods can also be shared by using the keyword "implements". In this example, "agentx" shares the methods of "agentz".

Apart from sharing and inheritance of properties from other agents, it is also possible to create new methods for agent with unique functionalities in the control system. The data layer, preceptor layer and effector layer can be inherited by "agentx" using the object creation syntax,

New_object_name = new exist_layer_template_object;

Using this syntax, new objects will be created from the template exist_layer_template. The newly created object exists in the new agent only and it is independent from other agents. The contents, program logic and execution of the new agent's object will be executed without interference with other agents inherited from the same layer template.

7. Conclusion

In the current implementation of the prototype multi-agent system, agents are installed in different controlling computers in a LAN through the TCP/IP interface. In the corresponding agent control computers, graphic user interfaces (GUI) are established to provide the real-time status of the FAC and the agents. Using the object-oriented models in the design and analysis stages, necessary models and interfaces are determined before the actual program codes are developed. This can help to reduce the redundant codes at the development stage. This can also reduce the investment cost because most of the program design problems are found at the early stage of program development. The MAS has demonstrated its flexibilities in different cases in dealing with different configurations of example work-pieces. Fast response to the system preceptors (sensors, robot controller messages) is also shown by using distributed agents in the FAC.

For future research, the agent-based infrastructure can be expanded to a more complex FMS control system. More objects or agents can be added to the control system. The existing agent infrastructure can be implemented and tested thoroughly by more agents implemented into the control system. For instance, it is possible to add in the process planning agent, product design agent, and machining agent to cope with additional planning and control functions and activities.

REFERENCES

1. **BASRAN, J.S.; PETRIU, E.M.; PETRIU, D.C.** Flexible agent-based robotic assembly cell. Proceedings of the 1997 IEEE International conference on Robotics and Automation, 1997, 3461-3466

2. **NAGATA, A.; HIRAI, J.** Distributed planning for assembly tasks by multiple manipulators. Proceedings of the IEEE International conference on Robotics and Automation, 1994, 3522-3529

3. **FRANKLIN, S.; GRASSER, A..** *Is it an agent, or just a program?* Proceedings of the third international workshop on Agent theories, architectures and languages Springer-Verlag, 1996, 54-66

4. **PARUNAK, VAN, WARD.** *Distributed A.I. and Manufacturing Control: some issues and insights.* In Y. Demazeau and J. Muller, eds., Decentralized A.I., North-Holland, 1990, 81-104

5. **OUELHADJ. D.; HANACHI, C.; BOUZOUIA. B.** *A multi contract net protocol for dynamic scheduling in flexible manufacturing systems.* ICRA'99 IEEE International Conference on Robotics and Automation, Detroit, Michigan, USA, 1999.

6. **DUMITRACHE, I.; CARAMIHAI, S.I.; STANESCU A.M.,** *Intelligent Agent-based Control Systems in Manufacturing.* Proceedings of the 15th IEEE International Symposium on Intelligent Control, Rio, Patras, Greece. 17-19 July, 2000 (ISIC 2000)

7. **FRIEDRICH, HOLGER, ROGALLA, OLIVER AND DILLMANN, RUDIGER,** *Integrating skills into multi-agents systems.* Journal of Intelligent Manufacturing 1998, 9, 119-127

8. TSO, S.K., LAU H.C.W., HO, J.K.L., ZHANG W.J. *A framework for developing an agent-based collaborative service-support system in a manufacturing information network.* Engineering Applications of Artificial Intelligence 12, 1999, 43-57

9. SIEMIATKOWSKI, MIECZYSLAW, PRZYBYLSKI, WIODZIMIERZ, *A system solution for integrating of process planning and control in flexible manufacturing. Integrated Manufacturing Systems.* Integrated Manufacturing Systems 1997, 8/3 173-180

10. MONOSTORI, L.; KADAR, B. *Agent-based Control of Manufacturing Systems.* Intelligent Processing and Manufacturing of Materials, 1999. IPMM '99. Proceedings of the Second International Conference on Volume: 1 , 1999 , Page(s): 131 -137 vol.1

11. D'AMBROSIO, JOSEPH; DARR, TIMOTHEY; BIRMINGHAM, WILLIAM. *Hierarchical Concurrent Engineering in a Multiagent Framework.* Concurrent Engineering: Research and Applications, March, 1996, Vol.4 No.1 47-57

12. MATURANA, F.P., NORRIE, D.H. *MetaMorph: an adaptive agent-based architectures.* International Journal of Intelligent Manufacturing, 1996, Vol.7 257-270

13. FRALIE, JUAN-CARLOS; PAREDIS; CHRISTIAAN J.J; WANG CHENG HUA; KHOSLA, PRADEEP K., *Agent-Based Planning and Control of a Multi-Manipulator Assembly System.* Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on , Volume: 2 , 1999 1219 -1225

14. REVELIOTIS, S.A. *Production planning and control in flexibly automated manufacturing systems: current status and future requirements.* Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on ,Volume: 2 , 1999

15. GAMBIN, A.J.; PIERA, M.A.; RIERA, D. *A Petri nets based object oriented tool for the scheduling of stochastic flexible manufacturing systems.* merging Technologies and Factory Automation, 1999. Proceedings. ETFA '99. 1999 7th IEEE international Conference on , Volume: 2 , 1999 1091 -1098

Integrating intelligent agents with legacy manufacturing information systems

C W LEUNG and T N WONG
Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong,
Hong Kong

ABSTRACT

The paper presents a framework of an agent-based system which aims to enrich the supply chain management functions with existing legacy manufacturing information systems. The key idea is to establish a distributed, reusable and extensible agent template which can be readily deployed into any company-wide Information Technology (IT) architecture. The object-oriented system design concept has been adopted in the framework, while the encapsulated template will be implemented using Java and XML standards for constructing the applications and message structures respectively.

The agent template is being used in an agent-based order integration system. This system is able to link up with the legacy information systems and provide a channel for order bidding and order integration between the buyer's and supplier's systems. With an involvement of multiple systems throughout the processes, the proposed design will reduce the time and resources in agent construction and implementation.

1   INTRODUCTION

According to Smith et al. (1), company-wide manufacturing information systems such as Enterprise Resources Planning (ERP) systems exhibit some inherent problems. For instance, most ERP systems lack software mobility, and they are not designed to support e-business in the domain of supply chain management. In view of this, they have designed a web-based system COMIREM (Continuous, Mixed-Initiative Resource Management) for collaborative resource management which provides direct support for special operation forces planning.

In recent years, numerous researchers have attempted to apply agent technology to manufacturing enterprise integration, supply chain management, manufacturing planning, scheduling and control, materials handling, and holonic manufacturing systems. With agent technology, it is possible to open up a new dimension of behaviours and capabilities for manufacturing information systems to make them more competent in supporting today's

business and manufacturing activities. Wortmann et al. (2) further summarized that the agent-based technologies (mainly featured by their empowerment and cooperative interaction) will bring a new paradigm shift in the conceptualization and design of future information and communication technology systems. They point out that enterprise-to-enterprise integration will be a painful experience to many companies. However, companies which pursue this goal will finally benefit from creating new opportunities. For example, companies which carry out a normal commercial transaction will avoid checking and retrieving data twice or even more in their systems. This type of inter-organizational integration is of great value to many companies which have employed an ERP system.

One main application of agent technology is in the integration of standalone applications into highly distributed computational systems. A number of research projects focusing on system integration have been reported in literature recently (3) (4) (5). They have primarily focused on the agent-based system integration with some high-end ERP systems such as SAP, PeopleSoft, and Baan.

This paper describes an ongoing research on the integration of intelligent agents and manufacturing information systems. In consideration of Hong Kong industries and practices, an agent-based framework for customer order integration (6) has been proposed for local small and medium sized enterprises (SMEs). The framework ABOIS (Agent-Based Order Integration System) is established as an extended part of a legacy manufacturing information system, so that the entire system will be capable of generating bids and executing order-bidding. Moreover, it can also communicate with the supplier's systems so that exchange of bids and messages can proceed. The agent-based system will have a synergic effect on traditional information systems so that they can serve better in the era of e-commerce.

In this paper, the architecture ABOIS is reviewed and the objective of this system is described. Besides, the design of the generic agent template is also presented. This template structure is based on the analysis of the functionalities of each individual agent. With the generic agent template, the implementation of the whole system is simplified and enhanced.

2 SYSTEM OVERVIEW

The proposed agent-based framework is designed to bring a new paradigm to traditional information systems, in particular to the in-house or locally developed ones which have been implemented in SMEs for years. In reality, most of these systems are restricted to operating and supporting transactions and functions within an organization only. With a prevailing atmosphere of globalized business and supply chain management, the framework is expected to enable existing legacy information systems to couple with current proprietary systems from business partners. The focus of Agent-Based Order Integration System (ABOIS) is on the procurement process of a typical manufacturing company. This is based on a scenario between a customer-side Purchasing Department and multiple supplier-side Sales and Marketing Departments. The scope of the agent-based framework will be extended to other inter-organizational activities.

The procurement process of the customer in a typical supply chain network involves order acquisition, order-bidding with multiple suppliers and finally order transformation to an appropriate supplier's system. In the framework, the buyer-side agents aim to place the

orders to the appropriate suppliers. These two agents are (1) Customer Interface Agent (CIA) which queries and generates bids from the ERP database and (2) Manager Agent (MA) which represents the customer throughout the order-bidding process. The MA will invite bids from Bidder Agents (BA) which represent each supplier. After the bidding process, the winning BA connects with a Supplier Interface Agent (SIA) to transform the order to that supplier database.

Figure 1 depicts a typical flow among these agents, with an abstraction of the bidding process. As shown in the figure, each type of agents serves different areas throughout the order-integration process. The CIA and SIA are mainly concerned with the database connectivity with the customers and suppliers respectively. They will invoke the MA and BA, whereas their core functions are the bidding process and the agent-to-agent communication. Moreover, these agents will be mounted on different platforms in their host locations. The communication between systems should be platform-independent. Hence, it is a daunting task to construct each type of agents and then deploy the agents onto their corresponding hosts.
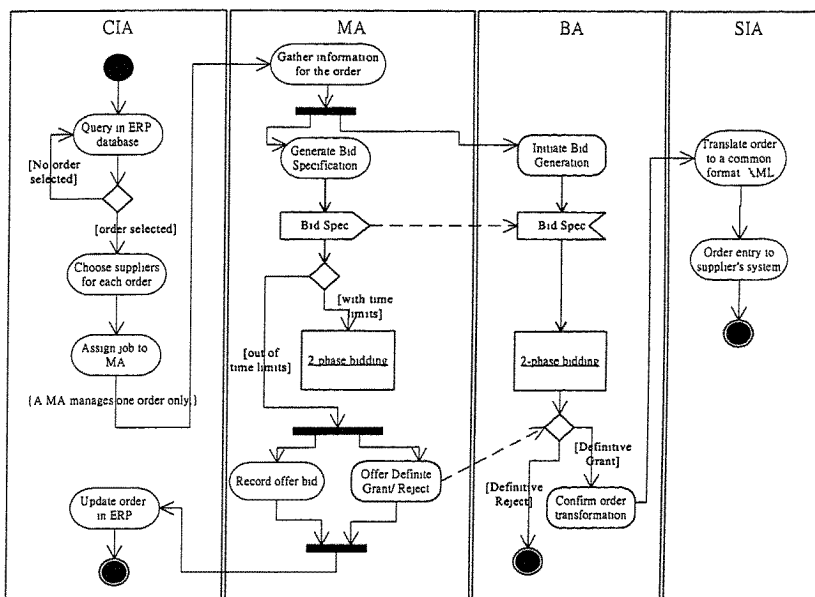

Figure 1. Typical Flow of ABOIS

Some researchers have raised similar issues in their agent-based system projects. Their common approach to address this problem is by developing some generalized agents which only keep a few configurations during implementation in order to fulfill the goals: Bullinger et al. (7) designed a Business Object (BO) model to interface with the legacy systems; Bui et al. (8) developed a generic software agent architecture which composes of an User-Agent Interface, Processing Engine, Procedure Repository and View; Hildum et al. (9) designed an Integrating Process Planning and Production Scheduling (IP3S) architecture which

demonstrates the reuse of knowledge sources across different architectures; Karacapilidis et al. (10) built two prototypes of e-market agents – Purchaser Agent and Seller Agent to take part in the electronic commerce. Other researchers have discussed the benefit of reusing agents. Brugali et al. (11) described that the effectiveness of agent technology concepts may be better exploited, if the development process is guided by a consolidated reuse methodology.

A generic agent is able to give a lot of conveniences in an agent-based system. Figure 2 illustrates the initial class diagram of the ABOIS, which demonstrates the number of agents and their links in the system. For example, each customer (one CIA) will generate an order for bidding (one MA). Assuming that this customer has sourced the purchase part from seven suppliers (seven SAs), whereas each of them has its own system (one SIA, because the ABOIS does not support order decomposition at this stage). It turns out that a single order will instantiate ten agents which may be located at different locations. To facilitate the implementation, a Generic Agent TEmplate (GATE) has been designed.
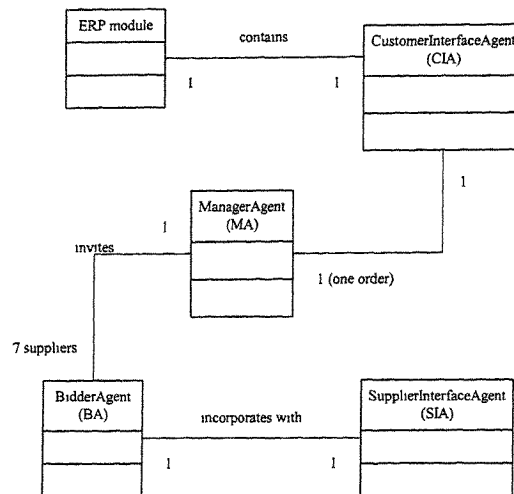

Figure 2. Class Diagram of ABOIS

3   AGENT STRUCTURE

Kendall et al. (12) studied the relationship between agents and objects with a case study on discrete parts manufacturing. They summarized the similarities as: abstraction, encapsulation, identity, state, behaviour and inheritance. In our proposed ABOIS, GATE is developed based on an object-oriented (OO) design method. The basic idea of designing this agent is to make it possible to "plug" into any platform and be used to accomplish the order-integration process. For these reasons, the abstraction, encapsulation and inheritance behaviours of OO-concept are important to GATE. Table 1 summarizes an analysis of the functional components of each individual agent in the architecture of ABOIS. The analysis is helpful to identify the core components of GATE.

It is apparent that the above classification has distinguished the four individual agents into five modules. Figure 3 depicts the core component model of GATE, based on the static diagram of the Unified Modeling Language (UML). GATE is actually a class of an abstraction of a group of the five modules (objects). Class users can easily instantiate the necessary objects, and do not need to know the internal implementations.

A deployment diagram showing the configuration of the agent-based system defining in terms of GATE and other elements is presented in Figure 4.

Table 1. Analysis of Agents in ABOIS

|	CIA	MA	BA	SIA
Graphical User Interface (GUI)	- Users select orders and vendors	- Users modifies on the bid spec	- User modifies replying bid	- An acknow-ledging message displayed
Database Connection	- Query buyer-side relational database - Update order details after bidding	- Query buyer-side database for gathering order info.	- Query supplier-side database for gathering bidding history	- Update Supplier-side database for order importing
Communication (Both message transfer and file-type transfer)	- Initiate MA with the selected orders	- Dispatch bid spec. to the selected BAs - Acknowledge BAs with grant / reject offer after bidding - Reply CIA the updated order status	- Reply MA with bid - Reply SIA with granted bid	N/A
XML Parsing	- Interpret Bid Spec from MA	- Generate Bid Specification into XML format (from SQL recordset)	- Interpret Bid Spec from MA - Generate Bid	- Interpret granted Bid Spec from BA
Business Logic	N/A	- Select an appropriate bid(s) based on predefined rules	- Determine "To bid" or "Not to bid"	N/A
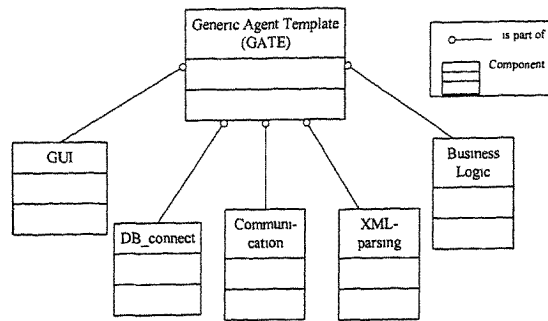
Figure 3. Core Components of GATE
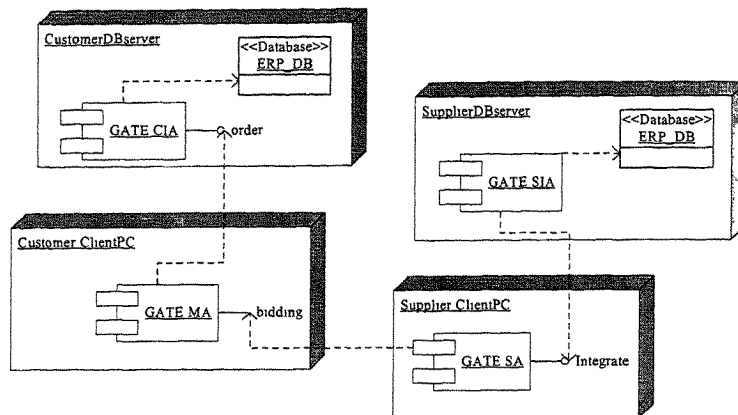
Figure 4. Deployment Diagram of GATE

4 IMPLEMENTATION

The proposed framework and agents are being implemented in the Java platform, which consists of the Java Virtual Machine (JVM) and Java Application Programming Interface (JavaAPI), thus ensures that the agents can be run on any platform (due to JVM). In addition to traditional programming language, there exist some powerful agent development tools. Examples are ABLE (IBM; www.alphaworks.ibm.com/tech/able), AgentBuilder (Reticular Systems, Inc.; www.agentbuilder.com), Aglets (IBM, Japan; www.trl.ibm.co.jp/aglets), Concordia Java2 (Mitsubishi; www.merl.com/projects/), JAFMAS (University of Cincinnati; www.ececs.uc.edu/~abaker/JAFMAS), JATLite (Stanford University;

http://java.stanford.edu/) and ZEUS (British Telecommunications Labs; www.labs.bt.com/projects/agents.htm). All of these tools are purely Java-based and being developed by academic or industrial organizations.

In this study, the tool Agent Building and Learning Environment (ABLE) has been selected for building our agents, especially for the bidding process. ABLE is a 100% java-based tool kit for building intelligent agents. There are component libraries for rule-based inferencing using Boolean and fuzzy logic, and for machine learning techniques such as neural networks etc. With ABLE as a foundation developing environment, the bidding process can be simplified.

Another important implementation issue concerns the standard and requirement of information exchange. Traditionally, the EDI (Electronic Data Interchange) standards have been applied in applications relating to business transactions, such as in e-commerce and supply chain management. One of them is UN/EDIFACT (United Nations Electronic Data Interchange for Administration, Commerce and Transport). However, the EDI protocols are too complex and costly to use (13). In recent years, the emerging XML (eXtensible Markup Language) has been widely accepted as a standard for information exchange. The XML standard allows users to customize the presentation and description of the data. This is particularly useful for generating bids and sharing them with other business partners easily in the Internet environment.

In the implementation of ABOIS, the orders, bids and messages being transmitted will take an advantage of the XML properties such that the attributes of them will be extended freely. As a result, the order-bidding process can be changed according to the bid format. Besides, in the final step of the flow, the confirmed bid could be easily converted into the recordset in the database.

5 CONCLUDING REMARKS

This paper describes an ongoing research on integrating intelligent agents with legacy information systems. Current development work of the proposed framework, ABOIS, is focused on the order-integration scenario in the supply chain. It is believed that the framework will bring a instant improvement to the systems which will be finally beneficial to the SMEs, in terms of minimizing their investment in upgrading their traditional systems. In future, the agent-based approach will be extended to other supply chain management activities.

REFERENCE

[1] **Smith, S. F., Hildum, D. W.,** and **Crimm, D.** (2001) Toward the design of web-based planning and scheduling services. Proceedings of the ECP-01/ Planet Workshop on Automated Planning and Scheduling Technologies in New Methods for Electronics, Mobile and Collaborative Work.

[2]   **Wortmann, H., and Szirbik, N.** (2001) ICT issues among collaborative enterprises: from rigid to adaptive agent-based technologies. Production Planning & Control, Vol. 12, pp. 452-465.

[3]   **Parunak, H. V. D.** (1998). Practical and industrial applications of agent-based systems. Industrial Technology Institute (www.erim.org/~vparunak/apps98.pdf).

[4]   **Kwon, O. B., and Lee, J. J.** (2001) A multi-agent intelligent system for efficient ERP maintenance. Expert Systems with Applications, Vol. 21, pp. 191-202.

[5]   **Turowski, K.** (2002) Agent-based e-commerce in case of mass customization. International Journal of Production Economics, Vol. 75, pp. 69-81.

[6]   **Wong, T.N., and Leung, C. W.** (2002) An agent-based framework for customer order management. Proceedings of the 30th international Conference on Computers & Industrial Engineering.

[7]   **Bullinger, H. J., Fahnrich, K. P., and Linsenmaier, T.** (1998) A conceptual approach for an architecture of distributed objects for the integration of heterogeneous data processing systems in manufacturing companies. International Journal of Production Research, Vol. 36, pp. 2997-3011.

[8]   **Bui, T., and Lee, J.** (1999) An agent-based framework for building decision support systems. Decision Support Systems, Vol. 25, pp. 225-237.

[9]   **Hildum, D. W., Sadeh, N. M., Laliberty, T. J., McA'Nulty, J., Smith, S. F.,** and **Kjenstad, D.** (1997) Blackboard agents for mixed-initiative management of integrated process-planning/ production-scheduling solutions across the supply chain. Proceedings of the Ninth Conference on Innovative Applications of Artificial Intelligence.

[10] **Karacapilidis, N., and Moraitis, P.** (2001) Building an agent-mediated electronic commerce system with decision analysis features. Decision Support Systems, Vol. 32, pp. 53-69.

[11] **Brugali, D, and Sycara, K. P.** (2000) Towards agent-oriented application frameworks. ACM Computing Surveys, Vol. 32, pp. 21.

[12] **Kendall, E. A., Malkoun, M. T., and Jiang, C. H.** (1997) The application of object-oriented analysis to agent based systems. Journal of Object-oriented Programming, Vol. 5, pp. 45-52.

[13] **Aberdeen Group.** (1998) Tightening the links in the extended supply chain: the extraordinary potential of XML. An Executive White Paper, Aberdeen Group Inc. (www.aberdeen.com)

Incremental induction based on logical network

R-L SUN, Y XIONG, and H DING
School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, People's Republic of China

SYNOPSIS

When used incrementally, generalized learning may result in inconsistency and inaccuracy for concept learning. This paper presents a logical network that is used for incrementally inductive learning. Induction is treated as two phases: accurate learning and generalization. The algorithm is based on logical operations thus it is solid and comprehensive. Also it is parallelism therefore it may be computational efficient.

1 INTRODUCTION

Dispatching rules have played a significant role within the dynamic scheduling because of their ease of implementation and compatibility with the dynamic nature of manufacturing systems (Chan, 1999; Newman and Maffei, 1999; Park et al., 1997). Dispatching rules are essentially sequencing ones. The job with the highest priority is selected and sent to an idle machine for processing. Panwalker and Iskander (1977), Blackstone, Phillips and Hogg (1982) provided often-quoted survey articles about known research results with sequencing rules. Studies have shown that the selection of the dispatching rules has significant impact on system performance (Crowe and Stahlman, 1995). Given its importance, there has been extensive research done in identifying dispatching rules that are superior. The major finding is that there is no one rule that clearly dominates others (Park et al., 1997). This is because the performance of a dispatching rule depends on the shop configuration, production tasks and scheduling objective (Goyal et al., 1995; Baker, 1984), which means one should select a suitable dispatching rule according to a particular manufacturing system.

Adaptive scheduling (Park et al, 1997; Jahangririan and Conroy, 2000) steps forwards along with the direction. It adopts scheduling knowledge, which is normally expressed as "*if p then r*", to select a dispatching rule r according to the current state p of manufacturing systems. Knowledge base (KB) plays a central role in this scheduling paradigm. A typical method to obtain the scheduling knowledge about a particular manufacturing system is to consult from scheduling specialists and then formalize their expressions. Generally speaking, this is a

difficult task known as the *knowledge acquisition bottleneck*. Machine learning, however, provides a promoting solution to the problem, which intends to automatically acquire knowledge about a domain problem from experience.

AQ (Hong, 1997) algorithm is one of the famous learning programs. It based on logical operations thus it is solid and comprehensive to human being. A core concept used in the AQ algorithm is *star*. Suppose that E^+, E^- represent positive and negative event set, respectively. $E^+ \cap E^- = \varnothing$. Let $e \in E^+$ be a positive event. The star of the event e against the E^-, denoted by $G(e|E^-)$, is the set of all *maximally general conplexes* covering the e and not covering any event in the E^- (Hong, 1997; Michalski, 1983). According to the AQ algorithm, the star $G(e|E^-)$ not only covers the event e and other positive events belonging to the E^+, but also covers the events occurring neither in the E^+ nor in the E^-. That is, the star being the result of induction covers some unknown events. It is a generalized form of learning result and may include some learning error. In the AQ algorithm, generalization is performed during the formation of the star. Generalization and learning are tightly coupled each other. Such a learning scheme is called *generalized learning* so as to be distinguished from the *accurate learning* scheme presented in this article.

Non-incremental learning systems may not be computationally efficient. Thus incremental learning fashion attracts extensive researchers (Bloedorn and R. S. Michalski, 1998; Maloof and Michalski, 1995; Michalshi et al, 1986). As the star lies at basis of all incremental learning systems of the AQ family, the generalized learning is a basic learning scheme used by the family. In the incremental learning fashion, however, the generalized learning has the following shortcomings. (1) Learning accuracy. Generalization is such a procedure that may introduce learning error. For the incremental learning fashion, new epoch of inductive learning is based on the result of last learning epoch. It starts with those that has some learning error introduced in previously learning. The new learning epoch itself will also introduce new error. Consequently the incremental learning is a procedure that accumulates learning errors, which may decrease accuracy of the learning systems. (2) Inconsistency. The learned result is a generalized one so it may contain some learning error. Although each epoch of learning can maintain the concept consistency over the observed events being learned, the new observed events used in the following learning epochs may conflict to the concept already learned. Therefore consistency maintenance becomes one of the central problems in the incremental learning fashion.

Neural network learning is essentially based on feature-similarity. Parallelism, non-linear mapping and noise tolerance are the major merits of this paradigm. Implicit representation of knowledge and difficulty of architecture design are the principal shortcomings. Additionally, background knowledge can hardly be used in learning process, as well as in architecture design. How to integrate the advantages of the above learning algorithms and exclude their disadvantages provides a non-trial challenge to the researchers involved in machine learning area. This paper focuses on logic-based parallel and incremental induction

The remainder of the paper is organized as follows. First in the next section, preliminaries and methodology are briefly introduced. In section 3, logical network is defined, which can be easily derived from and transformed to a logical expression. Inductive learning algorithm and generalization algorithm are proposed in section 4 and 5, respectively. Finally, in section 6, we conclude the whole paper.

2 PRELIMINARIES

Let $E = D_1 \times D_2 \times \cdots \times D_n$ be n-dimension logical space. E^+ and E^- are subsets of E and satisfy:

$$E^+ \cap E^- = \varnothing, E^+ \cup E^- = E$$

where $\varnothing$ represents empty set; E^+ and E^- represent positive and negative example set, respectively.

$e \in E$ is called an example (or an event) which is denoted by

$$e = (u_1, \ldots, u_n)^T$$

where $u_j \in D_j, j = 1, \ldots, n$, represents attribute value of example e. If $e \in E^+$, e is called a positive example and is denoted by s. If $e \in E^-$, e is called a negative example and is denoted by c.

Inductive concept learning is then formalized as following.

Given: Background knowledge KB (expressed as logical expressions) and training example set E^+ and E^-.

Find: (1) A formula F^+ covering all examples in E^+ and excluding any example in E^-;
(2) A formula F^- covering all examples in E^- and excluding any example in E^+.

Scheduling knowledge can be expressed as "*if p then r*", where p denotes conditions that system state should satisfied and r represents the dispatching rule adopted under the conditions (Park *et al*, 1997; Jahangririan and Conroy, 2000). In dispatching rule-based scheduling, p is a logical expression. Any logical expression can be transformed to disjunctive norm form (DNF) according to Boolean Algebra. Therefore it is assumed that any logical expressions used in this paper are with the DNF.

As the generalized learning scheme has some limitations when used incrementally, we divide the inductive learning procedure into two separated phases: *accurate learning* and *generalization*. The aim of accurate learning is to record the examples exactly and simplify the learned results. This procedure emphasizes the logical equivalency between the learned results and the examples. The results obtained are called *accurate concepts*. After all the examples available at the current learning epoch have been learned, the generalization procedure starts and the *generalized concepts* are obtained. The learning system keeps both the accurate concepts and the generalized concepts. When new examples are available, new learning epoch starts with the accurate concepts and learns the new examples in the incremental fashion. The updated accurate concepts are then to be generalized. Because of logical equivalency between the accurate concepts and the examples, the proposed incremental learning model can certainly maintain the consistency over all examples and may achieve higher accuracy. In such a learning scheme, two important aspects should be taken into account. It should be guaranteed that knowledge already learnt does not be destroyed when learning new examples. On the other hand, if new examples being learnt conflict to the knowledge already learnt, there is an easy way to emend the original knowledge. Finally we wish to solve the incremental induction with network so as to benefit from its parallelism.

3 LOGICAL NETWORK

Consider a forward neural network with only one hidden layer $Net(\cdot)$: $\{0,1\}^n \rightarrow \{0,1\}$. It performs a mapping from n-dimension logical space to one-dimension logical space.

The mapping from the input logical space to the hidden layer is denoted by:

$$h(\cdot): \{0,1\}^n \rightarrow \{0,1\}^m$$

where:

$$h = (h_1, h_2, ..., h_m)^T$$

$$h_i(\cdot) = \begin{cases} 1, & W_i^T x - \theta_i \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$W_i = (w_{i1}, w_{i2}, ..., w_{in})^T$$

$$w_{ij} \in \{-1, 0, 1\}$$

$$\theta_i \in \{0, 1, 2, ..., n\}$$

$$i = 1, 2, ..., m$$

$$j = 1, 2, ..., n$$

In the formula w_{ij} represents connection weight from the input neuron j to the hidden neuron i; θ_i represents the threshold of hidden neuron i; x represents n-dimension input logical vector.

The mapping from hidden neurons to the output neuron is demoted by $O(\cdot)$: $\{0,1\}^m \rightarrow \{0,1\}$.

$$O(\cdot) = \begin{cases} 1, & (1^m)^T h - 1 \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where $1^m = (1, 1, ..., 1)^T$ represents a m-dimension vector which elements are all equal to one. Such a network is called *logical network*. In the logical network, a hidden neuron is uniquely determined by the weights and threshold therefore it can be expressed by

$$h_i(\cdot) = (w_{i1}, w_{i2}, ..., w_{in}; \theta_i) = (W_i^T; \theta_i)$$

It is easy to show that given any logical expression p in DNF, there exists a logical network $Net(\cdot)$ such that $p \leftrightarrow Net(\cdot)$ is a tautology, i.e., $p = Net(\cdot)$. And vice versa, given any logical network $Net(\cdot)$, there exists a logical expression p in DNF such that $Net(\cdot) \leftrightarrow p$ is a tautology, i.e., $Net(\cdot) = p$.

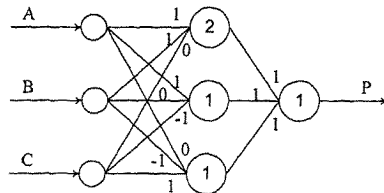Figure 1 presents a logical network corresponding to a NDF: $P = AB + A\overline{C} + \overline{B}C$.

Figure 1 The logical network corresponding to $P = AB + A\overline{C} + \overline{B}C$

4 LEARNING ALGORITHM

Defining operator $\triangleright(\cdot)$, which is used to transform a DNF to a logical network, as follow.

$$\triangleright(x) = \begin{cases} -1, & x = 0 \\ 0, & x = * \\ 1, & x = 1 \end{cases}$$

where $*$ represents *don't care term*, which means that the corresponding attribute does not appear in the DNF. If x represents a vector which consists of 0, 1 and $*$, $\triangleright(x)$ will operate on each element of x and then obtain a vector.

Operator $Ch(\cdot)$, which is used to create a new hidden neuron i according to an example e, is defined as following.

$$h_i(\cdot) = Ch(e) = (w_{i1}, w_{i2}, ..., w_{in}; \theta_i) = (W_i^T; \theta_i)$$
$$e = (u_1, u_2, ..., u_n)^T$$
$$W_i = \triangleright(e)$$
$$\theta_i = n\text{-}d_H(1^n, W_i)$$

where $d_H(\cdot, \cdot)$ represents *Hamming Distance* of two vectors.

Two hidden neurons

$$h_1(\cdot) = (w_{11}, w_{12}, ..., w_{1n}; \theta_1) = (W_1^T; \theta_1)$$
$$h_2(\cdot) = (w_{21}, w_{22}, ..., w_{2n}; \theta_2) = (W_2^T; \theta_2)$$

can be merged into one hidden neuron, say neuron i

$$h_i(\cdot) = (w_{i1}, w_{i2}, ..., w_{in}; \theta_i) = (W_i^T; \theta_i)$$

if $d_H(W_1, W_2) = 1$. Weights and threshold of the new hidden neuron are determined respectively as follows.

$$w_{ij} = \begin{cases} w_{1j}, & w_{1j} = w_{2j} \\ 0, & \text{otherwise} \end{cases}, \quad j = 1, 2, ..., n$$

$$\theta_i = n\text{-}d_H(1^n, W_i)$$

That two hidden neurons are merged into one is denoted by $Mh(\cdot, \cdot) = h(\cdot)$

Let $h_i(\cdot)$ be a hidden neuron corresponding to positive example(s). Let c be a negative example.

$$h_i(\cdot) = (w_{i1}, w_{i2}, ..., w_{in}; \theta_i) = (W_i^T; \theta_i)$$
$$c = (c_1, ..., c_n)^T$$

$h_i(c) = 1$ means that knowledge already learnt by $h_i(\cdot)$ conflicts to current negative training example c. Suppose that c is true therefore $h_i(\cdot)$ must be emended according to the negative example c.

Suppose that $d_H(\triangleright(c), W_i) = k$, $0 \le k \le n$. Let $c^1, c^2, \ldots, c^k$ be the components of c, each of which satisfies

$$c^l \in \{c_j \mid \triangleright(c_j) \ne w_{ij}, j = 1, 2, \ldots, n\}$$

$$l = 1, 2, \ldots, k$$

Delete the hidden neuron $h_i(\cdot)$ from the network and construct k hidden neurons as follows.

$$W_1 = (\ldots \triangleright (c^1) \ldots \triangleright (c^2) \ldots \triangleright (c^{k-1}) \ldots \triangleright (\overline{c^k}) \ldots)^T$$

$$W_2 = (\ldots \triangleright (c^1) \ldots \triangleright (c^2) \ldots \triangleright (\overline{c^{k-1}}) \ldots 0 \ldots)^T$$

$$W_{k-1} = (\ldots \triangleright (c^1) \ldots \triangleright (\overline{c^2}) \ldots 0 \ldots 0 \ldots)^T$$

$$W_k = (\ldots \triangleright (\overline{c^1}) \ldots 0 \ldots 0 \ldots 0 \ldots)^T$$

$$\theta_l = n - d_H(1^n, W_l), l = 1, \ldots, k$$

where ... indicates that weights of the new hidden neurons are as the same as the corresponding weights of the deleted neuron $h_i(\cdot)$. That a hidden neuron $h(\cdot)$ is emended according to a negative example c can be denoted by

$$Eh(c) = h_1(\cdot) + h_2(\cdot) + \ldots + h_k(\cdot)$$

Accurate learning algorithm

1. Construct a logical network $Net(\cdot)$ according to the background knowledge KB (a DNF).
2. Select a training example e and remove it from training sample set E.
3. If $e \in E^+$ and $Net(e) = 1$, or $e \in E^-$ and $Net(e) = 0$, go to step 7.
4. If $e \in E^+$ and $Net(e) = 0$, create a new hidden neuron $Ch(e)$ according to the positive example e and then go to step 6
5. If $e \in E^-$ and $Net(e) = 1$, emend the hidden neuron $h(\cdot)$ activated by the negative example e according to $Eh(e)$.
6. Merge hidden neurons of the new network if possible.
7. If the training sample is empty, terminate learning; else go to step 2.

It will be notified that we can obtain two logical networks. One corresponds to the positive examples and the other to the negative examples, which are hereafter denoted as $Net^+(\cdot)$ and $Net^-(\cdot)$, respectively. $Net^+(\cdot)$ is called positive concept while $Net^-(\cdot)$ is called negative concept.

5 GENERALIZATION

The $Net^+(\cdot)$ obtained last section is logically equivalent to the example set E^+ while $Net^-(\cdot)$ is logically equivalent to E^-. Now let us deal with the generalization. The strategy employed here is to generalize both $Net^+(\cdot)$ and $Net^-(\cdot)$ alternately.

Consider two hidden neurons.

$$h_1(\cdot) = (w_{11}, w_{12}, \ldots, w_{1n}; \theta_1) = (W_1^T; \theta_1)$$

$$h_2(\cdot) = (w_{21}, w_{22}, \ldots, w_{2n}; \theta_2) = (W_2^T; \theta_2)$$

if $|w_{1i}-w_{2i}| \leq 1$, w_{1i} and w_{2i} are *generalization equality*, which is denoted by $w_{1i} \Leftrightarrow w_{2i}$. Note that generalization equality does not have *transitivity*.

Generalization algorithm

Suppose that $Net^+(\cdot)$, $Net^-(\cdot)$ be logical networks with n input logical variables. $Net^+(\cdot)$ has m hidden neurons and $Net^-(\cdot)$ has l hidden neurons. $Net^+(\cdot)$ and $Net^-(\cdot)$ represent accurate concept obtained from accurate learning algorithm, respectively. w_{ik}^+ represents weight from kth input to ith hidden neuron of $Net^+(\cdot)$. w_{ik}^- represents weight from kth input to ith hidden neuron of $Net^-(\cdot)$. Generalization algorithm is described as following.

1. If $\exists k \in \{1, 2, ..., n\}$, such that: $\forall i, j \in \{1, 2, ..., l\}$, $w_{ik}^- \Leftrightarrow w_{jk}^-$.
   And if $\exists h \in \{1, 2, ..., m\}$, such that $\forall i \in \{1, 2, ..., l\}$, $w_{hk}^+ \Leftrightarrow w_{ik}^-$.
   Then perform generalization operation: $w_{hk}^+ = 0$.

2. If $\exists k \in \{1, 2, ..., n\}$, such that: $\forall i, j \in \{1, 2, ..., m\}$, $w_{ik}^+ \Leftrightarrow w_{jk}^+$.
   And if $\exists h \in \{1, 2, ..., l\}$, such that: $\forall i \in \{1, 2, ..., m\}$, $w_{hk}^- \Leftrightarrow w_{ik}^+$.
   Then perform generalization operation: $w_{hk}^- = 0$.

Repeat step 1 and 2 until $Net^+(\cdot)$, $Net^-(\cdot)$ can not be generalized any more.

6 CONCLUSIONS

Generalized learning may result in inconsistency and inaccuracy for concept learning when used incrementally. This paper presents a logical network that is used for incrementally inductive learning. Induction is treated as two phases: accurate learning and generalization. The algorithm is based on logic operations thus it is solid and comprehensive. Independent learning and generalization processes, which guarantee consistency and accuracy during incremental learning, make the research different from other related work. Also it is parallelism therefore it may be computational efficient.

Although the learning scheme has the above merits, some problems remain unsolved. Further research can be concentrated on optimisation of network architecture and more reasonable generalization algorithm.

7 ACKNOWLEDGEMENT

REFERENCES

1. **Baker, K. R.** (1984) Sequencing rules and due-date assignments in a job shop. *Manag. Sci.* Vol.30, No.9, pp.1093-1104.

2. **Blackstone, J. J. H., Phillips, D. T. and Hogg, G. L.** (1982) A state-of-the-art survey of dispatching rules for manufacturing job shop operations. *Int. J. of Production Research*, Vol.20, No.1, pp.27-45.

3. **Bloedorn, E. and Michalski, R. S.** (1998) Data-driven constructive induction in AQ17-PRE: a method and experiments. *IEEE Intelligent Systems & Their Applications*, Vol.13, No.2, pp.30-37.

4. **Chan, F. T. S.** (1999) Evaluations of operational control rules in scheduling a flexible manufacturing system. *Robotics and Computer Integrated Manufacturing*, Vol.15, pp. 121-132.

5. **Crowe, T. J. and Stahlman, E. J.** (1995) A proposed structure for distributed shopfloor control. *Integrated Manufacturing Systems*, Vol.6, No.6, pp.31-36.

6. **Goyal, S. K.; Mehta, K.; Kodali, R. and Deshmukh, S. G.** (1995) Simulation for analysis of scheduling rules for a flexible manufacturing system. *Integrated Manufacturing Systems*, Vol.6, No.5, pp.21-26.

7. **Hong, J.** (1997) Inductive learning: algorithms, theory and applications. Science Press of China.

8. **Jahangririan, M., Conroy, G.** (2000) Intelligent dynamic scheduling system: the application of genetic algorithms. *Integrated Manufacturing Systems*, 11(4), 247-257.

9. **Maloof, M. A. and Michalski, R. S.** (1995) A method for partial-memory incremental learning and its application to computer intrusion detection. *Proceedings of Seventh International Conference on Tools with Artificial Intelligence*, pp.392 –397.

10. **Michalski, R. S. and Stepp, R. E.** (1983) Learning from observation: conceptual clustering. In Michalski, R. S. *et al* (Eds.), *Machine learning: an artificial intelligence approach*, pp.331-363, Tioga, CA.

11. **Michalski, R. S.; Mozetic, I.; Hong, J. and Lavrac, N.** (1986) The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. *Proceedings of the Fifth National Conference on Artificial Intelligence*, pp.1041-1045, Philadelphia, PA: Morgan Kaufmann.

12. **Newman, W. R. and Maffei, M. J.** (1999) Managing the job shop: simulating the effects of flexibility, order release mechanisms and sequencing rules. *Integrated Manufacturing Systems*, Vol. 10, No. 5, pp. 266-275.

13. **Panwalker, S. S. and Iskander, W.** (1977) A survey of scheduling rules. *Operations Research*, Vol. 25, pp. 45-61.

14. **Park, S. C., Raman, N. and Shaw, M. J.** (1997) Adaptive Scheduling in Dynamic Flexible Manufacturing Systems: A Dynamic Rule Selection Approach. *IEEE Trans. Robotics and Automation*, Vol. 13, No. 4, pp. 486-502.

Application of genetic algorithm to computer-aided process planning in distributed manufacturing systems

L LI, J Y H FUH, Y F ZHANG, and A Y C NEE
Department of Mechanical Engineering, National University of Singapore, Singapore

ABSTRACT

In a distributed manufacturing environment, factories possessing various machines and tools at different geographical locations are often selectively used to achieve the highest possible production efficiency. When jobs requiring a number of operations are received, several feasible process plans are produced from factories capable and available in meeting the specifications. Therefore, obtaining an optimal or near-optimal process plan becomes the main objective. This paper presents a modified genetic algorithm (GA), which, according to some prescribed criteria like minimizing processing time and/or cost, could swiftly search for the optimal process plan from several distributed manufacturing systems, and can perform multi-objectives optimization. One example is included to demonstrate the feasibility and robustness of the approach.

1. INTRODUCTION

Computer-aided process planning (CAPP), the key technique for computer-aided design and manufacturing (CAD/CAM) integration, aims at automating process planning tasks in an attempt to overcome some of the problems occurring in manual process planning. In general, process planning involves establishing the necessary manufacturing process and operation sequence in order to produce a given part economically and competitively.

The generation of optimal process plans for jobs is expressed using optimization or approximation approaches. Amongst those addressed approaches, genetic algorithms based on the mechanism of natural selection now became the most widely known type of evolutionary algorithm compared to the more traditional optimization methods. Especially when dealing with complex engineering optimization problems, genetic algorithm does outperform the other optimization approaches in many aspects.

2. DISTRIBUTED MANUFACTURING SYSTEMS

As described in Figure 1, in a distributed manufacturing environment, factories possessing various machines and tools at different geographical locations, and having different manufacturing capabilities are often selected to achieve the highest production efficiency. When jobs requiring several operations are received, the feasible process plans are generated by the available factories according to the precedence relationships of those operations. The final optimal or near-optimal process plan will emerge after comparison of all the feasible plans.
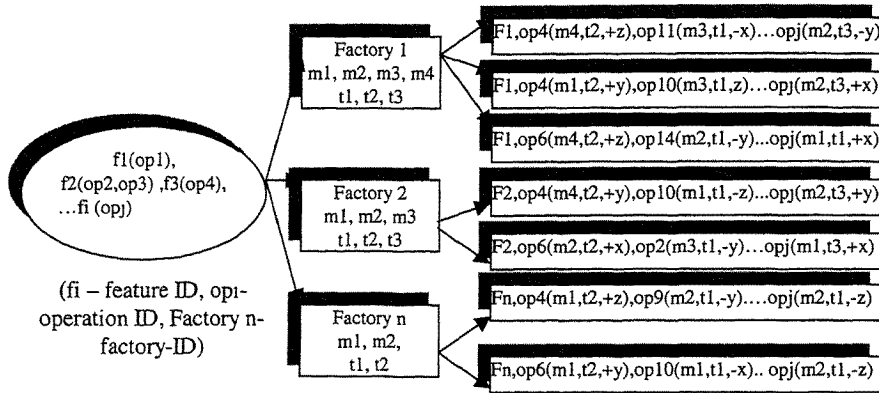
Figure 1. Description of a distributed manufacturing system

3. PROPOSED GENETIC ALGORITHM

3.1 Representation of process plans

When dealing with the distributed manufacturing system, a chromosome must not only represent the sequence of the operations but also indicate which factory this process plan should come from. Therefore, the identity number of the factory will be placed as the first gene of each chromosome. Each other gene comprises the operation ID and corresponding machine, tool and tool access direction (TAD). Figure 2 shows the representation of a six-operation process plan.

	Op4	Op1	Op5	Op6	Op3	Op2
001	m-02	m-03	m-02	m-01	m-03	m-01
	t-04	t-02	t-03	t-02	t-01	t-04
	+x	-y	+z	-x	+x	-y

Figure 2. Representation of a process plan

3.2 Reproduction

A genetic search starts with a randomly generated initial population; further generations are created by applying GA operators. This eventually leads to a generation of higher performing individuals.

3.2.1 Crossover

In this study, a crossover operator is designed by modifying the operator described in Bhashara et al. (1999). The implementation procedure of the crossover operation and an illustrative example are depicted below :

➤ Randomly choose two chromosomes as parent chromosomes.

➤ Two crossover points are randomly generated to select a segment in one parent. Each string is then divided into three parts, the left, the middle and the right.

➤ Copy the left side and right side of parent 1 to form the left side and right side of child 1. According to the order of operations in parent 2, the operator constructs the middle segment of child 1 with operations of parent 2, whose IDs are the same as operations of the middle segment in parent 1.

➤ The role of these parents will then be exchanged to generate offspring child 2.

➤ Re-assign machines and tools to the operations in the middle segment to legalize the offspring chromosomes according to factory id.

The crossover example is shown in Figure 3, where the selected points are x=2, y=5.

Parent 1		4	1	6	2	5	3
	003	m-06	m-06	m-07	m-08	m-08	m-07
		t-09	t-10	t-10	t-12	t-14	t-13
Patent 2		2	6	1	3	4	5
	005	m-16	m-15	m-17	m-17	m-16	m-18
		t-22	t-22	t-20	t-23	t-21	t-21
Child1 (after exchange)		4	2	6	1	5	3
	003	m-06	m-16	m-15	m-17	m-08	m-07
		t-09	t-22	t-22	t-20	t-14	t-13
Child2 (after exchange)		2	1	6	3	4	5
	005	m-16	m-06	m-07	m-07	m-16	m-18
		t-22	t-10	t-10	t-13	t-21	t-21
Child1 (after legalization)		4	2	6	1	5	3
	003	m-06	m-08	m-08	m-07	m-08	m-07
		t-09	t-11	t-12	t-10	t-14	t-13
Child2 (after legalization)		2	1	6	3	4	5
	005	m-16	m-16	m-16	m-17	m-16	m-18
		t-22	t-23	t-21	t-22	t-21	t-21

Figure 3. Crossover example

3.2.2 Mutation

In the proposed GA, mutation happens to the chromosomes twice. The procedure of the mutation operation is described as follows:

(1) Randomly select a factory ID from the factory ID list.

(2) If the selected factory is the same as the original one in the chromosome, repeat step (1) until a different factory ID is chosen (mutation1).

(3) In order to legalize the chromosome, machines and tools will be re-assigned to all the operations according to the new factory-id.

(4) Randomly choose several pairs of genes excluding the first gene, and exchange their positions (mutation2).

Example of mutation 1 and mutation 2 operations are shown in Figure 4:

Chromosome (before mutation 1)	**003**	1	2	6	4	5	3
		m-06	m-06	m-07	m-08	m-08	m-07
		t-09	t-10	t-10	t-12	t-14	t-13
Chromosome (after mutation1)	**005**	1	2	6	4	5	3
		m-06	m-06	m-07	m-08	m-08	m-07
		t-09	t-10	t-10	t-12	t-14	t-13
Chromosome (after legalized)	**005**	1	2	6	4	5	3
		m-16	m-16	m-18	m-18	m-17	m-18
		t-25	t-22	t-23	t-22	t-24	t-21
Chromosome (before mutation 2)	005	**1**	2	6	**4**	5	3
		m-16	m-16	m-18	**m-18**	m-17	m-18
		t-25	t-22	t-23	**t-22**	t-24	t-21
Chromosome (after mutation 2)	005	**4**	2	6	**1**	5	3
		m-18	m-16	m-18	**m-16**	m-17	m-18
		t-22	t-22	t-23	**t-25**	t-24	t-21

Figure 4. Mutation example

3.3 Chromosome evaluation

When all the individuals in the population have been determined to be feasible, they can be evaluated based on the objective functions. In this research, minimum processing time, minimum production cost and minimum combination of weighted time and cost will be employed to calculate the fitness of each process plan. The calculation of production cost (PC) is the same as Zhang (1999), which will not be described in detail in this paper.

3.3.1 Minimizing processing time

Processing time (PT) generally comprises machining time, machine-change time, tool-change time and set-up change time. Here, four time indices are used to evaluate a process plan: machine change time index (MCTI), tool change time index (TCTI), set-up change time index (SCTI) and machining time index (MTI). Different from the first three time indices (referred to Zhang (1999)), machining time index (MTI) is assumed to reflect the importance of machining time in the overall processing time. In this research, it is assumed that the machining time of a unit volume is treated as a fixed amount for a certain type of operation no matter what kinds of machine and tool are used. The MTI for each type of operation is a user-defined parameter and can be specified by a data file input by the user. Thus, given the removed volume for manufacturing the feature, this feature's machining time will be the result of MTI multiplied by the removed volume.

3.3.2 Combination of weighted time and cost

In practice, it is desirable that a job can be manufactured and finished in minimum processing time with a minimum production cost. Although current approaches for process planning can

seldom realize such a target, it is possible to obtain a reasonably good combination of both. In this work, a combination of weighted processing time and production cost (*WTC*) is calculated based on the following equation:

$$WTC = W_t \times PT \times \beta + W_c \times PC \tag{1}$$

Where W_t refers to the weightage of processing time; W_c the weightage of production cost; and β is the transfer coefficient from a unit of time to cost and it can be estimated according to the real process planning problems.

4. ILLUSTRATIVE CASE STUDY

In this case study, production cost, the most commonly used criterion, is selected as the optimization objective. A prismatic part shown in Figure 5 is used to evaluate the capability of the proposed GA. The part was used by Chi-Cheng and Rajit (1996) to test a rule-based approach to minimize set-up time. The precedence relationships of the part are shown in Table 1. The details are reported in (Chi-Cheng and Rajit, 1996). It is assumed that each feature can be finished in one operation, i.e.

F1 (-y), F2 (-y), F3 (-z), F4 (+x, -x), F5 (-x), F6 (+x, -x), F7 (-x)

It is also assumed that there are three factories, 001, 002 and 003, possessing different machines and tools in the distributed manufacturing system.

F1 Machining feature to create F1'	F1' Rib	F2 Slot
F3 Machining feature to create F3'	F3' Boss	F4 Chamfer
F5 Blind hole	F6 Chamfer	F7 Through hole
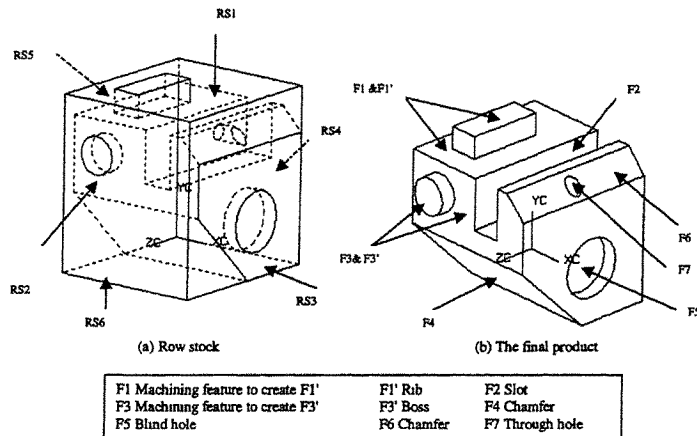
Figure 5. A prismatic part used in (Chi-Cheng and Rajit, 1996)

Table 1. The precedence relations for the case study (Chi-Cheng and Rajit, 1996)

Operation ID	Predecessor	Successor
Op1	op3	op2 op4
Op2	op1	op4 op6 op7
Op3	--	op1 op4
Op4	op1 op2 op3 op5 op6	--
Op5	--	op4
Op6	op2	op4 op7
Op7	op2 op6	--

Table 2 gives the cost indices of machine, tool and set-up changes (MCCI, TCCI and SCCI), in each factory and all the available manufacturing resources in the three factories, and values in parentheses are the cost indices of the machines and tools. In order to make the case study easily understandable, the cost indices in factory 001 are purposely set higher than the others. As shown in Table 3, columns 2 and 3 contain machine and tool candidates for every operation, in which machines (tools) from different factories are separated by columns, for example: "M-02 M-03 | M-05 M-06 | M-08 M-10" are different machines from factory "001", "002" and "003" respectively. The final process plan is shown in Table 4 with its total production cost and number of machine, tool and set-up changes, in which the precedence relationships shown in Table 1 are maintained. From Table 3, it can be seen that most of the operations can be accomplished on M-02 or M-03, however, the cost indices of M-03 is a little lower, so it is chosen instead of M-02.

Table 2. Manufacturing resources in the three factories

Factory-id	MCCI	TCCI	SCCI	Available machines	Available tools
001	**70**	**25**	**30**	M-01 (20), M-02(26),M-03(15)	T-01(3), T-02(2), T-03(4), T-04(6), T-05(5)
002	90	40	60	M-04(40), M-05(30), M-06(35), M-07(40)	T-06(15), T-07(10), T-08(12), T-09(9), T-10(13)
003	80	30	40	M-08(25), M-09(30), M-10(15)	T-11(4), T-12(3), T-13(5), T-14(7), T-15(6)

Table 3. Available resources for the operations

Operation-id	Machine candidates			Tool candidates		
Op1	M-02 M-03	M-04 M-05	M-08 M-10	T-03 T-05	T-07 T-09	T-12
Op2	M-02 M-03	M-04 M-05	M-08 M-10	T-03 T-04	T-07 T-09 T-10	T-12 T-13
Op3	M-02 M-03	M-04 M-05	M-08 M-10	T-03	T-07 T-09	T-12 T-13
Op4	M-02 M-03	M-04 M-05	M-08 M-10	T-01 T-03	T-07 T-10	T-12 T-15
Op5	M-01 M-02 M-03	M-04 M-07	M-08 M-09 M-10	T-02	T-06 T-08	T-11 T-14
Op6	M-02 M-03	M-05 M-06	M-08 M-10	T-01 T-03	T-09 T-10	T-12 T-13
Op7	M-01 M-02 M-03	M-04 M-07	M-08 M-09 M-10	T-02	T-06 T-08	T-11 T-14

Table 4. The process plan-1 against criterion-1 (minimum production cost)

Factory-id: 001			
Operation-id	Machine-id	Tool-id	TAD-id
Op3	M-03	T-03	-z
Op1	M-03	T-03	-y
Op2	M-03	T-03	-y
Op6	M-03	T-03	-x
Op7	M-03	T-02	-x
Op5	M-03	T-02	-x
Op4	M-03	T-01	-x
Number of machine change: 0	Number of tool change: 2		Number of set up change: 2
The total production cost is 238			

Since production cost is chosen as the optimization criterion, the distributed CAPP system chooses factory "001" with the lowest cost index for this job instead of factories "002" and "003". This finding indicates that the proposed GA can choose the most suitable factory for generating an optimal or near-optimal process plan with the lowest cost.

5 CONCLUSIONS

A modified GA for a distributed CAPP system has been developed based on geographically dispersed resources such as machines and tools in several factories. An optimal or near-optimal process plan for distributed manufacturing can be produced based on several optimization criteria such as minimum production cost, minimum production time or a combination of both based on pre-defined weightages. The final process plan would allow the most suitable factory or factories to be selected for a part. The findings of this paper would further enhance the functioning of distributed manufacturing systems.

REFERENCES:

Bhashara, R.S.V., Shunmugam, M.S. and Narendran, T.T., (1999), "Operation Sequencing in CAPP Using Genetic Algorithms", *International Journal of Production Research.*, Vol 37(5), pp.1063-1074.

Chi-Cheng, P.C. and Rajit, G., (1996), "Feature-based Approach for Set-up Minimization of Process Design from Product Design", *Computer-Aided Design*, Vol28(5), pp.321-332

Zhang, F., (1997),"Genetic Algorithm in Computer-Aided Process Planning", *M.Eng. Thesis*, National University of Singapore.

Technology Management

Quality function deployment – how it can be extended to incorporate green engineering objectives

H K WONG and J JUNIPER
School of International Business, University of South Australia, Australia

SYNOPSIS

Quality Function Deployment (QFD) is a well-established methodology that incorporates customer requirements and technology attributes into new product development (NPD) and continuous improvement activities. This paper aims to develop a new QFD for the achievement of green engineering objectives. QFD systems that are integrated with green engineering objectives may provide a robust and efficient methodology to resolve particular problems in waste minimisation, design-for-recyclability, the achievement of greater energy efficiencies, more optimal utilisation of renewable and non-renewable resources, and enhanced pollution control.

1 INTRODUCTION

This paper aims to develop a new Quality Function Deployment (QFD) for the achievement of green engineering objectives[1]. Through an industry placement the research will be based on a case study involving both the development and implementation of such a system. It is intended that risk management-related aspects of QFD system implementation will be accommodated through the use of some forms of multiple-attribute decision modelling.

QFD is a widely known technique for setting priorities and targeting activities that relates to new product development and strategic product and process planning The development of competitive green products and processes requires the reduction of waste and toxic materials, on one hand, and the streamlining of manufacturing on the other, thus saving money in the long term: this is core of the green engineering concept.
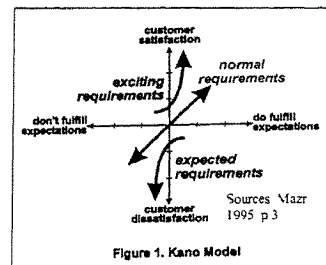
Figure 1. Kano Model

---

[1] Green Engineering objective can be defined as a framework embodying environmentally conscious attitudes, values and principles in association with relevant science, technology and engineering practices, all directed towards improving local and global environmental quality GE thus encompasses all of the engineering disciplines and is consistent and compatible with sound engineering design principles

During the implementation stage the project will investigate opportunities for using fuzzy logic as a vehicle for translating ambiguous and overlapping linguistic categories into quantitative measures. It is envisaged that this technique will better support efforts to enhance customer satisfaction and to categorise customer needs, as in the Kano quality model (see Figure 1), and may also assist in the management of risk.

The widely used Kano model fosters product improvements through functional innovations that are seen to best answer to the customer's unvoiced needs. This is because the basic quality line approach is not viewed as doing enough to create fully satisfied customers. As such, the model should better enable firms to churn out more desirable new products with greater efficiency, in less time and with lower manufacturing costs.

This paper sets out a framework for work in progress. The authors are actively involved in implementation of the methodology within a local, export-oriented, wine packaging and bottling plant, but at this early stage we cannot use this research as the basis for a detailed case study. As such, implementation issues are discussed in quite general terms and the actual case study of a restaurant, to be discussed later on in the paper, is provided solely for illustrative purposes.

The following section defines QFD and the Kano Model. The next section addresses the question of why QFD implementations, on their own terms, need to address issues related to green engineering. Then we consider how to extend QFD so that it becomes Green QFD (GQFD). A related section articulates the linkage between customer and community requirements and technology attributes. An illustrative case study is examined in the subsequent section of the paper. In this section we also suggest approaches that can be adopted to assist in implementation of the approach set out within the paper. Conclusions then follow.

1.1 Defining quality function deployment (QFD) and the Kano Model

In this paper, I follow Mazr's lead by investigating whether the Kano quality model can be incorporated into the QFD structure to better satisfy customer requirements (Mazr, 1994). Kano's model can be dynamic through its focus on product attributes that are best able to excite customers. They can also have additional dimensions in relation to different customer segments within a target market, including those to be created in the future.

QFD is essentially a customer-driven planning process to guide the design, manufacturing and marketing of goods. Throughout actual applications of QFD, every design, manufacturing and control decision is made to meet the specific needs of customers. Nowadays, QFD has been successfully used leading by some manufacturers of clothing, motors, construction equipments, electronic equipments, appliances by Toyota, Kodak, IBM, Xerox, Motorola, Ford, General Motors, AT&T and Hewlett-Packard (Mazr, 1994).

QFD originated in Mitsubishi's Kobe shipyard sites in 1972 (Akao, 1995, Mazr, 1995 & Evans, 1999) and has been used for thirty years in Japan as a quality system. Evans (1999) mentions that Toyota began to use the QFD concept in 1977. They achieved very good results, including a 20 percent reduction in start-up cost associated with the launching their new van.

These cost reductions were on-going so that by 1982 start-up costs had fallen by roughly 38 percent compared with 1977. And 1984 had achieved a 61 percent overall cost reduction.

In order to satisfy customers, we must understand clearly what customer's are seeking. In the Kano quality model, three different types of customer requirements are considered, as shown in Figure 1.

- **Normal requirements** are what we get by just asking customers what they want (Mazr, 1994). For the requirements satisfaction is directly derived from what is visibly present in the product or service.

- **Expected requirements** are so basic that customer may fail to mention them until the business fails to meet them (Mazr, 1994). They are basic expectations without which the products or service may lose much of their value.

- **Exciting requirements** are harder to discover. They are always situated beyond the customer's expectations (Mazr, 1995). Their absence does detract from satisfaction; but beyond this, their presence excites most customers.

Kano's model can readily be discussed in QFD terms: First, Kano's model can be used to highlight what excites us today. This implies that ultimately, the features responsible for the firm's competitors will imitate excitation so that customer will eventually come to expect them from everybody. Second, as we have seen, Kano's model can incorporate additional dimensions in relation to the key customer segments within the target market including. Third, Kano's model can be used to identify the sources of excitation that are tied to adding value, but are hard to identify both for the customer and the provider.

2 QFD AND GREEN ENGINEERING

In the past decade, many companies worldwide have made considerable progress in reducing the environmental impact of their manufacturing operations. It has increasingly been recognised that all-engineering disciplines can contribute to the process of improving environmental performance (Mackenzie, 1997). The incorporation of green engineering concepts relating to design in manufacturing is now a major focus of interest as more consumers are incorporating environmental considerations into their purchasing decisions and examining how they can safely dispose of products when they are no longer of use (Wong, 2000). In a discussion of environmental issues, Mackenzie (1997) states that

> We can deal with the problems we face and begin to turn things around in
> the next 40-50 years if we begin now (p. 84).

There is a pressing need for a fundamental change in the decision making process for manufacturers. A new QFD incorporating green engineering activities can employ the standard conventions of a step-by-step approach moving from customer needs and expectations through the four planning phases, namely:

- Product planning;
- Product development;
- Process planning; and
- Production planning through to manufactured products and delivered services.

QFD can therefore be more effective to the extent that it incorporates green engineering objectives, because it adds sustainable environmental advantages onto the existing sustainable competitive advantages for the company, which can also act as sources of value, uniqueness, and barriers to imitation while affording the organisational resources to preserve these sources of competitive advantage.

2.1 Transforming QFD into GQFD by incorporating green engineering activities

QFD is a concurrent engineering and management-planning tool driven by customer and stakeholder expectations. Arguably, environmental attributes represent an increasingly essential component of stakeholder value (i.e. employees and the community). Broadly speaking, GQFD could well become one of the leading methodologies amongst major corporations, who may well attempt to impose these systems on their suppliers. Akao (1995) states that:

> For companies to attain customer satisfaction, it is important that all
> employees acquire customer focused thinking through the value chain
> created by the awareness that the next process is your customer. It is with
> QFD that companies will be able to accomplish this future challenge. QFD
> will serve as a tool for creating this alignment, where true partnership can
> sprout. (p. 5)

The house of quality framework of GQFD (see Figure 2) includes a *WHATs* row for customer attributes and environmental attributes, a *HOWs* column to reflect relevant technology attributes and relative competitiveness, and a *HOW MUCH* matrix to account for production costs, while the "roof" of the QFD house of quality captures any interactions between the so-called "voice of the customer" and technology attributes. The customer perception section, plays an important role as the focal point for market evaluation for the evaluation of the technical requirements of competitive products, in the development of targets and priorities, and lastly, in the selection of the technical requirements that are to be specifically deployed in the remainder of the manufacturing process.

In the case of GQFD, customer attributes will be broadened to incorporate other stakeholder values and then related to the respective technology attributes, which include the design of products and processes that reduce or eliminate waste, minimise energy consumption, achieve optimal (renewable and non-renewable) resource usage, and control the release of hazardous substances. Hopefully, one major outcome of such an environmentally sensitive design and production procedure will be increased profitability and competitive advantage, not merely achieved through lower waste disposal, treatment and environmental compliance costs, but also through higher levels of customer satisfaction, employee commitment, and reduced production costs.
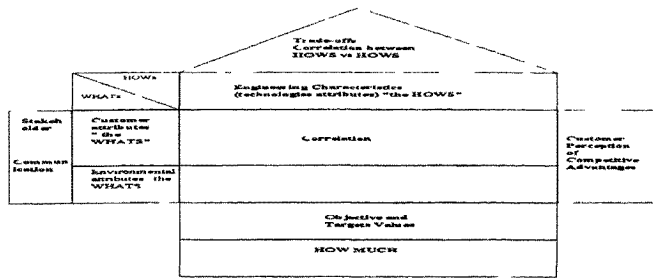
Figure 2. The framework of GQFD methodology

2.2 GQFD extended to encompass the four houses of quality

Green Quality Function Deployment (GQFD) not only starts with the analysis of customer requirements, but it also relate to environmental issues as well, by focussing on new product development and service delivery.

Customer requirements (the "voice of the customer") can be decomposed into a number of different categories; namely, quality, technologies, performance, functionality, reliability, cost, and process methods. Quality requirements (customer attributes) are based on the customer's wants and needs in which can be represented using fuzzy and qualitative forms of analysis. However, as we have argued above, environmental attributes (green issues) also have to be considered as a type of "customer need". Each of these requirements has to be related to technology attributes. Accordingly, managers and workers will have to gather together the requisite information so that customer requirements and technology attributes (engineering characteristics) can be discussed in relation to specific design targets.

In summary, the green quality function deployment matrix is used to articulate the relationships obtaining between customer attributes, environmental attributes, and technology attributes and on the basis of strategic importance, to identify priorities for the design of future improvements.

A comprehensive implementation of QFD actually involves a sequence of four HOQ to carry the voice of the customer (VOC) over into later stages of the overall production process. Thus, three additional HOQ relate the VOC to the characteristics of component production, process planning and production planning (see Figure 3). As indicated above, these matrices must be used for more than the mere identification of customer attributes, technology attributes, and environmental attributes. They must also be used to relate customer and competitive rivalry attributes to the relevant technology attributes, thereby supporting an evaluation of competing products, the development of targets and the determination of which technology attributes to deploy in the remainder of the production process.
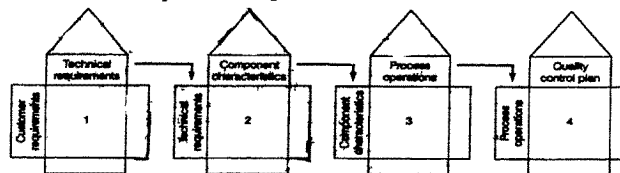
Figure 3. Four House of Quality (Adapted from Evans 1999, p. 414)

Along these interlinked matrices, the customer and environmental attribute priorities will be matched to those relating to technology attributes (engineering characteristics). The technology attributes will be benchmarked against competitors to identify design targets for the most critical technology attributes i.e. those, which can be set to achieve more highly valued benchmarks. In this way, the business can hopefully achieve more sustainable competitive advantage within the global market.

In summary, the GQFD system drives product development from the planning through to the production stage (including end-of-product-life), based on what matters most to the customer, identifying where there is an opportunity for competitive positioning with the strategic objectives and general vision of the organisation.

3. IMPLEMENTATION ISSUES

Implementation of this methodology within an actual organization requires sensitivity on the part of the practitioners and the full support of senior management. Information must be gathered from all the key functional areas of the firm including production, R&D, marketing, finance, human resources, distribution and logistics. Tangible concerns must be addressed alongside those that are fairly intangible and subjective. Practitioners will inevitably be involved in some form of iterative, action-learning process. For this reason, we favour the application of a soft-systems methodology (SSM) such as that proposed by Checkland (2000) and Churchman (1971). Common to all these strands of management thinking is an emphasis on the ill-defined nature of most human, problem-orientated situations. Objectives are often ambiguous, and susceptible to a multiplicity of interpretations. This is because orientations towards a given problem situation were typically characterized by a marked divergence in actor interests, values, beliefs, priorities, and goals. A series of relevant *human activity systems models* should be chosen, reflecting the range of world views or *Weltanschauung* upon which they are based (Checkland, 2000). Each world-view is seen to determine both relevance of the model and its content. Moreover, rather than working with a clearly defined "problem" emphasis is placed on problematical *situations*. Therefore, any approach to the resolution of these problem situations is itself regarded as an emerging, *organized learning system* (p. 15). As such, the models are not so much representations of a certain set of external relationships in the world outside the researcher but rather, accounts of concepts about purposeful activity, based on declared world-views, which can be used to stimulate debate around what is perceived to be problematic and what can be done to change things (p. 26).

3.1 An Illustrative Case-Study
The following case study draws on one initially outlined in Mazr (1994). This paper described activities within a restaurant showing how the owners had to improve the overall process in a useful way that went beyond the mere preparation of food for the customer, to consider how the customer was treated after he or she had finished dining. In this study, Mazr (1994) adopted QFD techniques to examine customer needs and service activities, but without including any environmental attributes. Thus, for purely illustrative purposes, we have incorporated certain environmental attributes such as waste treatment and energy-related costs

of food preservation and preparation. Our extended GQFD version of Mazr's matrix is shown in Figure 4 below:

		Food preparation	Prepare	Cook	Convey	Serve	Clean	Wash	Customer Perception of Competitive Advantages							
									Importance	US	Hotel S	Plan	Improvement Ratio	Sales Point	Raw Score	Demand d quality weight
Customer attributes	Food in fresh		3	3	9				4	3	4	4	1 3	1 2	6 2	23 9
	Local delicacies		3	9		1			1	3	4	4	1 3	1 2	1 6	6 3
	Elaborate presentation		3	3	9	1			2	3	3	4	1 3	1 0	2 6	10 0
	Task seasoning		9	9					2	3	3	3	1 0	1 0	2 0	7 8
	Proper temperature		3	9	9				4	3	4	5	1 7	1 0	6 8	26 9
	Polite service					9			3	4	4	4	1 0	1 0	3 0	11 3
	Clean	3					1	9	3	3	3	4	1 3	1 0	3 9	16 6
Environmental Attributes	Packaging waste	5	3	3	9	9	9		3	3	3	4	1 3	1 2	4 8	24 5
	Difficult to compost	3	3					3	3	3	3	3	1 0	3 0	15 3	
	Access to cleaning	9		5		5	5		2	4	3	4	1 0	1 0	2 0	10 2
	Chemical waste	3	3				5		2	4	4	3	0 8	1 2	1 8	9 2
	Low cost in food preservation	9		3			3		4	3	3	4	1 3	1 5	8 0	40 8
	Absolute Weight	976 70	617 97	793 66	338 58	410 47	512 53									
	Process Priority	26 76	16 93	21 74	9 28	11 25	14 04									

Figure 4. Prioritisation Matrix for restaurant (modified original from Mazr, 1994 & 1997)

The best way to fully explain the GQFD matrix is to follow a couple of threads. The most important customer attribute is "proper temperature": the fifth row on the left side of the matrix. Amongst the environmental attributes the most important is "low (i.e. energy-related) cost in food preservation", which appears in the last row of the environmental attribute section. The main technology attributes relating to these "customer requirements" are "prepare", "cook", "convey" and "clean", which identify the relevant design characteristics that must be considered in relation to both the given environmental cost factor and the customers' concern with the overall cost of food preparation.

Figure 4 shows that the importance rating for the environmental attribute, "low (energy-related) cost in food preservation," is 4, which means that the restaurant has to pay more attention to this specific factor. Now, the vision of the company is reflected in the sales point score. Sales points relating to each of the attributes are given values of 1.0 (weak), 1.2 (moderate), and 1.5 (strong). The improvement ratio gives a subjective evaluation of how successful a design change is likely to be in achieving improvements in the given attribute. The raw score for a given attribute is then calculated as the importance rating × improvement ratio × sales point (4 × 1.3 × 1.5 = 8.0). The demand quality weight appearing in the final column is a relative ranking determined by converting each of the raw scores into a percentage within each of the relevant attribute segments—customer and environmental—of the HOQ matrix. For example, the demand quality weight of "low (energy-related) cost in food preservation" is 40.8% [(8/(4.8 + 3 + 2 + 1.8 + 8) × 100% = 40.8%]. This weight identifies the criticality of "low cost in food preservation" based on its combined importance in terms of the customer, competitive position and company vision.

Each of the cells appearing below a given technology attribute has assigned to it a score representing the level of interaction between the technology attribute and the customer attribute (i.e. it indicates the contribution that the specific technology attribute makes to each of the customer attribute). The absolute weights for each technology attribute are the calculated by multiplying each of these cells by its respective demand quality weights appearing at the end of the row. In the process priority row these absolute weights are then converted into relative percentages. As with the demand quality weights, these percentage scores can now be used to assign priorities to the various product improvement tasks. In

reference to Figure 4, it can be seen that the "prepare" column is ranked as having the most importance in term of technology attributes, as given by the absolute weight (966.5) and resulting process priority (26.85%).

4 CONCLUSION

GQFD differs from other design approaches in that it focuses not merely on inspecting problems "out of the product", but also on understanding customer requirements and valuing the contributing that green engineering activities make in meeting our needs. Those potential features and functions that are critical to customer satisfaction can be "designed in", while potential failures can be anticipated and "designed out", so that savings in both time and money can be realised by minimising last minute design changes, reducing product introduction delays, and preventing market failures. Thus, GQFD ensures that customer satisfaction can be achieved in the quickest, least costly, and most efficient way. It is the great strength of QFD that it focuses on sustainable competitive advantage. GQFD includes both competitive and environmental sustainability. As such, GQFD is a valuable tool that can be used to complement other forms of environmental management, including as those based on systems of quality assurance such as Life Cycle Assessment.

REFERENCES

1. **Akao, Y.** (1995) "QFD toward development management", Proceedings Of The International Symposium On Quality Function Deployment' 95 Tokyo. pp. 1-8.
2. **Checkland, P.** (2000) "Soft systems methodology: A thirty Year Retrospective," *Systems Research and Behavioural Science Systems Research*, Vol. 17, pp. S11-S58: reproduced from *Soft Systems Methodology in Action*, John Wiley, Chichester, 1999.
3. **Churchman, C. W.** (1971) *The Design of Inquiring Systems, Basic concepts of Systems and Organisation*, Basic Books, New York.
4. **Dale, B. G.** (1999), *Managing quality*, Blackwell Publishers, Malden.
5. **Evans, J. R. & Lindsay, W. M.** (1999) 4th Edition, *The Management & Control Of Quality*, South-Western College Publishing, Cincinnati, Ohio.
6. **Mackenzie, D.** (1997) *Green Design: Design for The Environment*, Laurence king, London.
7. **Mazr, G.** (1994) *Initiating Service Quality Function Deployment*. Ann Arbor, MI: Japan Business Consultants LTD.
8. **Mazr, G.** (1995) "Elicit service customer needs using software engineering tools." *Transactions from the seventh symposium on quality function deployment*, Ann Arbor, MI: QFD Institute.
9. **Mazr, G.** (1997) "Task Deployment: The human side of QFD", *Transaction from the Ninth symposium on quality function deployment/ international symposium on QFD 97*, ANN Arbor, MI: QFD Institute.
10. **Monden, Y. & Sakurai M.** (1997) *Japanese management accounting: a world-class approach to profit management*, Productivity press, USA.
11. **Wong, H. K.** (2000) Green design and green manufacture techniques: a state of the art review, M Eng thesis, University Of South Australia.

Flexibility management and measurement of flexibility in Australian manufacturing industry

B KAYIS and K SKUTALAKUL
School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, Australia

ABSTRACT

Flexibility is another competitive weapon that has proved its significance in this evolving manufacturing world of coping up with the ever changing and increasing customer requirements. Numerous industries of different nature have been aware of adopting principles of flexibility management to stay competitive. This study established a conceptual framework on the interrelationships of flexibility indicators as well as evaluating the flexibility levels currently being practiced by Australian industries. The outline of the mechanism of how the customer-supplier relationships have an impact on total chain of manufacturing would enable us to understand how flexibility could be built. As a result of the overall flexibility assessment in response to the current level of customer-supplier relationship, the current manufacturing flexibility of Australian industries according to the survey is approximated to be at medium level.

1 INTRODUCTION

The current thinking amongst those driving manufacturing is that industry is moving away from mass production towards mass customisation (O'Kane, 2000). The response to the market tends to be dramatic to cope with the ever-changing demand of customers in order to be or remain competitive. Nowadays, all the efforts of restructuring, reengineering, and reinventing are simply to make a company flexible enough to respond to the fast-changing customer needs, as flexibility, another competitive weapon used in today's competitive markets, is defined as 'the ability to respond effectively to the ever-changing and increasing needs of the customer' (Mandelbaum, 1978; Sethi & Sethi, 1990).

The traditional thought was that there were so many conflicts in the multiple demands on the operations function that trade-offs were made in achieving excellence in one or more dimensions. From a strategic perspective, the dimensions of cost, quality, time and flexibility are not to be traded off against one another but need to be simultaneously prioritised (Erenguc et al., 1999; Kayis et al., 2002).

At acquisition stage, the relationship with suppliers is critical in relation to supplier flexibility. Since a supplier is conceived primarily as another set of people, the improvement of suppliers for flexibility is seen as the improvement of people of supplier, either on their own and particularly with the assistance and cooperation from manufacturing industries. Distance in

transit times from suppliers has much less to do with the flexibility and lead times than the relationships with them (Association of Manufacturing Excellence, 1990). The participation could be incorporated in terms of different aspects within the industries, such as design, equipment, maintenance, policy planning, etc. This relationship is the basis for any member of the supplier family working out of a jam whenever it is in one. To achieve flexibility at the supplier interface, firms should look at some of their major suppliers as partners, and where possible encourage them to take part in strategic activities such as product and process design and development. In many cases, suppliers should have complete responsibility of component testing and quality control (Kamath and Liker, 1994).

The capability of a manufacturer to offer a rich variety of products is dependent on the supplier's capability to produce a variety of component parts, i.e. the supplier's flexibility in several dimensions: delivery time, mix, volume, new products, etc. The development of a responsive, integrated supply base leads to changes in production, engineering and marketing strategies. Manufacturers can create and support competitive order-winning competencies by targeting specific manufacturing flexibilities with suitable supply-base strategies (Narasimhan and Das, 1999).

At distribution stage, widespread distribution is the ability to effectively provide widespread distribution coverage, intensive distribution coverage, or both (Vickery & Calantone, 1999). Despite the distribution aspects, customer's participation also plays a critical role in distribution flexibility since customers determine the mechanisms of what and where the products are needed. Customers also could participate in different manufacturing aspects like the suppliers do, namely design, workforce, equipment selection, policy planning, etc. In practical terms, the network could be a set of customers, a set of distributors or a set of product dealers.

Flexibility is important at every stage of the total chain for realising vast improvements in the overall flexibility of organisations. Management attention towards a number of plant and machine based flexibility strategies is important for developing the flexibility of the processing stage, and hence improving the total flexibility of the chain. At the processing stage of the supply chain, the direct inputs from the suppliers are transformed into the final product (Erenguc et al., 1999). Being fast enough to be the first in the market with the right product, is worth more to the prosperity of most businesses than any other single manufacturing function.

After all, to achieve flexibility, companies need to make each stage of the acquisition-processing-distribution chain as flexible as possible. This can be accomplished by adding flexibility to each resource without spending too much money, while at the same time, attaining sufficient capability to satisfy all customer needs.

Each organisation has a structure in which it operates. With varying requirements to respond to the market, the way in which the organisation operates to respond to its market will depend on the degree of flexibility that is required (Webster, 2001). Bearing in mind of the various possible approaches to respond to the market, a question still lies on 'How to do it without excessive costs, time, organisational disruption, or loss of performance?', which conforms to the description of flexibility as stated by Aggarwal (1997), Knudsen (1996), and Upton (1994). Therefore, the study attempts to measure, analyse, investigate, and test the survey

data as well as the interactions of different parameters collected to suggest a broad idea of the best path to flexibility improvement.

2 METHODOLOGY

In accordance to different classifications of organizations in Australia, the sampling population is comprised of manufacturing companies with different Standard Industry Codes (SIC), such as food and kindred products, industrial and commercial machinery, fabricated metal products, chemicals and allied products, etc. A sample of 249 companies was randomly selected for the mail survey, and 71 companies have completed the survey, which were accounted for 28% of the total survey sample. To analyze the flexibility management of the organization, the survey questions were categorized into four dimensions, namely the organizational structure, manufacturing practices, supplier's contribution, and customer's contribution. The first two dimensions could be considered as 'internal' aspects of the organization, where the other two dimensions could be called 'external' aspects, respectively.

2.1 Organisational Structure

This section of the survey deals with the details of organisation regarding their structure and background. The outline of information gained from the questions is detailed below.

- Number of employees (organisational size)
- Ownership distribution (percentage of Australian and foreign)
- Annual gross sale in Australian dollars
- Number of plants in Australia
- Process choice
- Number of products produced

Most of the data under this category is already quantitative in different units. However, the process choice could be considered either as nominal or ordinal type of data, where each choice would be assigned a number, preferably 1 to 4.

2.2 Manufacturing Practices

This section of the survey deals with the existing manufacturing practices, objectives, and problems. The outline of information gained from the questions is detailed below: -

- Production volume
- Bottleneck identification
- Extra capacity of the facilities
- Manufacturing objectives
- Scale of context and depth of manufacturing jobs
- Manufacturing problems (the degree of seriousness)

The data obtained under this category is both qualitative and quantitative. The manufacturing problems, for instance, could be considered as ordinal data type, where the degree of seriousness is limiting from 1 to 3 (Not serious to very serious). Other data could be organised similarly through ordinal scale, for the ease of further quantitative formulation.

2.3 Supplier's Contribution

This section concentrates mainly on the acquisition stage of the supplier chain, emphasising how supplier's role is reflected to other stages of the chain. The outline of information gained from the questions is detailed below.

- Contact regarding design/ product availabilities
- Supplier's participation under different categories
- Number of suppliers over past 5 years (increase, decrease, or stay the same)

The possible answers to the supplier's participation level were defined on a three-point rating scale. The higher score was attributed to definite participation in different aspects that could benefit manufacturing organisation. Generally, the data type under this section is based on the ordinal scale, where the ratings of scale are limited, such as participation level of 1 to 3 (No participation to definite participation).

2.4 Customer's Contribution

Similarly, this section contains information with the same format as the supplier's contribution, making the sensitivity of effect with one another become more obvious. The questions further took a brief look if customers are demanding flexibility and the response of the companies towards the demand. The questions also address the current awareness of the company to flexibility, which is reflected in terms of change in product design and market demand. The outline of information gained from the questions is detailed below.

- Contact regarding design/ product availabilities
- Customer's participation under different categories
- Number of customers over past 5 years (increase, decrease, or stay the same)
- Product market spread (Domestic, Industry, Military, or Export)
- Change in market demand over the past five years
- Change in product design over the past five years
- Demand for product customisation/variation
- Ability of the company to cope with such customisation/variation

Same rating scale was used for the customer's participation level as those of supplier's. However, the change in market demand and product design was based on a five-point rating scale, representing no change at all (1) to completely different status (5). Under the demand for customization/ variation, the question addresses how often the customers have asked for product change in nominal scale, namely 'never', 'occasionally', 'regularly', and 'always'. Such demand could be compared with the company's ability to respond using 'current system', 'slightly modified system', 'highly modified system', or 'new system'. Asking the organizations if they are able to customize or add variation to the product helps determine the current level of flexibility practices and possibly how much would it take to make operations flexible enough to cope with demand.

3 RESULTS

The customer-supplier relationship has been a major challenge for manufacturers in terms of their ability to think strategically. The relationship over design, assembly, and marketing of finished products with both customers and suppliers could possibly become a competitive advantage. The approach of customer-supplier relationship is to create value-added partnerships in a particular industry, with great amount of issues regarding vertical integration. The stronger customer-supplier relationships has been increasingly importance due to the concentration of the firm on what it does best, rather than being responsible for all activities in the supply chain. This then means that greater dependence upon suppliers and greater corresponding response to customers are the requirements.

The study further looks into the individual components of customer-supplier relationship, which are supplier flexibility and customer flexibility. The evaluation is nevertheless done in a compromising manner where some subjective consideration would be drawn to reflect the overall manufacturing situation of Australian industries. The impact of customer-supplier flexibility on the total chain of manufacturing could also be introduced regarding the inner core of manufacturing flexibility that constitutes of its process, design, human resources, and policies. The performance assessment framework for this survey is also drawn to connect the interlinking factors and contribution regarding customer, supplier, and manufacturing flexibility of Australian industries. With the framework, we would be able to see the performance level of those industries regarding the topics of suppliers, customers, and manufacturing side. Thus, the subjective ratings are assigned to each flexibility factor to give a clearer quantitative idea of the flexibility performance.

As the different elements under manufacturing flexibility have suggested, the manufacturing flexibility of Australian industries as affected by customer-supplier participation is 'MEDIUM'. The overall assessment in response to the flexibility assessment framework could be seen in Table 1.

As a result, the manufacturing flexibility and the elements under them (design, process, policy, HR) are considered to be at medium level of flexibility corresponding to the estimated customer-supplier flexibility

Even though the flexibility level under different criteria is evaluated to be at medium level according to the survey data, it does not mean that the performance of Australian industries in such area is good and need no more improvement. As flexibility is defined as the ability to cope up with the ever-changing and increasing demand, the industries need to keep in mind of the continuous improvement methodology including the aspect of flexibility.

Table 1 Results of manufacturing flexibility assessment

Flexibility elements	Flexibility level
SUPPLIER FLEXIBILITY	Medium
CUSTOMER FLEXIBILITY	Medium
MANUFACTURING FLEXIBILITY • Process flexibility • Design flexibility • HR flexibility • Policy flexibility	Medium Medium Medium Medium Medium

REFERENCES

Aggarwal S. (1997) Flexibility Management: The Ultimate Strategy, *Industrial Management*, Vol. 39, No. 1 Jan-Feb, 5pp.

AME (1990) Association of Manufacturing Excellence, Flexibility: Manufacturing Battlefield of the 90s, Sponsored by the Association of Manufacturing Excellence and Systems Science Institute, Waseda University.

Erenguc S.S., Simpson N.C., Vakharia A.J. (1999) Integrated Production/Distribution Planning in Supply Chains: An Invited Review, *European Journal of Operations Research*, No. 115, pp. 219–236.

Kamath R.R., Liker J.K., A Second Look at Japanese Product Development, *Harvard Business Review*, pp. 154–170.

Knudsen D.C. (1996) The Transition to Flexibility, Kluwer Academic Publishers, Massachusetts, USA.

Mandelbaum, M. (1978) Flexibility in Decision Making: An Exploration and Unification, *Ph.D. Thesis, Department of Industrial Engineering*, University of Toronto, Toronto, Canada.

Narasimhan R., Das A. (1999), Manufacturing Agility and Supply Chain Management Practices, *Production and Inventory Management Journal*, Falls Church, Vol. 40, No. 1, First Quarter, pp. 4-10

O'Kane S. (2000) Flexibility Management: Analysis of the Impact of Flexibility Management in the total chain of Acquisition, Processing, Distribution Function of Australian

Manufacturing Companies, BE Thesis, The University of New South Wales, School of Mechanical and Manufacturing Engineering, Sydney, Australia.

Sethi, A.K. and Sethi, S.P. (1990) Flexibility in Manufacturing: A Survey, International Journal of Flexible Manufacturing Systems, Vol. 2, pp. 289–328.

Upton, D.M. (1994) The Management of Manufacturing Flexibility, *California Management Review*, Vol. 36, No. 2, Winter, pp. 72–89.

Vickery S., Calantone R., Droge C. (1999) Supply Chain Flexibility: An Empirical Study, *The Journal of Supply Chain Management*, National Association of Purchasing Management, August/Summer.

Webster O. (2001) Value Stream Management, BE Thesis, The University of New South Wales, School of Mechanical and Manufacturing Engineering, Sydney, Australia.

Dynamic management of assembly constraints for virtual disassembly

P CAO, J LIU, and **Y ZHONG**
CAD Center, Huazhong University of Science and Technology, Wuhan, People's Republic of China

ABSTRACT

This paper presents a dynamic management mechanism of assembly constraints for virtual disassembly. The mechanism comprises acquisition, representation and maintenance of assembly constraints and management of tool positioning constraints as well. Assembly constraints are acquired by constraints reconstruction combined with constraints transfer and are represented using an assembly constraint graph. An indirect deletion method of assembly constraints is presented to maintain assembly constraints during disassembly. Management of tool positioning constraints is investigated so as to enable disassembly with tools.

1 INTRODUCTION

Virtual reality (VR) technology provides natural and direct manner for interaction between designers and the virtual environment. Virtual disassembly, which provides a powerful means for design for disassembly, is one of the most important engineering applications of VR. To implement virtual disassembly, the assembly model of the product designed in CAD system must be transferred into virtual environment through some special interfaces. Nevertheless, the display model of objects of commercial VR software is generally polygonal model, which results in the loss of data like assembly constraints that is necessary for the analysis of the disassemblability of products in virtual environment. In addition, fasteners usually exist in an assembly and disassembly them with tools makes virtual disassembly more authentic and reliable. This paper presents a dynamic management mechanism of assembly constraints that can support disassembly either with or without tools. The mechanism is addressed and its implementation approach is discussed.

2 RELATED WORK

The research about disassembly has focused on disassembly sequencing, evaluation tool development and disassembly method investigation (1-6). With the flourish of VR technology, academia has paid much attention to virtual assembly·disassembly. de Sa' investigated the steps needed to apply VR for virtual prototyping to verify assembly and maintenance processes (7). Fernando indicated that a common weakness of the existing environments is the lack of efficient geometric constraint management facilities (8), and conducted such work as *allowable motion* and *automatic constraint recognition* to support constraint management within virtual environment. He also developed a software framework to support constrained-based assembly and maintenance operations (9). Jayaram et al. developed the virtual assembly design environment (VADE) that allows engineers to plan, evaluate and verify assembly of mechanical products (10-12). The tool-part-human interactions are modelled in VADE to enable the use of tools for disassembly of fasteners in the assembly.

3 DYNAMIC MANAGEMENT MECHANISM OF ASSEMBLY CONSTRAINTS

From the point of view of constraints space, disassembly can be regarded as the process in which components move from a constraints space to a constraints-free space gradually, with constraints becoming invalid and being deleted. It is likely that some assembly constraints will generate during disassembly. However, these constraints are temporary constraints because they will become invalid and be deleted later when disassembly proceeds. Tool positioning, tool use and tool removal are the three steps of disassembly with tools. In tool positioning, the tool is taken as a component and is assembled onto the fastener to be disassembled. Tool use is the process in which the tool is operated to disassemble the fastener. In the final step of tool removal, the tool is disassembled from the fastener. Practically, a set of assembly constraints is satisfied after tool positioning and they keep no change in tool use but are deleted in tool removal. In this paper, the assembly constraints between the tool and the fastener that are satisfied in tool positioning are defined as **Tool Positioning Constraints (TPCs)**. Thus the three steps of disassembly with tools correspond to three processes, that are satisfaction of TPCs, preservation of TPCs and deletion of TPCs respectively.

A dynamic management mechanism of assembly constraints is presented to support virtual disassembly. In this paper, assembly constraints are abbreviated to ACs and dynamic management of assembly constraints to DMAC. As shown in Fig.1, DMAC is composed of four modules. They are *acquisition of ACs, representation of ACs, dynamic maintenance of ACs* and *management of TPCs. Dynamic maintenance of ACs* refers to tracking and updating ACs during disassembly and it includes *dynamic deletion of ACs* and *dynamic generation of ACs*. The *management of TPCs* module includes *satisfaction of TPCs, preservation of TPCs* and *deletion of TPCs*. Moreover, the DMAC can be divided into two sections. One is *DMAC for disassembly without tools* and the other one is *DMAC for disassembly with tools* To sum up, the DMAC can support disassembly either with or without tools.
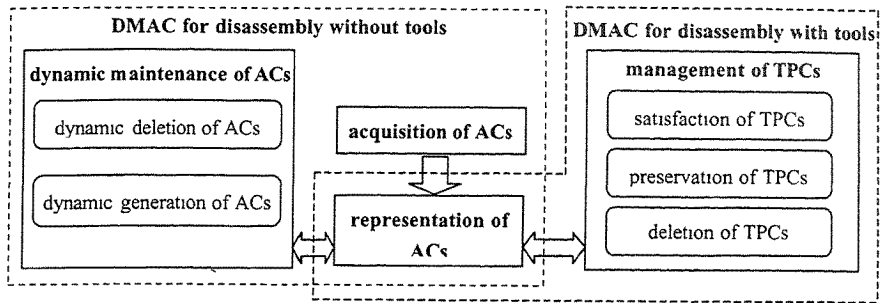
Fig. 1 Dynamic management mechanism of assembly constraints

4 DMAC FOR DISASSEMBLY WITHOUT TOOLS

Generation of ACs is a problem of constraints identification and it will not be discussed in this paper. Acquisition and representation of ACs and its dynamic deletion will be addressed.

4.1 Acquisition of ACs

ACs are designated at the time of assembly modelling. They can meet the demand for assembly but are not sufficient for disassembly. As shown in Fig. 2, the simple assembly constitutes of three components, A, B and C. It is supposed that B and C are assembled first. Then only the assembly constraints between A and B need to be designated for assembly of A. Consequently, there exists a mate AC between A and B and it is transferred by the assembly constraints between B and C. So a mate AC between A and C exists, which is necessary for disassembly of component C for it restricts the movement of C along the upward direction.
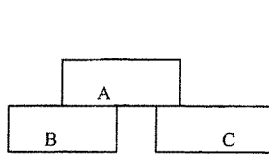
Fig. 2 A simple assembly
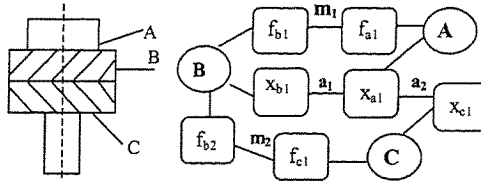
Fig. 3 An assembly and the assembly constraint graph

Thereby, acquisition of ACs in virtual environment can be achieved by combination of constraints reconstruction and constraints transfer. Constraints reconstruction is to reconstruct ACs that have been designated in CAD system. Constraints transfer is to obtain the connotative ACs via the transfer of those ACs already designated. The advantage of this approach is that constraints identification can be avoided which can greatly enhance the efficiency of acquisition of ACs. The method of constraints transfer is obvious. Means of constraints reconstruction is described in succession. The basic elements of ACs are either topological geometry elements (such as the mate planes in mate ACs) or assembly features (such as the axes in axes align ACs). In other words, ACs are based on topological geometry elements and assembly features. Therefore, they can be reconstructed in virtual environment

(VE) through the following steps. Firstly, extract data in CAD system and save it in a file in neutral format. The data to be extracted include topological and geometric elements of components, assembly features, assembly constraints and so on. Secondly, reconstruct the topological relationship of the geometry elements of components in VE. That is to map topological relationship of the geometry elements of components from CAD system to VE according to data in the saved file. Finally, form the assembly constraints in VE, which is automatically done based on the reconstructed data in VE and data extracted in CAD system.

4.2 Representation of ACs ·

Assembly constraints graph (ACG) is used to represent ACs and the ACG can represent various ACs as well as their geometry elements. This paper will focus on representation of mate ACs and axes align ACs for they are the most common ACs in an assembly. The ACG is a non-directed graph and is formulated as $ACG=<V, E>$, where V formulated as $V=\{P, GE\}$ is the set of nodes and E formulated as $E=\{Element, Mate, Align, \bullet \}$ is the set of edges. Nodes in the graph represent either components described as $P=\{A, B, C, D, \bullet \}$ or geometry elements of components described as $GE=\{F, X, \bullet \}$ which can be classified to plane nodes, axis nodes and so on. The plane nodes are described as $F=\{f_{a1}, f_{a2}, \bullet \; f_{b1}, f_{b2}, \bullet \}$, where f_{a1} indicates the first plane of component A. The axis nodes are described as $X=\{x_{a1}, x_{a2}, \bullet \; x_{b1}, x_{b2}, \bullet \}$, where x_{a1} indicates the first axis of component A. Edges in the graph represent either geometry elements belonging to components or assembly constraints. To be specific:

Element$=\{e|e=(p, ge), p\bullet P, ge\bullet GE\}$, geometry element ge belonging to component p.

Mate$=\{m|m=(f_1, f_2), f_1, f_2\bullet F\}$, mate AC with two mate planes f_1 and f_2.

Align$=\{a|a=(x_1, x_2), x_1, x_2\bullet X\}$, axes align AC with two axes x_1 and x_2.

As shown in Fig.3, B and C are joined together by A, and the corresponding ACG is also shown. The kind of Element edges is not labeled for purpose of simplification.

For a simple assembly, its ACG can be directly generated. But for a complicated assembly, ACGs of each subassembly are generated first, and then these ACGs are synthesized to generate the ACG of the assembly based on the hierarchical information of the assembly.

4.3 Dynamic deletion of ACs

The ordinary approach to delete ACs is tracking ACs, judging their validity and deleting invalid ACs in the ACG. This approach is called *direct deletion of ACs* with judgment of the validity of ACs conducted for each AC. Hence it is computationally time consuming.

In fact, during disassembly there exist two situations of invalidity of ACs. On one hand, for a specific disassembly operation, some ACs in the assembly will become invalid simultaneously. Again as shown in Fig.2, if A is disassembled by translation along the upward direction, the mate ACs between A and B, between A and C will become invalid simultaneously. On the other hand, disassembly is always supposed to be monotonic disassembly, in which a component is fully disassembled from the assembly only by a single disassembly operation. Therefore, ACs of a component will become invalid orderly during its disassembly. As illustrated in Fig.3, when A is disassembled by translation along the axis direction, its mate AC m_1, axes align AC a_2 and a_1 become invalid orderly. Conclusions can

be drawn that ACs may become invalid simultaneously or orderly. Symbols '=' and '>' are applied to indicate the relation of ACs to be invalid from the point of view of time. For example, $m_1=m_2$ indicates that AC m_1 and m_2 become invalid simultaneously, and $m_1>m_2$ indicates that AC m_1 becomes invalid before AC m_2. *Reference constraint* (RC) is introduced in this paper. If ACs become invalid simultaneously, namely if $c_1=c_2=c_3$, where c represents any AC, then any of the ACs can be chosen as RC. If ACs become invalid orderly, namely if $c_1>c_2>c_3\cdot$ •then c_3, the last AC to become invalid can be chosen as RC. During disassembly, only RC needs to be judged for invalidity. The reason is that if RC becomes invalid, the other ACs related to it will have already been invalid, which is assured by the simultaneous relation and the sequential relation of ACs to be invalid. It shows that there is a propagation effect of the invalidity of ACs. Based on the above analysis, a so-called *indirect deletion of ACs* approach is proposed with its algorithm described as follows. The approach implements dynamic deletion of ACs by judging the invalidity of RC and deleting invalid ACs based on the propagation effect of the invalidity of ACs, so that it can greatly enhance the efficiency of deletion of ACs compared with the *direct deletion of ACs* approach.

Algorithm of indirect deletion of ACs

Known: assembly constraints graph *ACG*, component *P* to be disassembled and the disassembly operation *M*

Step1 Establish simultaneous relation R_1 and sequential relation R_2, choose reference constraints RC_1, RC_2 for R_1 and R_2 respectively.

Step2 Disassemble *P* by *M*, track RC_1, RC_2 and judge their invalidity.

Step3 If RC_1 or RC_2 becomes invalid, then delete the related ACs in *ACG* and update *ACG* according to R_1 and R_2.

Step4 Execute step2 and step3 until both RC_1 and RC_2 become invalid.

In the above algorithm, for a disassembly operation, rules are used to establish R_1 and R_2. Specifically, the simultaneous relation R_1 are established according to Rule1 and Rule2, and the sequential relation R_2 are established according to Rule3 and Rule4.

    Rule1   For mate ACs, build R_1 for those ACs whose normals of the mate planes are parallel to the disassembly direction.

    Rule2   For axes align ACs, build R_1 for those ACs whose axes are parallel to the disassembly direction and the align length are equal.

    Rule3   For mate ACs, those ACs whose normals of the mate planes are parallel to the disassembly direction become invalid firstly.

    Rule4   For axes align ACs, establish R_2 according to align length. Typically, the shorter the align length, the earlier the ACs become invalid.

5 DMAC FOR DISASSEMBLY WITH TOOLS

DMAC for disassembly with tools is to manage TPCs when disassembly with tools. The representation and deletion of TPCs are similar to that of DMAC for disassembly without tools. The preservation of TPCs is achieved by considering the tool and the fastener as a

whole and keeping the TPCs no changes. Satisfaction of TPCs is the corresponding process of tool positioning, and it will be discussed at length subsequently.

User can manipulate the objects in VE by data gloves. Nevertheless, because the movement of virtual hand in VE is inaccurate and unsteady, it is difficult to achieve accurate positioning of tools in VE. Here, a three-step positioning manner including orientation of tools, approximate placement of tools and automatic positioning of tools is presented to accurately position tools in VE. In orientation of tools, the pose of the tool is adjusted based on that of the fastener. In approximate placement of tools, the tool is moved by virtual hand to access to the fastener based on constraints matching. For example, a mate AC is termed 'matched' if normals of two planes are almost parallel and are in the opposite directions, and the distance between the planes is within the tolerance limit. In automatic positioning of tools, transformation matrix of the tool is calculated to get the translation vector and the rotation vector, and then the tool is translated and rotated according to the vectors. The transformation matrix of the tool can be calculated by the following formulae:

$$M_{t2} = M_{t1} \cdot M_{trans} \qquad M_{t2} = M_p \cdot M_{tp} \qquad M_{trans} = M_{t1}^{-1} \cdot M_{t2} = M_{t1}^{-1} \cdot M_p \cdot M_{tp}$$

Where M_{t1} is the azimuth matrix of the tool after its approximate placement. M_{t2} is the azimuth matrix of the tool after its accurate positioning. M_{trans} is the expected transformation matrix of the tool in the global reference frame. M_p is the azimuth matrix of the fastener. M_{tp} is the transformation matrix of the tool relative to fastener after its accurate positioning in the local reference frame of the fastener.

6 IMPLEMENTATION AND CASE STUDY

An assembly of bell-roller (shown in Fig.4) made up of thirteen components including fasteners like screws and nuts is disassembled in virtual environment based on the dynamic management mechanism of assembly constraints. The assembly is modelled in Pro/Engineer[TM], and such data as topological geometry elements of components and assembly constraints are extracted using Pro/ToolKit[TM]. Envision[TM] of Deneb Company is the software platform, and the disassembly is accomplished in virtual environment using CyberGlove[TM]
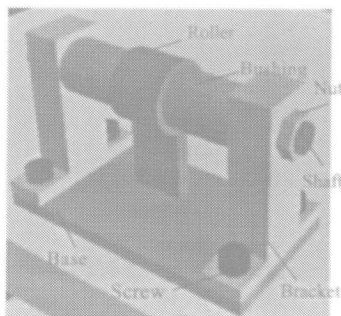
Fig. 4 Assembly of bell-roller
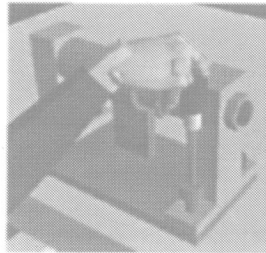
Fig. 5 Disassembly of screw with screwdriver
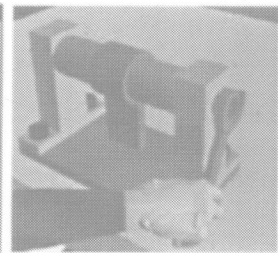
Fig. 6 Disassembly of nut with wrench

data glove and V6 HMD. Fig.5 shows the disassembly operation of screw with screwdriver. Fig.6 shows the disassembly operation of nut with wrench.

Fig.7 shows the ACG before component K is disassembled. Here, the number of ACs of K is three, which are m_2 between K and J, m_8 between K and M, and a_{13} between K and L. Apparently, they will become invalid in the order $m_2 > m_8 > a_{13}$ when K is disassembled by translation along axis direction. So, a_{13} is chosen as the RC. Track a_{13} and judge its validity. When a_{13} becomes invalid, delete m_{12}, m_8 and a_{13}. The updated ACG is shown in Fig.8.
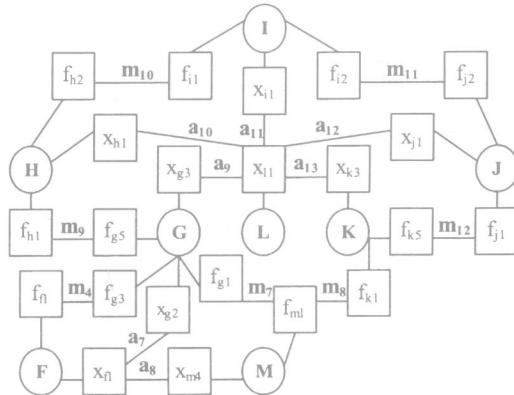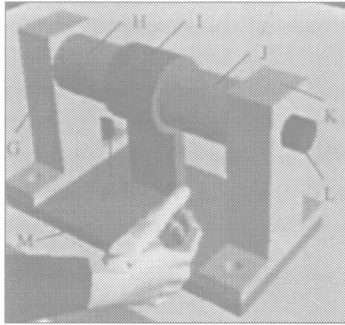
Fig. 7 ACG during disassembly (before component K is disassembled)
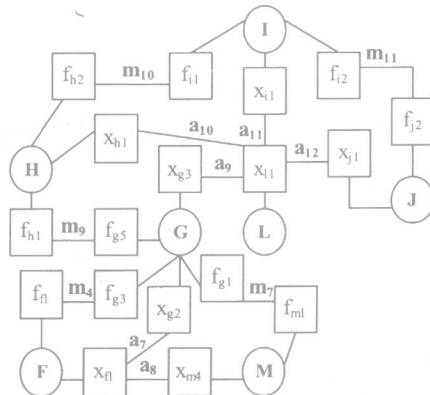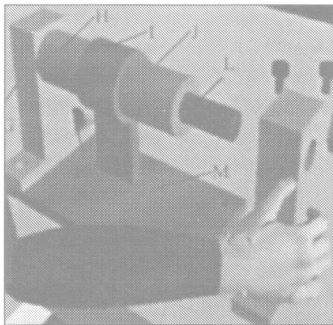
Fig. 8 ACG during disassembly (after component K is disassembled)

7 SUMMARY AND FUTURE WORK

The dynamic management mechanism of assembly constraints can support virtual disassembly either with or without tools. Assembly constraints acquired by constraints reconstruction combined with constraints transfer avoids constraints identification. The approach of indirect deletion of assembly constraints enhances the efficiency of deletion of

assembly constraints during disassembly. The management of tool positioning constraints enables disassembly with tools. Our future work will focus on analysis of accessibility of the tool and stability of the assembly during disassembly to make virtual disassembly more authentic and reliable.

ACKNOWLEDGEMENTS

REFERENCE

1 **Kuo, T. C.** Disassembly sequence and cost analysis for electromechanical products. Robotics and Computer Integrated Manufacturing 2000; 16(1): 43-54.

2 **Siddique, Z., Rosen, D. W.** A virtual prototyping approach to product disassembly reasoning. Computer-Aided Design 1997; 29(12): 847-860.

3 **Gadh, R., Srinivasan, H., Nueeghalli, S., et al.** Virtual disassembly - a software tool for developing product dismantling and maintenance systems. Proceedings of the IEEE Symposium on Reliability and Maintainability 1998; 120–125.

4 **Srinivasan, H., Shyamsundar, N., Gadh, R.** A virtual disassembly tool to support environmentally conscious product design. Proceedings of the IEEE Symposium on Electronics and the Environment 1997; 7-12.

5 **Srinivasan, H., Figueroa, R., Gadh, R.** Selective disassembly for virtual prototyping as applied to de-manufacturing. Robotics and Computer Integrated Manufacturing 1999; 15: 231-245.

6 **Lee, k., Gadh, R.** Destructive disassembly to support virtual prototyping. IIE Transactions 1998; 30: 959-972.

7 **de Sa' , A. G., Zachmann, G.** Virtual reality as a tool for verification of assembly and maintenance processes. Computers & Graphics 1999; 23: 389-403.

8 **Fernando, T., Fa, M., Dew, P. M., et al.** Constrained-based 3D manipulation techniques within virtual environment. Virtual Reality Applications. ed. Earnshaw R A. Academic Press, London, 1995; 71-89.

9 **Fernando, T., Wimalaratne, P., Tan, K.** Constraint-based virtual environment for supporting assembly and maintainability tasks. Proceedings of ASME DETC99/CIE-9043.

10 **Jayaram, S., Connacher, H., Lyons, K.** Virtual assembly using Virtual Reality Techniques. Computer-Aided Design 1997; 29(8): 575-584.

11 **Jayaram, S., Wang, Y., Jayaram, U., et al.** A Virtual assembly design environment. Proceedings of IEEE Virtual Reality 1999; 172-179.

12 **Jayaram, U., Tirumali, H., Jayaram, S.** A tool/part/human interaction model for assembly in virtual environments. Proceedings of ASME DETC2000/CIE-14584.

Some issues in LCA for manufacturing industries

C DENG and **P LI**
Huazhong University of Science and Technology, Wuhan, China

ABSTRACT

This paper provides a discussion of some key issues in LCA for manufacturing industries. LCA plays an important role in the field of manufacturing industries. For the use of LCA in the industry, a specific feature is a meaningful approach to continuously and routinely use LCA for one product instead of conducting LCA on a project basis just once in a while. The realities of LCA research and application reveal that some inconsistent issues exist, such as methodology development versus application of LCA, application dependency of LCA, research and investment of LCA. To meet the needs of manufacturing industries, LCA must not be used as a stand-alone-tool. It has to be integrated to the regular business decision-making context of a particular company. Two strategies are proposed to put LCA research and application into right direction: implementation of LCA into CIMS, and collaborations in the region and internationally.

1 INTRODUCTION

Life Cycle Assessment (LCA) is an information and knowledge intensive science, which is now well established both as an interdisciplinary science discipline and an important environmental management tool for companies. In addition, LCA is part of the international standardization of Environmental Management System (EMS) which is documented by the ISO 14000 series. On the surface LCA looks deceptive simple – the measurement of the physical flow of resources and pollutants. However the further into LCA people progress the more questions that arise concerning:
  • The nature of processes, products and production systems;
  • What represents the environment and how can it, and should it, be measured;
  • The interplay of market forces, supplies and consumers;
  • Decision making theory.
Therefore, the establishment and stabilization of LCA should be very decisively concerned, because many industries, including manufacturing industries are currently still in the process of evaluating the use of LCA, to decide whether they are going to use LCA, how they are going to use it and the purposes and applications for which they will use it.

This paper tries to address some key issues in LCA for manufacturing industries. The role of manufacturing industries in the field of LCA is important. First, manufacturing industries have a large physical material output. Second, manufacturing industries are customers of material suppliers. The intermediate role in the supply chain between material suppliers and consumers is relevant for promoting LCA. In other words, the production system behind the products society consumes should be explained. How the systems interaction with the environment can be tracked by measuring hundreds of environmental flows, also should be explained. From the viewpoint of LCA, an important feature is that products are developed and produced on a regular basis. Because a meaningful application of LCA will require a continuous execution of LCA case studies as well. However, some inconsistent issues exist within LCA research and application.

2 REALITIES OF LCA RESEARCH AND APPLICATION

2.1 Methodology development versus application of LCA

It is evident that the separation of methodology development and application of LCA is suggested to promote LCA in the manufacturing industries. As shown in Figure 1, there is both synergy, but also tension between the development and application of LCA methodology. The point is, that exactly the joint of research and application often leads to disappointment of the capabilities and usefulness of LCA. Usually, the focus on methodological discussions has to be critically evaluated. Not necessarily the choice of the "best" methodology is a crucial point for a successful project, but a meaningful application of LCA, which adapts pure methodology to the real world, copes with real industries, with real products, with real data, with conflicting interests and a large number of stakeholders from basis for application. However, to apply pure research results is not realistic. The improved mixing methodology development and application should be studied.

Figure 1. Methodology development versus application of LCA

2.2 Application dependency of LCA

In recent LCA methodology development, it is found that the application dependency of LCA has direct and beneficial consequences for the practical application of LCA. One of the main reasons for disappointment of the usefulness of LCA was that in the past, one conducted one comprehensive LCA to use for everything – from comparison to optimisation. It primarily

aimed at describing reality as accurate as possible to answering a certain question of technosphere. However, there is a second important requirement for LCAs – the suitability and validity to answer certain questions - which was neglected for LCAs. By using a particular single LCA, which aimed at a very detailed description of the given system, one tried to answer the most different questions for all kinds of applications. As a result, none of the purposes was achieved in satisfactory manner. LCA was promoted from its practitioners as a tool that can do almost everything, if just enough time and money is spent for making a complete or full LCA. However, this did not work, because the questions to be answered by LCA were too broad and general.

The application dependency reflects a mechanistic philosophy. The more high quality of an LCA increases, the more processes, parameters or emissions are considered. However, the quality of a model is not determined by its complexity or quantity of parameters, but by the quality and validity of the conclusions drawn from the model. As a consequence, more emphasis has to be put on the actual questions, that have to be tackled with a particular LCA. This decision-oriented approach can lead to a higher quality and validity of the conclusions from the model. Actually, this concept is a direct consequence of taking goal and scope definition serious and highlights, which is the central important phase of LCA. However, using a single LCA for different purposes is neither efficient nor effective. It is not efficient, because the model might provide information that is not necessary to answer a specific question. It is not effective, because if main requirements of a specific question are neglected this model might even lead to wrong conclusions.

2.3 Research and investment of LCA
As we had known, individual researchers can answer some of LCA questions in a project specific manner. Others require much broader input and consensus if the answer is to be of any value in the overall scheme of decision-making. Impact assessment, allocation methods, and the critical process are the examples, which require some level broad acceptances. So, there is a clear need for LCA researchers, developers, practitioners and end users to work together in answering many of these questions.

Environmental performance is often touted as a competitive edge for companies in an increasingly environmentally aware market place. However, the commercial realities of LCA research and practice counteract this need. Realities, such as the commercial confidentiality of company products and services, often limit the release of information and particularly basic inventory data. Also, knowledge and information sources developed by LCA research organizations are usually required to make some commercial return, which often limits the free exchange of information and data within the research environment. This problem is not unique to LCA, and is encountered in all areas of science and engineering.

Even then, in the new era of globalization and the information economy, there are the beginnings of a shift away from this problem. In the computer world products such as Linux, Adobe, Netscape, Encyclopaedia Britannica and many more, are given away to consumers, with profits being made by providing services to consumers, assisting in the use and customization of the product and through advertising. This shift from a product to service approach has been one of the challenges of LCA and Design for Environment advocates in recent years, so maybe its time to practice what we preach in manufacturing industries.

3 PUT THE LCA RESEARCH AND APPLICATION INTO RIGHT DIRECTION

3.1 Implementation of LCA into CIMS

From an industry perspective, LCA will not survive as a stand-alone-tool. It is a normal process that companies evaluate new tools and concepts as such on a project basis. After this evaluation process, a particular company will decide, if the new tool or concept will be abandoned or whether it will be implemented. Therefore, it is crucial to position LCA as a useful tool for already existing processes and procedures within companies. There is definitely no single solution, how LCA can be best applied within the business decision making context. Each company has to solve and decide that case by case. It depends on the size and culture of the company, the products produced, the strategy, the internal tools and procedures, and many other things. Therefore, if LCA as a tool in Computer Integrated Manufacturing System (CIMS), the applications of LCA will have the largest potential , as shown in Figure 2.
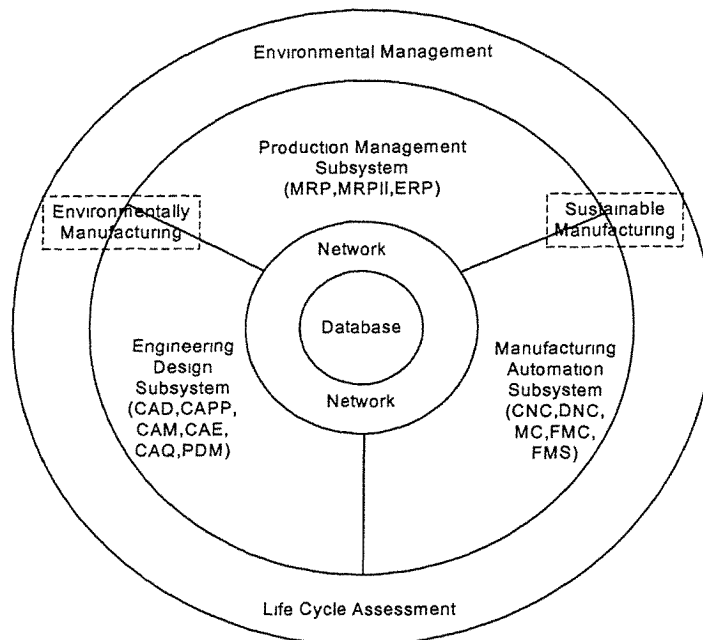
Figure 2. The basic framework of CIMS with LCA

In the past century, none of the definitions of CIMS was concerned with environmental issues. However, resource saving, environmental protection and labour protection have become increasingly important. Moreover, in many countries regulations have become legally effective restrictions to the manufacturer with the responsibility for the complete life cycle of a product, including reusing and recycling of used products. The environmental protection and minimum consumption of resources during the life cycle of products are vital for manufacturers. On the other hand, there has been an increasing interest in the implementation of CIMS in different types of enterprises all over the world. The two tendencies suggest that it would be a good idea to integrate sustainable development with CIMS. Such a green-

manufacturing strategy can help manufacturing enterprises to achieve an environmentally and sustainable manufacturing, which can considerably improve the competitiveness of manufacturing enterprises.

3.2 International collaborations

A couple of the major international activities which are underway or have been completed at an international level. The work in the international standards arena on LCA by ISO, has possibly been the most important piece of collaboration to take LCA beyond its home base in Europe. They allow, at least, the concept to be disseminated to the rest of the world. It has given LCA legitimacy amongst governments and companies, and combined with the other standards on environmental management (including environmental management systems and eco-labelling), which has given LCA some very tangible applications. In addition, the Society of Environmental Toxicology and Chemistry (SETAC) working groups have given us much of our agreed methodology, and have produced a great deal of the information included in the standards. Further, they have contributed beyond the scope of the standards in areas such as the best practice work on the impact assessment indicators, data working groups and many other methodological developments. There are some possible collaborations on LCA.

3.2.1 To initiate collaborative projects

With the increasing global hazards, the issues of environmental protection become globalization. International collaborative projects facilitate the sharing of expertise and experience across the region. Further, it improves the understanding of the challenges faced by different countries in applying LCA. Working with people from other countries also gives insights into how they have solved, or failed to solve, some of the major problems areas of LCA data availability, allocation, impact assessment.

3.2.2 To establish regional LCA forum

Many countries have established some form of National forum or society for networking LCA issues. Establishing a regional forum is of benefit to reduce duplication and helps in coordinating activities. Most national fora have looked at issues of data availability and generation and localized impact assessment methods as they are faced with basic problems in the lack of data supply and therefore need to draw heavily of European studies. Formal links and cooperation between these groups, would be beneficial to each of the groups, and to the co-ordinated development of LCA in the region.

3.2.3 To share LCA inventory data

Obtaining inventory data is one of the hardest tasks facing LCA practitioners in the region. Obtaining data on material, or products from other countries is even more difficult. While completed inventory data is rarely available, in many countries there are public sources of data available from environment departments and some individual companies, such as the International Energy Agency could also be referenced here to find data for different countries. As public LCA data becomes available, it will be important for users of the data from other countries, to be able to check assumptions and system boundaries with the country where the data was generated. This local knowledge is not always conveyed with the data, and strong voluntary networks of LCA practitioners would assist in the dissemination of this knowledge.

3.2.4 To develop toolkits and resources guides

There is an immediate need for an easily understood, practical toolkit for undertaking and integrating LCAs for use by industry. The target audience should be small and medium-sized companies, as these are the companies that find entrance to LCA more problematic. Notwithstanding that, such a guide would also be useful for academics new to the area, as well as larger industries.

3.2.5 To run workshops and seminars

For countries that are still developing their understanding of the basics of LCA, seminars and workshops on undertaking and understanding LCA would be a good introduction. They could provide opportunity to share skills and promote LCA in general.

4 CONCLUSION

LCA, as a methodology, deals with manufacturing industries indispensably. Notwithstanding there are some inconsistencies in the development methods, application and investment of LCA, the LCA network for integration into CIMS, and for regional/international cooperation will be developed and promoted.

ACKNOWLEDGEMENT

REFERENCES

1 **J. Gertsakis, H. Lewis** and **C. Ryan.** (1997) A Guide to EcoReDesgin: Improving the environmental performance of manufactured products, The Centre for Design at RMIT with Environment Australia.

2 **Genbao Zhang, Ying Liu, Wei Tan** and **Zhongquan Huang.** (2001) S-CIMS: Towards Integrated and Environmentally Conscious Manufacturing Systems. Proceedings of DETC'01, Pittsburgh, Pennsylvania, USA, Sept. 9-12.

3 **E. Westkämper, L. Alting** and **G. Arndt.** (2000) Life Cycle Management and Assessment: Approaches and Visions Towards Sustainable Manufacturing. Annals of the CIRP Vol.49/2, pp. 501-522.

4 **Mary Ann Curran.** (1999) The Status of LCA in the USA. The International Journal of Life Cycle Assessment, Vol.4/3, pp. 123–124.

5 **R.Margent, et.al.** (1999) Integrated models for environmental management: issues of process and design. Environment International, Vol.25, No6/7, pp. 693–699.

A hierarchical approach to assembly sequence planning

X NIU and H DING
School of Mechanical Engineering, Shanghai Jiaotong University, China
Y XIONG
School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, People's Republic of China

ABSTRACT

This paper presents an approach to find feasible and practical plans for mechanical assemblies based on the hierarchy partition of the assembly. The basic idea of the approach is to construct the hierarchical structure of a product by using group and subassembly cluster methods, then get the optimal of assembly sequence for each structure level by genetic algorithm. As a result, a better assembly sequence of the product can be generated by combining the assembly sequences of all hierarchical structures, which provides more parallelism and flexibility for assembly operations. An industrial example is solved by this new approach.

1 Introduction

Assembly sequence planning aims to identify and evaluate the different ways to construct a product from its components. A product may always naturally be arranged into a hierarchy, with the large, high-level structures containing smaller substructures. If a product is first divided into substructures, which is also continually divided into small substructures, assembly sequences of the product then can be decreased largely, which always is also adopted in industry. Heemskerk[1] proposes some heuristics in assembly sequence planning. Based on the geometry of parts, Moradi[2] develops the method of part grouping for assembly planning. Ong[3] describes a methodology for automatically extracting subassemblies from a product. Sugato Chakrabarty and Jan Wolter[4] present a structure-oriented approach to assembly sequence planning, but the structure hierarchy of a product needs to be identified by the user.

To find good assembly sequences is even more important, most researchers[5,6] use traditional AI tree search or graph search methods to find an optimal assembly sequence. Essentially, they perform part-by-part search. Therefore, theses methods often have the problem of getting trapped in combinatorial complexity. To overcome this complexity, other researchers use sequence-by-sequence search, such as Hopfield neural networks (NN) approach[7] or the simulated annealing (SA) method[8]. As an alternative search method, genetic algorithm (GA) has proved to be highly effective in solving NP-hard problems and some successful applications to assembly and process planning have been already proposed[9][10][11].

Combined with hierarchical strategy and genetic algorithm, this paper proposes a novel approach to assembly sequence planning. According to the geometric and non-geometric data for assembly, a product is automatically partitioned into certain hierarchical structures, including groups and subassemblies. Then GA is used to search out the best assembly sequence for each hierarchical level.

2 Description of an Assembly

In the paper, the process of assembly is assumed to be the reverse of disassembly, and the assembly sequences can be obtained by reversing the disassembly sequences. An assembly can be described to a computer by specifying its components and their relationships in the assembly. In this work, the components themselves are specified by their boundary representations and their relationships in an assembly are specified by the mating conditions between all the mating components.

Let a product $A=(P, L)$ consists of n parts interconnected by r liaisons, where $P=\{p_i|i=1, 2, \ldots, n\}$, $L=\{l_{ij}|i, j=1, 2, \ldots, n. \ i \neq j\}$, and $(n-1) \leqslant r \geqslant n(n-1)/2$. A liaison l_{ij} represents the mating conditions between a pair of parts p_i and p_j. The mating conditions can be divided into against conditions and fit ones. The against mating condition, in general, constrains two planar faces in such a way that their normal vectors align in the opposite directions. The fit mating condition aligns the center axes of two cylindrical faces. In order to derive the blocking relations of the parts by a fit mating condition, it is necessary to extend the fit mating condition. Here we consider four typical fit conditions: fit, shaft-in-hole(SIH) fit, thread fit, and taper fit. Of course, the implied mating conditions between parts can be inferred from ones explicitly defined in the assembly model.

3 Hierarchical Structure Extraction

Components clusters are very effective to reduce the disassembly complexity. According to the geometric and non-geometric data of a product model, two categories of components can be defined: groups and subassemblies, then a hierarchical structure of the assembly may be obtained by the extraction of groups and subassemblies in turn.

3.1 Groups Extraction

Definition 1: A *group* of components is a set of components in the assembly, generally without contact between them, which can be assembled/removed only in one direction, but in any order with respect to each other.

Grouping parts and treating each group as a single part can speed up the search for an optimal plan. The implicit result of group representation is the reduction in the number of reorientations in assembly plans, which is one of the optimality criteria in assembly sequence planning. Typically, a set of screws constraining the same parts may be grouped together. The characteristics of a group of components are:

- Each element is a single part
- All parts are of the same type
- All parts must have similar relation (with similar disassembly direction) at similar relative positions with the same parts

The principal rule in choosing a group is that the assembly of a group may not be interrupted by any part outside the group. In other words, the translation of parts in a group is not blocked by any part outside the group in the given direction. Thus, the group recognition process can be described as follows:

1) Check the Bill Of Material (BOM) for parts that occur more than once. Each groups of these parts is a cluster candidate.
2) Check the relations of each part:
   a) Make subgroups of parts with the same number of relation
   b) Make subgroups according to corresponding related parts
   c) Make subgroups according to corresponding main disassembly direction

3.2 Subassemblies Extraction

Definition 2: A *subassembly* is a set of composed of two or more components that can be handled as a whole. The geometric and technological characteristics of its components may be different from each other.

In order to detect subassemblies successfully, consideration of a base element is necessary. In a practical product, there will be one component which will serve as a base for the assembly or disassembly of the product. Using the criteria of maximum number of connections, maximum weight, maximum volume, and so on, the base part can be determined. According to the geometric and technological characteristics of components, four heuristic rules have been set up to automate the determination of subassemblies:

1) The contact fits provide an easy and efficient way to create subassemblies. Such typical subassemblies can be a housing or a shaft with ball bearing tightly fitted onto it.
2) The case that some components are not easily separated leads to the definition of one or more subassembly clustering rules. Typical subassemblies can be a cluster of screws and all components in contact with it by thread fit mating conditions.
3) The block relationships between components forming loops also imply the

existence of subassemblies. It is possible to define a matched contact for each two adjacent components and hence their relative mobilities. Therefore, if no translation exists between two components, a subassembly will be created.

4) The base component is the last component left after the complete disassembly process has been carried out, remaining not entered into any subassembly cluster.

Based on the extraction of groups and subassemblies in the assembly, the hierarchical structure of a product finally can be constructed. It is obvious that no hierarchical relation exists between group clusters, whereas a subassembly may contains smaller subassemblies and groups. A product can be described with single parts, groups of some parts and subassemblies. Assembly sequence planning can be performed at each structure level from high to low order, where both a subassembly and a group will be treated as a single part, and then these plans are merged to get the final plans. It can be proved that the proposed hierarchical approach is both correct and complete.

4 Precedence Constraints

Assembly operations cannot be implemented in a random order because some operations may prevent the execution of others owing to geometric or other constraints. These operation constraints are called precedence constraints, which can be separated into geometric and non-geometric constraints. Geometric constraints are those which relate to product geometry, such as contact, or interference. Non-geometric constraints represent the information that is not related to product geometry but which affect assembly methods. In this paper, only geometric constraints are taken into account, which may be deduced from the mating relations model of an assembly.

The assembly sequences that satisfy precedence constraints are called the feasible assembly sequences. Therefore, the set of precedence constraints for all components is considered as a doorsill, which can judge an assembly sequence is feasible or infeasible. Under the conditions of hierarchical strategy, the concept of precedence constraint can be extended to the case that subassemblies and groups occur in the assembly.

5 Generation of Assembly Sequences Using GA

For assembly sequence planning of each structure level, GA is used to get the best assembly sequence. If an disassembly sequence is represented as an chromosome, which is assigned a fitness related to the disassembly cost, then as a result of GA evolution, the disassembly sequences represented by the fittest chromosomes will be finally found.

5.1 Codifying chromosomes

The first step in developing GA for assembly sequence planning is to map the problem solutions (disassembly sequences) to chromosomes. In our system, a genetic chromosome (individual) C is represented as $(v_1, v_2, \ldots, v_N)$, where v_i expresses a component part that may

be a single component, a group or a subassembly, and N is number of component parts at certain structure level.

5.2 Fitness

In general, chromosomes may correspond to unfeasible disassembly sequences. In order to identify whether each chromosome represents a feasible disassembly sequence or not, precedence constraints for every component should be generated. For infeasible disassembly sequences, the feasible degree, which is the number of the feasible genes in the chromosome, can be used to evaluate them.

Each chromosome is assigned a fitness, which means the probability that the chromosome will survive and create offspring in the next generation. The fitness associated with a given chromosome is calculated as a weighted sum:

$$f(c) = w_1 I + w_2(N-1-D) + w_2(N-1-M)$$

Where N is the number of component parts at certain structure level. I is the feasible degree of a chromosome, whose maximal number is N. D is the direction metric, which is represented as the number of direction changes of disassembly. M is the manipulability metric, which takes into account how hard a component is handled during disassembly operation. The weight w_i, $i = 1\sim3$, may be chosen as appropriate. In general, w_1 will be greater than the others to reflect the importance of the sequence feasibility.

5.3 Genetic operations

Initially, a set of individuals is randomly generated. Then, the population is evolved by natural selection, which produces the next generation through reproduction, crossover and mutation operations. Reproduction creates a new population from a population according to fitness scores of the original population. The traditional biased roulette wheel method is applied in our system. Random pairs of chromosomes mate by a crossover operation, by swapping portions to make new chromosome. In this study, the partially matched crossover (PMX) is adopted. Mutation swaps elements within a single chromosome, and swap operation is used in this paper.

6 An example

To illustrate the performance of proposed approach, we consider grinder vice shown in Fig 1. The mating relations model between parts is shown in Fig 2. In order to decrease the number of nodes in the model, the grouped parts are described as single nodes, and they are filled in ellipses. The symbols of a, f, s, th and ta for the liaisons respectively represent the mating conditions of against, fit, SIH fit, thread fit and taper fit. In the model, the symbols writed by boldface denote the implicit mating conditions, such as the SIH fit between part 1 and 2.

Using the approach of hierarchical extraction above mentioned, grinder vice is partitioned into four structure layers, as shown in Fig 3. In the first layer, the product contains only itself.

In the second layer, the assembly contains five single parts {1, 2, 3, 9, 10}, where part 3 is chosen as the base part, and one subassembly Sub1. In the third layer, subassembly Sub1 contains five single parts {5, 6, 8.1, 8.2, 14.2}, where part 6 is chosen as the base part, and four group component parts: G1{4.1~4.4}, G2{7.1~7.2}, G3{7.3~7.4}, G4{15.1~15.2}, and one subassembly Sub1.1. In the final structure layer, the subassembly Sub1.1 contains three single part {12, 13, 14.1}, where part 13 is chosen as the base part, and one group component part G5{11.1~11.2}.

It is noted that the example product contains four hierarchical levels, thus in order to get the best disassembly sequences of the product, four GAs should be used. The main difference of them is the length of chromosomes. Starting from a randomly set initial population composed of 100 chromosomes, at each generation chromosomes are selected for reproduction with a probability. We set the crossover and mutation probabilities to 0.95 and 0.90, respectively. The weighting factors in fitness are assumed to be w1=0.8, w2=0.5, and w3=0.5. Obviously, the first layer need not considered. As a result, by use of the genetic algorithm, the best disassembly sequences for every levels from the second layer to the fourth layer are generated: {10→1→2→9→Sub1→3}, {4.1~4.4→15.1~15.2→5→7.3~7.4→7.1~7.2 →8.1→8.2→14.2→Sub1.1→6} and {11.1~11.2→14.1→12→13}.
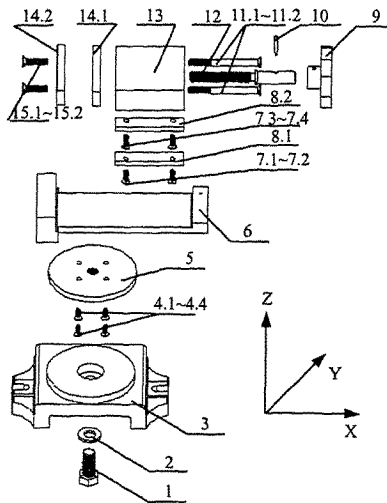
Fig 1. An exploded view of 25-part grinder vice
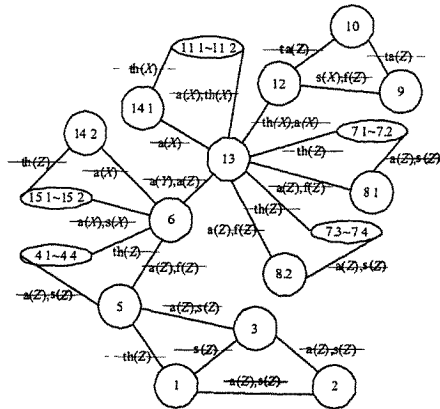
Fig 2. The mating relations model of grinder vice
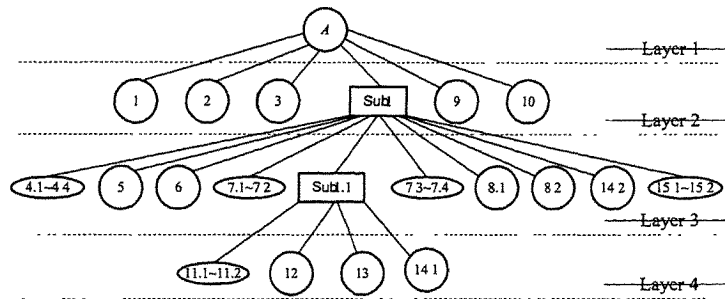
Fig 3. The hierarchical structure of the example product

6 Conclusions

This paper introduces a novel approach to solve assembly sequence planning problem by hierarchical structure extraction, which decreases the complexity of assembly sequence planning, and is especially suitable to plan some complex assemblies. Using genetic algorithm, the best disassembly sequence of each structure level can be easily generated. Consequently, a better disassembly sequence of a product can be found quickly. The disassemble sequences of each structure level may contain many parallel operations on the grouped parts or subassemblies, which can provide more parallelism and flexibility for assembly/disassembly planning.

REFERENCES

1   **Jr. Heemskerk C. J. M.** (1989) The use of heuristics in assembly sequence planning. Annals of the CIRP, 38(1), pp. 37–40.

2   **Moradi Hadi, Golberg Ken, Sukhan Lee.** (1997) Geometry-based part grouping for assembly planning. IEEE International Synposium on assembly and task planning, CA, pp. 281–286.

3   **N. S. Ong, Y. C. Wong.** (1999) Automatic subassembly detection from a product model for disassembly sequene generation. Int J Adv Manuf Technol, Vol. 15, pp. 425–431.

4   **Sugato Chakrabay, Jan Wolter.** (1997) A structure-oriented approach to assembly sequence planning. IEEE J. Robot. Automat, 13(1), pp. 14–29.

5   **De Fazio T L and Whitney D E.** (1987) Simplified generation of all mechanical assembly sequences. IEEE Journal of Robotics and Automation, 3(6), pp. 640–658.

6   **De Mello L. S. H. and Sanderson A. C.** (1991) A correct and complete algorithm for the generation of mechanical assembly sequences. IEEE J. Robot. Automat, 7(2), pp. 228–240.

7   **Hong D. S., Cho H. S.** (1995) A neural–network–based computational scheme for generating optimized robotic assembly sequences. Engng Appl. Artifical Intelligence, 8(2), pp. 129–145.

8   **Millner, J. M., et al.** (1994) Using simulated annealing to select least-cost assembly sequences. Proceeding of IEEE International Conference on Robotics and Automation, 2058-2063.

9   **Hong D. S., Cho H. S.** (1999) A genetic-algorithm-based approach to the generation of robotic assembly sequences. Control Engineering Practice, Vol. 7, pp. 151–159.

10 **Fujimoto H., Sebaaly M. F.** (2000) A new sequence evolution approach to assembly planning. Journal of Manufacturing Science and Engineering, Vol. 122, pp. 198–205.

11 **De Fazio T. L., Rhee S. J., and Whitney D. E.** (1999) Design-specific approach for assembly (DFA) for complex mechanical assemblies. IEEE J. Robot. Automat., 15(5), pp. 869–881.

Research on virtual/practical integrated material processing cell

H BIN, F XIONG, and J YANG
School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, People's Republic of China

SYNOPSIS

This paper presents a machining mode, named by Virtual/Practical Integrated Material Processing Cell (VPIMPC), which is based on existing machine and information technology on the shop floor. The purpose of the paper is focused on dealing with problems caused by complex machining operations. The configuration of VPIMPC is given and the characteristics are pointed out. Two case studies including taper ball end mill CNC grinding and selective laser sintering are offered to verify the feasibility and implementation of VPIMPC. The experimental results show that the VPIMPC may build a bridge to span the gap between practical environment and virtual environment.

1 INTRODUCTION

Recently, Virtual Manufacturing (VM) technology and system are widely used in manufacturing industry, which would simulate the machining process and be truly "WYSIWG" (What You See Is What You Get) where a machine tool or robot may be picked from a pull down menu and dragged into position into a manufacturing cell. Once in position the machine controller may be accessed by selecting the machine and be operated like the real thing. It requires CNC programs, offsets, stock material, tooling, and fixture details. A 3D virtual machining environment is displayed with CRT to control the simulation. Unlike the real thing, however, VM safely detects machine crashes, broken tooling, programming errors, and other problems.
The ability to produce an accurate CAD model from an "as-machined" solid is an emerging technology in NC verification that fills a void in today CAD/CAM machining processes.
Since the invention of the first NC machine tool, NC programming languages, and CAD/CAM systems, manufacturing engineers have needed an accurate electronic model of the initial and in-process state of material. It is difficult to confidently create tool paths for each subsequent operation without an accurate representation of the initial material. The operation accuracy, efficiency and correctness depend on the material initial geometric shape.

Until now the only way to get an in-process CAD model is with expensive, labor-intensive, error-prone, and inaccurate methods. These include manually modeling material state for each operation or using scanning equipment and surfacing software on the physical workpiece produced by the previous operation. But due to cost or time constraints these efforts are rarely undertaken. Without a good representation of the in-process material, NC programmers must visualize the geometrical shape of the material at each machining stage.

Access to the correct geometric workpiece shape in the CAD/CAM environment at each manufacturing stage, enables NC programmers to avoid problems when creating the subsequent cut. Some CAD/CAM developers attempt to provide this by using their existing CAD solid modeling tools or by producing a very rough approximation of the in-process material.

The CAD modeling approach is typically limited to $2^1/_2$ axis operations, and does not actually simulate the motion. Instead it makes a best guess at the region the operation is supposed to remove, creates a solid feature for the region, and subtracts it from the initial stock model. Often details produced by cutter radius, step-over, or step down are not included. And the size of the model can grow such that it becomes impossible to manage. The main difference between a verification model and a CAD model is verification software must be able to handle hundreds of thousands (or millions) of operations without failing or slowing. The software must also be able to allow for small inaccuracies in the data. Verification software is built to meet these requirements. After simulating an NC tool path, the resulting verification model can contain tens or hundreds of times more data than a traditional CAD solid model algorithm can process.

A material processor is capable of processing a part. Processing involves giving a new shape to the part by performing a series of machining operations whereby material is removed (or added) from (or to) the part. The resulting part may or may not have similar geometry as the original part, depending on the type of processing. For example, a grinding operation will remove minimal material thus retaining the shape as shown in Case Study I. On the other hand, a SLS (Selective Laser Sintering) operation may give a completely new geometry to the part as shown in Case Study II.

Computer simulation of the entire process enables a model of the entire system, allowing predictions of system performance. By making the simulation dynamic, effects of variability can be identified and any negative impact reduced. When this simulation is integrated with control design, the start-up occurs more quickly and with less pain.

Many large companies use CAD/CAM type simulation which can faithfully simulate actual manufacturing processes and as yet are beyond the reach of most small and medium size enterprises (SMEs), by nature of their complexity, high investment costs and high in-house technical skills required.

Even today, however, control engineers still build control programs based on information received mainly on paper as impulse diagrams, 2D drawings, and spread sheet tables. Unfortunately, there is no digital transfer of all that information already defined for simulation in a form adequate for the control engineer.

Furthermore, verification of control programs can only be made on the shop floor during the commissioning phase, where every error and lost day increases costs tremendously.

2 VIRTUAL/PRACTICAL INTEGRATED MATERIAL PROCESSING CELL

Based on the advances of virtual manufacturing environment and the conditions of small and medium sized enterprises (SMEs) especially in China, a machining mode, virtual/practical integrated material processing cell (VPIMPC) is proposed.

Here, material processing includes material removing and material increasing (or material additive) processes. The configuration of proposed VPIMPC is depicted as Fig .1. Several characteristics will be found from Fig.1.

• In the VPIMPC, CNC machining equipment exists on shop floor, so that it is unnecessary to design machine tool and fixture from CAD/CAM system. Thus, it will be suitable to SMEs and will enhance the productivity of the existing CNC machine tools by software development.

• In order to complete the digital transfer of component to be machined there are some engineering information needs to be digitized.
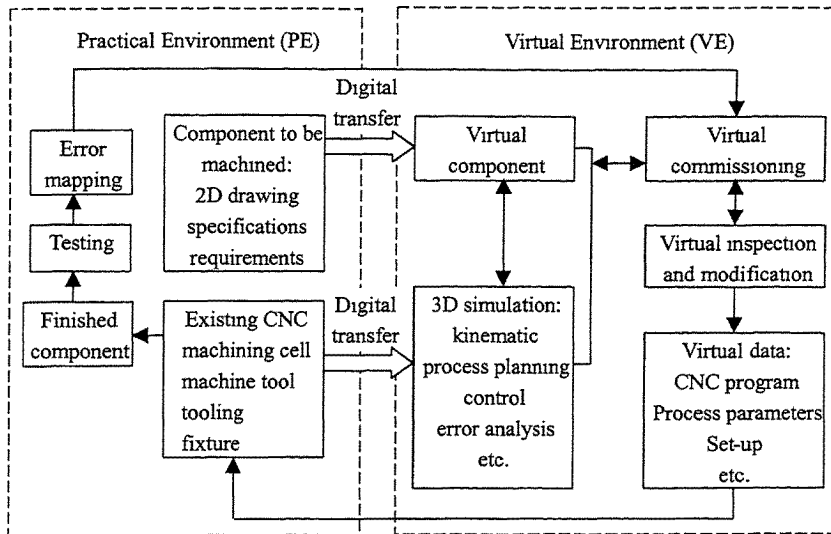
Figure 1. Configuration of VPIMPC

• The 3D simulation is used to make CNC verification and to show the operating procedure under a selected machining complex environment. The machining complex which is a closed-loop machining system consisting of machine-workpiece-tool-fixture may be variable because tooling systems may be different from each other. For example, using different types of grinding wheels, even though the machine tool structure is the same, the machining complex will be changed and the kinematics as well. Therefore, although the VPIMPC is specific to existing machines, lots of machining complex and operations will be created.

• Virtual inspection software is developed on the basis of virtual commissioning results and consideration only of geometric relationship, which is very useful to the grinding processes. In general, the grinding process is the finishing operation and determines the final accuracy of the part to be ground, and the grinding force is small enough to consider the deformation of the part.

• In many complicated machining operations, e.g. CNC spiral bevel gear cutting, CNC grinding of plunge gear shaver, precision leadscrew CNC grinding process etc., the error mapping from practical environment (PE) to virtual environment (VE) is of importance, through which the virtual commissioning process will be modified and the prototyping samples machined will be decreased greatly.

According to authors research experiences, the machining mode, VPIMPC, possesses the ability to build a bridge spanning the gap between PE and VE. It is necessary to deal with some physical or engineering problems in the future, revolving construction of a comprehensive system, especially to meet requirements from shop floor personals.

2.1 Case study I: Virtual/Practical Integrated Grinding Cell for Taper Ball End mill

A taper ball end mill is widely used in CNC machine, machining center etc.. Since this kind of tool has a complex cutting edge structure, it is often chosen to verify the performance of a five-axis simultaneous motion CNC tool grinder. Based on the mode indicated by Fig.1, a virtual /practical integrated grinding cell (VPIGC) is developed by several years efforts.

A self-developed CNC tool grinder, MMK6026, possesses 6 controllable axes, a photo of the grinder and a practical taper ball end mill are shown in Fig.2.
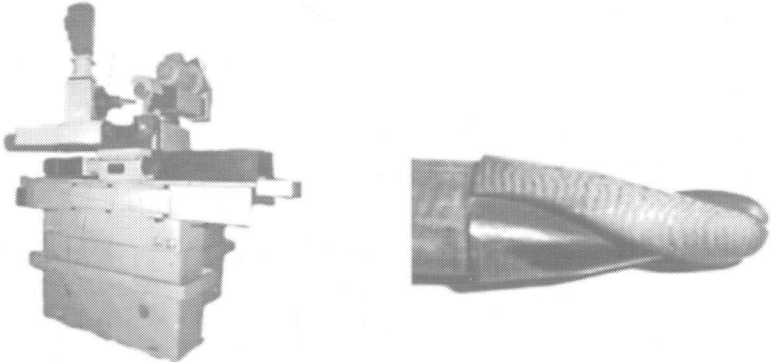
Figure 2. Practical grinder and component to be ground

Because the grinding wheel rotates in high speed so that the collision between the wheel and part to be ground will cause great damage. It is dangerous and expensive to verify CNC tool grinding process on a practical grinder, thus a virtual grinding environment is built, as shown in Fig.3. The tool grinder sub-assemblies (see Fig.3 (a)) which are related to the grinding kinematics are digitized, and the digitized tool is depicted in Fig.3 (b).

(a)   (b)
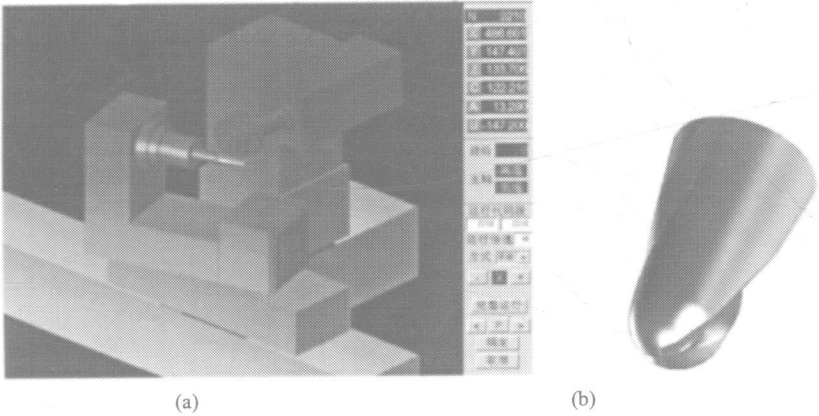
Figure 3. Virtual tool grinding environment

It should be pointed out that the VPIGC involves almost all of rotating tool grinding operations, the tool to be ground is easily digitized into virtual environment. The grinding wheel selected is also transferred into digital format, under 3D simulation environment, where the corresponding virtual commissioning will be carried out. There, of course, are a lot of software developments and theoretical analyses, especially, physical simulation to be made. Fortunately, the final tool grinding process will not produce great force induced and heat induced errors. The experimental results show that what obtained in VPIGC is just as that expected in virtual environment, as shown in Fig.2 and Fig.3.

2.2 Case Study II: Virtual/Practical Integrated SLS (Selective Laser Sintering)

SLS process is one of layered manufacturing processes, which may be used to directly produce real industrial products. Thus it is more important than others. The main factors influencing on SLS process properties include laser power, laser scanning feed rate, material composition to be sintered, and the scanning path. Since SLS process is a complicated process, it is hard to be accurately described by means of a mathematical model. According to the machining mode of VPIMPC, a Virtual/Practical Integrated SLS (VPI-SLS) may be built. Recently, we finished SLS process investigation by adopting fractal scanning path including self-developed experimental scanning device driven by steel wire rope, fractal path generation and filling of arbitrary regions, direct slicing from CAD model, 3D finite analysis on temperature field and residual stress field, and carrying out a great amount of experiments. The basic structure of the practical environment and the virtual environment are created respectively and then integrated with each other. Fig.4 indicates the material processing cell in PE [Fig.4 (a)] and VE [Fig.4 (b)]. As the SLS process is based on 2D production principles, the PE of material processing cell is transferred into VE digitally only concerning the scanning path of the laser beam.

(a)   (b)
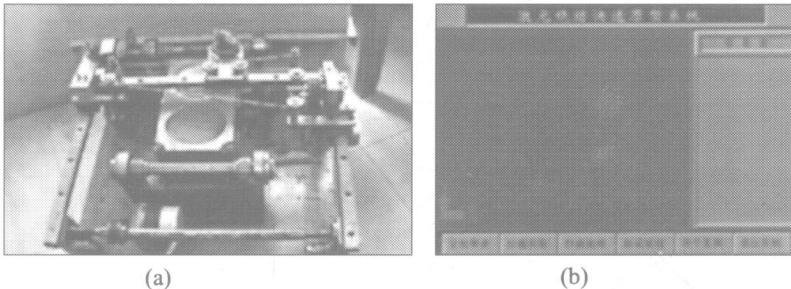
Figure 4. (a) the VPI–SLS device (b) the VPI–SLS virtual demonstration

Fig.5 shows the components to be processed in PE and VE respectively. In the VE, there exists a CAD solid model, direct slicing, filling desired area with scanning path. All of these mentioned above is limited in geometrical domain, In order to complete the virtual commissioning of SLS process, it is expected to analyze the physical performance of the process, Following research results obtained by the research group may help to deal with the problems mentioned.

(a)   (b)   (c)   (d)
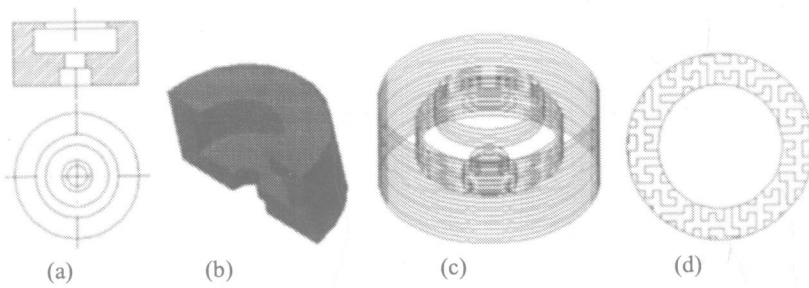
Figure 5. Components to be processed in PE and VE (a) 2D-drawing; (b) 3D model; (c) slicing; (d) filling of arbitrary layer

- Based on 3D finite element analysis, the temperature field and the residual stress field are calculated and displayed respectively, the results of which may be applied to determine selection of the scanning path, say, fractal curve, "S" line scanning, circular scanning etc..
- On the basis of error mapping principle, the width of sintered lines measured by a self-developed CCD-sensor-based measuring device corresponds to the laser power and feed rate of the beam scanning. ANN-based control algorithm is developed based on experimental results, so as to obtain an optimal combination of processing parameters.
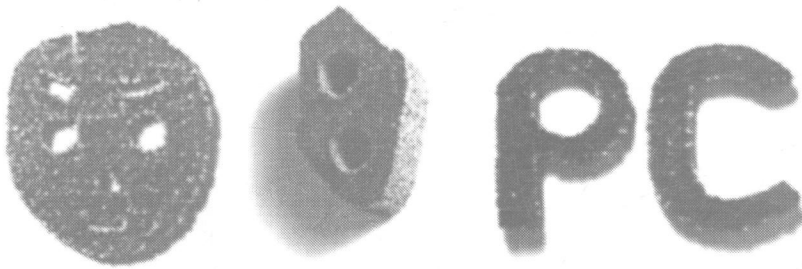
Figure 6. Sintered samples using the fractal scanning path

Many sintered samples shown in Fig.6 are offered to indicate the VPI-SLS feasibility.

3 CONCLUSIONS

The VPIMPC strategy is supported by two case studies, and proven to be successful and powerful for shop floor usage. The machining operations are in the foundation and the ground level of industrial production system. For existing CNC machines, VPIMPC may fully adopt the machine information and resources.

ACKNOWLEDGEMENTS

REFERENCES

(1) **Ravi Chawla, Amamath Banerjee.** (2001) A virtual environment for simulating manufacturing operation in 3D. Proceedings of the Winter Simulation Conference, pp. 991-997.

(2) **Gary A. Mintchell.** (2001) Model, Simulate, Execute Simulation in Discrete Control Control Engineering, April.

(3) Verification Update, Winter 2000. The virtues of Virtual Manufacturing.

(4) http://sst.tees.ac uk/external/u0000981/CRAFT/FPV, Virtual Environment for Manufacturing.

(5) **Xiong Feng, Bin Hongzan, Jin Jianxin.** (2001) 3D simulation of CNC grinding for complex shaped tools, China Mechanical Engineering, 12(12), pp. 1253-1255. (In Chinese)

(6) **Liu Zhengyu, Bin Hongzan, Zhang Xiaobo.** (1998) The influence of the fractal scanning path on the temperature field of the layer in material increase manufacturing. Journal of Huazhong University of Science and Technology, 26(8), pp. 32-34. (In Chinese)

(7) **Liu Zhengyu, Bin Hongzan, Zhang Xiaobo.** (1999) The influence of the fractal scanning path on the residual stress field of the layer in material increase manufacturing. China Mechanical Engineering, 10(8), pp. 848-850. (In Chinese)

(8) **J Yang, H Bin, W Liang.** (2001) A New Direct Slicing Strategy Based on B–rep In: IMechE Conference Transactions of Computer-aided Production Engineering, Wuhan, China, pp. 251-254

Research on the process model of product development with uncertainty based on activity overlapping

R XIAO and **S SI**
CAD Center, Huazhong University of Science and Technology, Wuhan, People's Republic of China

ABSTRACT

The essence of competition among modern enterprises is time; hence the time for product development must be greatly reduced. This paper presents a new process model of product development with uncertainty based on activity overlapping, borrowing ideas from the uncertainty model and activity overlapping model proposed by Loch and Krishnan respectively, to realize reduction on the time. Besides the fundamental framework of the proposed model, some derivations on the formula computing the total execution time of upstream and downstream activities are made. The effectiveness of the proposed model is verified through some further discussions and initial computational results.

1 INTRODUCTION

Product development is presently faced with most challenging environments due to fierce competition on the market. At the initial stage of industrialization, competitiveness mainly lies on the price of products. Only if the products were cheap and usable, would they be of competitive advantage in the market. Since the price of products is determined by the cost, this type of competition is called cost-based competition. With the improvement of people's livings, the quality as well as service turned to be the trumps, which lead the competition to be a quality-based one. Since 1980s, uncertainty of the business circumstance has increased and the enterprise competitions tend to be more drastically, when competitiveness lies greatly on the time as well as variety of products. Only those who respond to the market changes rapidly would occupy larger market share. Thus the pattern of competition turned to be time-based one.

To succeed in the time-based competition, it is necessary to shorten the time of product development greatly; hence, some new concepts, theories and technologies relevant to the issue emerged, like concurrent engineering and time compression. Regarding the product development process, this paper proposed a new time compression method via modelling approach, which is based on the idea of concurrent engineering, conforms to the current markets with uncertainty and the demands of the quick response to market changes for modern enterprises, accordingly, the work in this paper can be of effective support to the management of complicated processes of product development.

2 ANALYSIS OF ACTIVITIES IN PRODUCT DEVELOPMENT

2.1 Description of information relations in product development

There have come forth many models describing the process of product development, where there is a lot of species based on the Design Structure Matrix (DSM)(1). The concept of DSM is originated from the systemic design (2), and it can also be used to analyze the process of product development that has wider overlaying. In DSM, each row and its corresponding column are identified with one of the tasks or activities. Along each row, the marks indicate other activities that is dependent to fulfill the activity represented by this row, and what certain column indicates is the other activities that receive information from the activities represented by this column (3). Diagonal elements do not convey any meaning at this point, since an activity cannot depend on itself. Thus, DSM can be utilized to describe the information relations in product development activities.

Viewed from the way of information connection, there are three forms among activities in product development described by DSM, viz., serial, parallel and couple (4), as shown in figure 1. Of which, there are simple one-way connections among the serially dependent activities, which can be easily decomposed, however, it will result in time delay of product development. In order to shorten the total time for the execution of activities, the preliminary information of the activity being executed can be utilized to implement its partial overlapping with the coming up activities. The former is usually called upstream activity, and the latter as downstream activity. For relatively pure connection form of serial dependence, the result of time compression of product development can be obtained when taking proper strategies and through overlapping of upstream and downstream activities, which truly embodies the idea of the concurrent engineering and the research in this paper belong to this class.

| (a) Serial | (b) Parallel | (c) Couple |
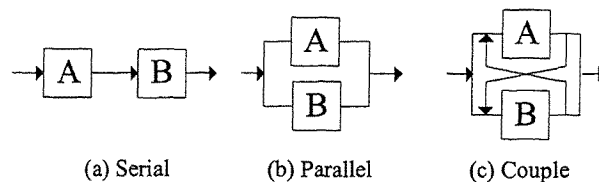
Fig.1 Three connection forms among activities

2.2 Review and analysis of the research on activity overlapping

2.2.1 The model of activity overlapping

The fundamental research framework on activity overlapping is given by Krishnan et al (5,6,7). Krishnan proposed four typical patterns when discussing information relations among the upstream and downstream activities. Viewed qualitatively, the information evolution speed of upstream activities differs greatly and the sensitivity degree of downstream ones is also not the same. Through combination of the two facts, four typical patterns form, where one kind of pattern has low sensitivity degree of downstream activity and slow speed of information evolution of upstream one. The preliminary information of upstream can be utilized early to trigger downstream activities. However, the information evolution of upstream activity is slow, leading to the fact that downstream activities must be iterated to be convergent to the final results. This pattern is emphasized by Krishnan and named Iterative Overlapping Problem (IOP). In the IOP model, t_i is the start time of the ith iteration, $T_{sum} = t_n + d_n$ is the total execution time of upstream and downstream activities. The number of information exchanges is $n+1$, which are finished instantaneously, and the following expressions are true:

$$t_i \geq t_{i-1} + d_{i-1} \tag{1}$$

$$t_0 \geq t_{As} \tag{2}$$

$$t_{n-1} < t_{Af} \tag{3}$$

$$t_n \geq t_{Af} \tag{4}$$

2.2.2 The consideration of uncertainty

Based on Krishnan's framework, Loch argued that the information of the upstream activity is variable, hence the arrival of the exchanged information of the upstream activity A can be supposed to be a non-homogeneous Poisson stochastic process, the rate of which is $\mu(t_A)$. Loch proposed the relevant mathematical expression (8) in a further step. On the other hand, after the arrival of the exchanged information of the upstream activity, the downstream activity will respond on the exchanged information through certain rework. Since there exits time consummation of information commutation between the upstream and downstream activities, Loch thought that it is unsuitable to trigger information commutation due to the downstream reaction on the arrival of the commute information of the upstream activity at any moment, whereas it is better to adopt appropriate batching police. However, Loch only considered the one shot batching instance (8), which does not consist with the real facts.

2.2.3 Analysis

From the introduction of sections 2.2.1 and 2.2.2, we know both the merits and drawbacks of Krishnan and Loch's works respectively. The former considers the repetitiveness and instantaneousness of communication that consequently triggers the iteration of the downstream activity, while it disregards information uncertainty. The latter considers the

uncertainty of information arrival from upstream activity, which satisfies the requirement of complicated product development presently with uncertainties, and is therefore much closer to the reality. However, it only considers the information commutation of upstream and downstream activity for once, which is accumulative due to batch processing operation, hence the iteration of downstream activity is only one time. Viewed from this point, the multi-iterative model proposed by Krishnan is, no doubt, more reasonable. It is obvious that the combination of the two complementary models has more advantages and is of more practical significance. This is the motivation to create models corresponding to the product development process with uncertainties based on activity overlapping.

3 THE UNCERTAINTY MODEL OF PRODUCT DEVELOPMENT PROCESS BASED ON ACTIVITY OVERLAPPING

3.1 The uncertainty model based on activity overlapping

3.1.1 Information arrival model of revised upstream activities

According to the above discussion, the arrival of the exchanged information of the upstream activity A can be supposed to be a non-homogeneous Poisson stochastic process whose intensity is $\mu(t_A)$. Loch (8) founded a model of the information arrival of such stochastic process adopting the linear function

$$\mu(t_A) = \mu[1 + e(2\frac{t_A}{T_A} - 1)] \tag{5}$$

where T_A is the executing time of the upstream activity and t_A is limited within $[0, T_A]$, e is a parameter reflecting the degree of the evolution of the upstream activity, where $e < 0$ corresponds to the fast evolution and $e > 0$ corresponds to the slow evolution; hence e is named as the evolution degree which is limited within $[-1, 1]$.

In formula (5), μ is the average Poisson intensity. Based on some experiential results, Loch thought that μ is determined by the following formula

$$\mu = \mu_0 \exp\{-B\alpha\} \tag{6}$$

where α is the total amount of information exchange before startup, parameter B represents the coordination capability of development group, μ_0 is the inherent technical uncertainty of the project.

3.1.2 The rework model of the downstream activity

The rework time of the downstream activity lies on the degree of the downstream activity reacting on the exchanged information and the evolutional status of the downstream activity. Therefore, Loch defines the influencing function $f(t)$ that denotes the rework time of the downstream activity. The more downstream has progressed in its work, the more cumulative work must be modified, hence $f(t)$ is non-decreasing function. In practice, $f(t)$ might be concave or convex. Adopting the linear function, Loch (8) gave the expression of $f(t)$ as follows in order to make it simple,

$$f(t) = kt \qquad (7)$$

where k is the sensitive degree of the downstream activity that is usually limited within [0.01, 0.09].

3.2 The model proposed in this paper

3.2.1 Fundamental framework

In the model proposed in this paper, the times for information exchange between upstream and downstream activities, viz. n+1 times iteration in all, consist with that of Krishnan's model, and since each iteration is accumulative, which is accordant to that of Loch's model, the iteration times of the downstream activity is $n+1$ too. A fragment of overlapping between upstream and downstream activities is given in Fig.2, which covers iteration for two times, where figure 2(a) denotes the physical structure reflecting real states and figure 2(b) gives the logical structure equivalent to the former one.
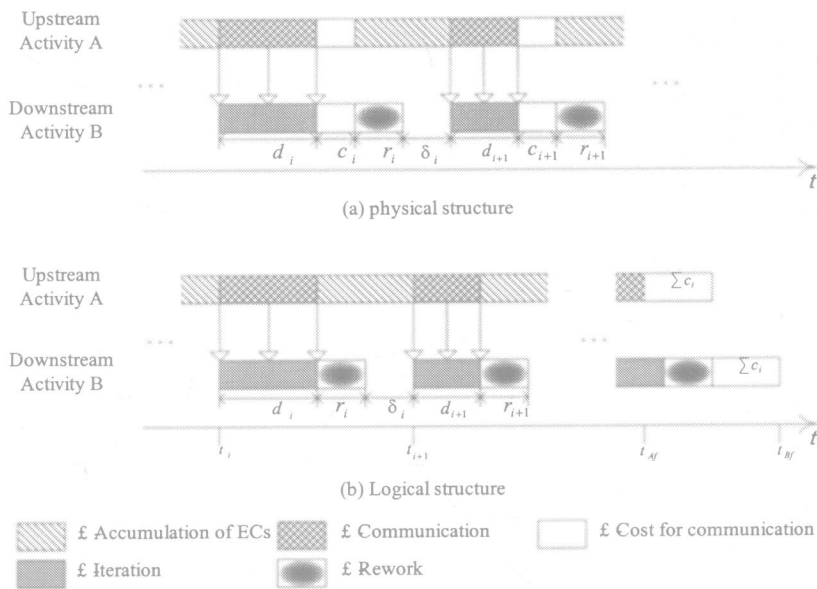
Fig. 2 Twice iterations in activity overlapping

It can be obtained from figure 2(a) that the ith iterative time of the downstream includes the following three parts: 1) the execution time d_i of iteration during the delivery period of upstream information, wherein d_0 corresponds to the planning iteration. 2) The cost of time c_i when exchanging information between upstream and downstream activities. 3) The rework time r_i after information exchange. The model proposed in this paper is more comprehensive

than Krishnan's model that considers only d_i, and Loch's model that only takes c_i and r_i into account.

To utilize Krishnan's model conveniently, activity A in figure 2 should be in the state of information processing all the time. Then the relevant T_A can not include c_i, viz., c_i exists physically instead of logically, which is not the same case for the downstream activity B. Supposed T_B be the executive time of activity B, T_B will include c_i, viz., c_i exists not only physically but also logically. All of the following discussion is based on the logical structure (In figure 2(b)). First, let t_{Bs}, t_{Bf} be the initial and terminative time of activity B respectively, T_0 be the overlapping time of activity A and B, λ be the overlapping rate, then we can get the following formulae,

$$T_B = t_{Bf} - t_{Bs} = t_{Bf} - t_0 \tag{8}$$

$$\lambda = \frac{T_o}{T_A} \tag{9}$$

$$t_0 = t_{Af} - T_o = T_A - \lambda T_A = (1 - \lambda)T_A \tag{10}$$

3.2.2   The amendment to Loch's model
3.2.2.1 Amendment to formula (5)

Since $T_A = t_{Af} - t_{As} = t_{Af}$, formula (5), alternatively, can be written as,

$$\mu(t_A) = \mu[1 + e(2\frac{t_A}{t_{Af}} - 1)] \tag{11}$$

where $0 \le t_A \le t_{Af}$. Supposed t_F to be the termination time of the upstream information, t_F is nearly correlative to the value of e. If upstream evolves slowly, the final form of exchanged information can no be obtained once the upstream activity is terminated, where $t_F = t_{Af}$ at that time. If upstream evolves fast, the exchanged information can get to the final form ahead where $t_F < t_{Af}$. There is no information arrival from t_F to t_{Af} for the latter instance which is not rational because $\mu(t_A)$ is not equal to 0 according to formula (5) or (11). Then there is the revise that substitute t_{Af} with t_A, therefore the model of the upstream information arrival that we adopt practically is the following,

$$\mu(t_A) = \mu[1 + e(2\frac{t_A}{t_F} - 1)] \tag{12}$$

3.2.2.2 Amendment to formula (6)

Our model has the repetitive iterations whereas Loch's model only considers only once iteration of the downstream activity. Therefore, formula (6) should be amended as follows,

$$\mu = \mu_0 \exp\{-B\alpha_i\} \tag{13}$$

where α_i is the total amount of the exchanged information when the ith iteration proceeds. $\alpha_0 = \alpha$, while α_i can be obtained from the following formula

$$\alpha_i = \alpha_{i-1} + \int_{t_{i-1}}^{t_{i-1}+d_{i-1}} \mu(t)\,dt \tag{14}$$

where $\int_{t_{i-1}}^{t_{i-1}+d_{i-1}} \mu(t)\,dt$ is the amount of revised information arrived at $(i-1)^{\text{th}}$ iteration from the upstream activity.

3.2.3 The derivation of T_{sum}

T_{sum} refers to the total project execution time including upstream and downstream activities. Considering the $(i+1)^{\text{th}}$ iteration of the overlapping of A and B that correspond to the ith iteration since i is calculated starting up from 0. According to reference (9),

$$d_i = d_0 \varphi_B^i \tag{15}$$

where φ_B^i is the decreasing ratio of the execution time at ith iteration of the downstream activity and it satisfies the following formula

$$\lim_{i \to \infty} \varphi_B^i = 0.5 \tag{16}$$

According to reference (8),

$$c_i = \int_{t_i}^{t_i+d_i} \sqrt{\frac{k\mu(t)\tau}{2}}\,dt \tag{17}$$

where τ is the coefficient of communication cost.

The rework at ith iteration is $\int_{t_i}^{t_i+d_i} k\mu(t)(t-t_i)\,dt$ disregarding communication cost and

the additional rework at ith iteration is $\int_{t_i}^{t_i+d_i} \left(\sqrt{\dfrac{k\mu(t)\tau}{2}} - \dfrac{k}{2} \right) dt$ caused by time cost

when exchanging information, thus the sum of the above two is

$$r_i = \int_{t_i}^{t_i+d_i} \left(k\mu(t)(t-t_i) + \sqrt{\frac{k\mu(t)\tau}{2}} - \frac{k}{2} \right) dt \tag{18}$$

Let δ_i be the time interval between i^{th} and $(i+1)^{\text{th}}$ iterations of downstream activity, then

$$t_{i+1} = t_i + d_i + r_i + \delta_i \tag{19}$$

When the upstream activity approaches to the end, there will be two cases:

1) The last execution of downstream iteration, started before t_F is finished before t_F, whose

rework of the iteration is ended at t_F or later. When there still exists certain accumulative information of the upstream activity, which has not been passed to the downstream activity. Therefore, a downstream iteration should be triggered, where there are only iterative execution and no rework. Thus t_{Bf}, the end time for downstream activity, can be given as

$$t_{Bf} = t_n + d_n + \sum_{i=0}^{n-1} c_i \qquad (20)$$

2) The last downstream iterative execution starting before t_F is accomplished at t_F or after that. Under such a situation, all of the upstream information has been transferred. So the new iteration is not required to trigger and t_{Bf} can be calculated by the following formula

$$t_{Bf} = t_n + d_n + r_n + \sum_{i=0}^{n} c_i \qquad (21)$$

where $r_n = \int_{t_n}^{t_F} \left(k\mu(t)t + \sqrt{\dfrac{k\mu(t)\tau}{2}} - \dfrac{k}{2} \right) dt$, $c_n = \int_{t_n}^{t_F} \sqrt{\dfrac{k\mu(t)\tau}{2}} dt$.

So far, the calculation formula of T_{sum} can be given as follows.

For the first case,

$$T_{sum} = \max\{T_A + \sum_{i=0}^{n-1} c_i, t_{Bf}\} \qquad (22)$$

For the second case,

$$T_{sum} = \max\{T_A + \sum_{i=0}^{n} c_i, t_{Bf}\} \qquad (23)$$

4 DISSCUSION

4.1 The expression of t_F

According to the above discussion, t_F is close relevant to the evolution degree e. Obviously, $t_F = t_{Af}$ when the evolution is the slowest that $e = 1$ and $t_F < t_{Af}$ when the evolution is fast that $e < 0$. Here r_A, an information terminative coefficient of the upstream A, is introduced as follows which denotes the advance ratio of the upstream information termination when $e = 0$.

$$r_A = \frac{t_{Af} - t_F}{t_{Af}} \qquad (24)$$

From the formula (24), we can obtain that,

$$t_{Af} - t_F = r_A t_{Af} \qquad (25)$$

$t_{Af} - t_F = 0$ when $e = 1$. Moreover, smaller e indicates faster evolution, thus smaller t_F and then larger $t_{Af} - t_F$. So $t_{Af} - t_F$ is the decreasing function of e. Based on the above points, a

simple expression of $t_{Af} - t_F$ can be constructed as follows.

$$t_{Af} - t_F = (1-e)r_A t_{Af} \qquad (26)$$

From the formula (26), t_F can be yielded as

$$t_F = [1-(1-e)r_A]t_{Af} \qquad (27)$$

4.2 The preliminary calculation results and analysis

The essential purpose of model construction in section 3 is to calculate T_{sum}, so as to quantify the effect of time compression. To make T_{sum} non-dimensional, the coefficient of time compression ρ is introduced and defined as follows.

$$\rho = \frac{T_{sum}}{T_A + d_0} \qquad (28)$$

According to the proposed model in this paper, the relation between time consumption of the information exchange and ρ is shown in figure 3 and the relation between the times of the advance information commutation and ρ is shown in figure 4, which results from the calculation. The parameters adopted in the calculation are: $T_A = 20$ weeks, $d_0 = 5$ weeks, $e = 1$, $k = 0.01$, $\lambda = 0.94$, $n = 4$, $r_A = 0.01$, $\mu_0 = 1$, $B = 0.1$ and $\alpha = 0$ in figure 3, $\tau = 3$ days in figure 4.
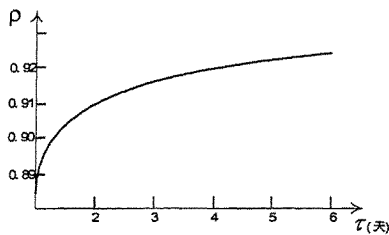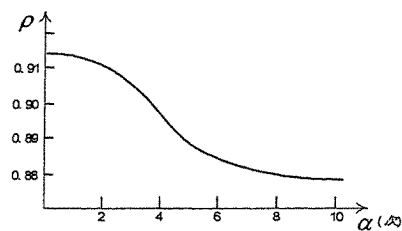
Figure 3   **Figure 4**

In figure 3, the total time to execute upstream and downstream activities may increase with the augment of communication costs. It can be drawn from figure 4 that the total executing time may decrease correspondingly with the increase of time for the advance information commutation before the project starts. Such results accord with our intuition. Both the results in figure 3 and figure 4 satisfy $\rho < 1$, which shows that the effect of time compression is achieved by utilizing the model in this paper.

5 CONCLUSIONS

In view of the complicate and non-deterministic environments of current markets and the

requirement of rapid response of modern enterprise to market changes, a kind of process model of product development with uncertainty based on activity overlapping is proposed in this paper, adopting the idea of concurrent engineering and taking the work of Krishnan and Loch as reference. The proposed model integrates the merits of both Krishnan's and Loch's, and the influence of many factors is generally considered, which make it closer to the actual process of the complex product development. The effectiveness of the proposed model is verified through preliminary calculation and analysis. The work of this paper is of important meaning to mastering in principle the inherent regularities of complex product development process and for effective compression of development time.

ACKNOWLEDGEMENTS

REFERENCE

1 **Smith R. P., Morrow J. A.** Product development process modeling. Design studies, 1999, 20(3): 237-261
2 **Steward D. V.** The design structure system: a method for managing the design of complex systems. IEEE Trans. on Engineering Management, 1981, EM-28(3): 71-74
3 **Ulrich K. T., Eppinger S. D.** Product design and development. Boston: McGraw-Hill, 2000
4 **Eppinger S. D.** A model-based method for organizing tasks in product development. Research in Engineering Design, 1994, 6(1): 1-13
5 **Krishnan V., Eppinger S. D, Whitney D. E.** A model-based framework to overlap product development activities. Management Science, 1997, 43(4): 437-551
6 **Krishnan V., Eppinger S. D., Whitney D. E.** Accelerating product development by the exchange of preliminary product information. ASME J. of Mechanical Design 1995, 117(4): 491-498
7 **Krishnan V.** Managing the simultaneous execution of coupled phases in circumvent product development. In 1996 IEEE Trans. on Engineering Management, EM-43(2): 210-217
8 **Loch C. H., Terwiesch C.** Communication and uncertainty in concurrent engineering. Management Science, 1998, 44(8): 1032-1048
9 **Cho S. H, Eppinger S. D.** Product development process modeling using advanced simulation. Proceedings of 2001 ASME Design Engineering Technical Conference and Computers and Information in Engineering Conference, DETC2001/DTM-21691.

A purchasing policy model based on components/parts unification

X SUN and **D MAN**
School of Economy and Management, Southeast University, Nanjing, People's Republic of China
D ZHONG
School of Economy and Management, Southeast University, Nanjing, People's Republic of China and Zhuzhou Institute of Technology, Zhuzhou, People's Republic of China

SYNOPSIS

This paper presents a mathematical model for components/parts unification (CPU) policy. This model considers two components/parts that are functional interchangeable but purchased from suppliers with differing prices and quality characteristics. Because of buyer's quality preference and suppliers' discount rates for bulky purchases, the model assist the procurement manager to determine how best to purchase the components/parts to meet its demand while minimize the total acquisition costs.

Keywords: *Components/Parts Unification, Purchasing, Parametric Example.*

1. INTRODUCTION

Mass Customization is an effective weapon for fighting the fierce competition in the global market where customer requirements are becoming increasingly diversified and individualized (Bovet and Martha, 2000). The enormous challenge in mass customization is to resolve the conflict between the product diversification, production cost and response speed (Anderson and Joseph Pine II, 1997). One of the effective tactics is to maximize the product variety without increasing the number of components and parts dramatically. By so doing, the enterprises are able to respond to the market demands fast with the right products and services of high quality and low costs. However, this is easy to say but difficult to implement. Even large reputable corporations have experienced difficulties in controlling the part varieties. For example, the company of the Nissan found in 1993, in the style that only produced at that time, use 110 kinds of different radiators, 300 kinds of different ashtrays, 437 kinds of gauge boards, the fasteners of 1200 kinds of inland blankets different from 6000 kinds (Anderson, 1997).

The kinds of components/parts increasing lead to the economic benefits reducing, which is not paid attention. The phenomenon of increased part varieties causing reduced performance is at first evidenced at the products design stage. The explosion of part variety in the product design leads to the explosion at the procurement stage. Purchasing engineers put undue emphasis on the price of purchase of the components/parts (the direct cost), ignoring the relevant cost that the components/parts is obtained and takes place in the course (the overhead costs). In fact, the proportion that the overhead costs of the parts is relatively great, depending on the demand. Pareto's law applied to inflexible plants would say that 80 percent of the material overhead

costs would be consumed on low-usage parts that may represent only 20 percent of total part volume (Anderson and Joseph Pine II, 1997). However, because overhead costs are often difficult to quantify accurately, the current practice of estimating overhead costs does not reflect the contribution of the cost involved in managing the part variety. Therefore, there is no management control over the ever increasing part variety.

Now the challenge is how can components/parts be unified to reduce the part variety during the procurement and under what condition such components/parts unification (CPU) should be best adopted? These problems have not been studied carefully in theory and practice. Only some preliminary qualitative analysis findings were brought forward (Bovet and Martha, 2000 and Anderson and Joseph Pine II, 1997). Only when the tendency of buyer reducing supplier quantity has been studied since 1990s, components/parts standardization was mentioned that follows above phenomenon in some literature (Trent, 1998 and Bakos and Brynjolfsson, 1993). The components/parts standardization was emphasized only as a kind of means or precondition that enterprises implement the strategy of customization manufacturing in other literature (Kolisch, 2000). . All these have not solved the problem of **CPU**.

In this paper, a method will be proposed to minimize total acquisition cost (**TAC**) by means of components/parts unification (**CPU**). The influence factors to **CPU** are discussed via qualitative analysis. The condition that the similar components/parts of two types of different prices are unified into one type of components/parts is confirmed via quantitative analysis. According to the analytical findings, some concrete suggestions to the management of **CPU** are suggested.

2. TOTAL ACQUISITION COST MODEL OF THE COMPONENTS/PARTS

For structure total acquisition cost (**TAC**) model for similar two different types of components/parts, we suppose as follows at first.

(1) There are two different types of same kindred components/parts that are u_1 and u_2 They are functionally interchangeable but purchased from suppliers with differing prices and quality characteristics. u_2 can be replaced by u_1 in terms of the design and quality requirement. The unit purchasing cost/price of u_1 is higher than that of u_2 (otherwise u_2 has no reason to be purchased).

(2) Annual demand quantity of u_1 and u_2 is stable.

(3) $\mathbf{TAC}_1$ and $\mathbf{TAC}_2$ stand for the total acquisition cost when components/parts u_1 and u_2 are purchased separately. **TAC** shows the acquisition cost after u_1 and u_2 have been unified into u_1 (For short: u_1 and u_2 unification).

(4) D_1 and D_2 stand for annual demand quantity of u_1 and u_2 separately. D stands for the total annual demand quantity of u_1 and u_2 unification. So, $D = D_1 + D_2$. Order: $D_2 = px$ D_1, therefore, $D = (1+p) x D_1$

(5) m_1 and m_2 stand for the unit price of u_1 and u_2 separately, and m stands for the unit price after u_1 and u_2 unification. Order: $m_2 = axm_1$, $m = bxm_1$, $(a, b \leqslant 1)$, where, "$a \leqslant 1$" stands for that the unit price of u_2 is lower than the unit price of u_1, and "$b \leqslant 1$" stands for that certain purchasing price discount (the preferential price) can be obtained owing to the increasing of purchased quantity by means of parts unification.

(6) Separately, k_1 and k_2 stand for the order cost (including production arranged cost) of u_1 and u_2 at a time. k stands for the order cost (including production arranged cost) after u_1 and u_2 unification at a time.

(7) Separately, h_1 and h_2 stand for the average unit stock cost of u_1 and u_2 per year, and h stands for the average unit stock cost after **CPU** per year. In addition, $h_i = r_i \times m_i$ and $h = r \times m$, where, r_i and r stand for the corresponding unit price percentages of the stock cost $(i=1, 2)$.

(8) The overhead cost (including order cost and stock cost) of two types of components/parts is shared according to the different unit rate separately. The unit rate is related to the different demand quantity. The smaller the demand quantity is, the greater unit rate is. So, we can set up,

$k = g (1+p) \times k_1$, $k_2 = g (p) \times k_1$, $g (p)$ is a strictly monotony declining function of "p"

$r = f (1+p) \times r_1$, $r_2 = f (p) \times r_1$, $f (p)$ is a strictly monotony declining function of "p" （1）

(9) The order point method and safe stock system are be adopted to managing u_1 and u_2 (Brosocs and Kelose, 1999, Chase and Aquilano, 1998). Namely, one specific order point "R_i" is stipulated for u_i. Components/parts should be ordered when the stock level of u_i is up to R_i, and order batch is Q_i $(i =1,2)$.

(10) The safe stocks of u_1 and u_2 separately are S_1 and S_2 The safe stock after **CPU** is S.

Based on supposition mentioned above, the **TAC** model of u_1 and u_2 can be structured while u_1 and u_2 are purchased separately.

$$TAC_i = m_i \times D_i + \frac{D_i}{Q_i} \times k_i + \left(\frac{Q_i}{2} + S_i \right) \times h_i , \quad i=1, 2 \qquad （2）$$

In formula (2), item 1 "$m_i \times D_i$" represents annual purchasing cost of u_i. Item 2 "$\frac{D_i}{Q_i} \times k_i$" represents annual order cost (including cost of order form and production arranged cost) of u_i. Item 3 "$\left(\frac{Q_i}{2} + S_i \right) \times h_i$" represents the annual stock cost (including the relevant management cost proportioned). Meanwhile, each item above all indicates the status of purchasing separate for u_1 and u_2

The **TAC** after **CPU** was represented as follows.

$$TAC = m \times D + \frac{D}{Q} \times k + \left(\frac{Q}{2} + S \right) \times h \qquad （3）$$

The meaning of each item on the right in formula (3) is similar to that of formula (2).

3. ANALYSIS OF CPU CONDITION

Whether or not the components/parts u_1 and u_2 should be unified can be established by testing the following formula:

$$\min \textbf{TAC} \leqslant \min \textbf{TAC}_1 + \min \textbf{TAC}_2 \qquad （4）$$

If formula (4) is valid, u_1 and u_2 should be unified into one type of component/part u_1.

Without considering the safe stock cost (viz. S_1, S_2, $S=0$), the validity of formula (4) is analysed below.

Proposition: If the unit purchasing price of u_2 (m_2) is not lower by v times than the unit purchasing price of u_1 (m_1), these two types of components/parts should be unified into one type of component/part u_1. And the **TAC** will be reduced by means of **CPU** (keeping the same only when $m_2 = v \times m_1$). Among them:

$$v = \frac{\left[-\sqrt{\psi(p)} + \sqrt{\psi(p) + 4\sqrt{w}\left(\sqrt{w}\times(b-1+p) - \sqrt{1-p} + \sqrt{b\times\psi(1+p)}\right)} \right]^2}{4w\times p} \tag{5}$$

Where

$$\psi(\bullet) \; \square \; f(\bullet) \times g(\bullet)$$

$$w \; \square \; \frac{D\times m_1}{2r_1 \times k_1} \;, \quad \text{which is called "Form Parameter" of \textbf{TAC} in \textbf{CPU}.}$$

Proof. Insert $S_i = 0$ into (2) and evaluate Q_i's derivative:

$$\frac{dTAC}{dQ_i} = -k_i \frac{D_i}{Q_i^2} + \frac{h_i}{2} = 0 \;, \qquad i=1,\ 2$$

$$Q_i = \sqrt{\frac{2D_i \times k_i}{h_i}} \;, \qquad i=1,\ 2$$

Therefore, $\min TAC_i = D_i \times m_i + \sqrt{2D_i \times k_i \times h_i} \;, \quad i=1,\ 2$ \hfill (6)

Analogously, $\min TAC = D \times m + \sqrt{2D\times k \times h}$ \hfill (7)

Insert "$D_2 = p \times D_1$, $D = (1+p)\times D_1$, $m_2 = a \times m_1$, $m = b \times m_1$, $k_2 = f(p)\times k_1$, $r_2 = g(p)\times r_1$, $k = f(1+p)\times k_1$, $r = g(1+p)\times r_1$" and formula (6) and (7) into formula (4). We can educe:

$$b\times D \times m_1 + \sqrt{2D \times m_1 \times k_1 \times r_1 \times b \times f(1+p)\times g(1+p)}$$
$$\leq (1-p+a\times p)\times D\times m_1 + \sqrt{2D\times m_1 \times k_1 \times r_1}\left(\sqrt{1-p} + \sqrt{a\times p \times f(p)\times g(p)}\right)$$
\hfill (8)

Define: $w \; \square \; \dfrac{D\times m_1}{2r_1 \times k_1}$. Then (8) can be changed to (9).

$$\sqrt{w}\times p \times a + \sqrt{a\times p \times f(p)\times g(p)} - \left(\sqrt{b\times f(1+p)\times g(1+p)} + (b-1+p)\sqrt{w} - \sqrt{1-p}\right) \geq 0$$
\hfill (9)

Define: $\psi(\bullet) \; \square \; f(\bullet)\times g(\bullet)$. Then we predigest (9) and get (10).

$$a \geq \frac{\left[-\sqrt{\psi(p)} + \sqrt{\psi(p) + 4\sqrt{w}\left(\sqrt{w}\times(b-1+p) - \sqrt{1-p} + \sqrt{b\times\psi(1+p)}\right)} \right]^2}{4w\times p} \tag{10}$$

Define the right function of (10) as $v(w, \psi, p, b)$, then $a \geq v$.

$$\because m_2/m_1 = a$$

$$\therefore m_2/m_1 \geq v$$

Therefore, if the unit purchasing price of u_2 (m_2) is not lower v times than the unit purchasing price of u_1 (m_1), u_2 should be replaced by u_1 (unification of u_1 and u_2).

In addition, because $(m_1 - m_2)/m_1 = 1 - a$, the condition of **CPU** is also expressed that the unit price difference rate of u_2 relative to u_1 cannot exceed $1 - v$. Shortly as follows, the permission scale of relative price difference rate is $1 - v$ if u_1 and u_2 should be unified.

4. ANALYSIS OF INFLUENCE FACTOR ON CPU

From (5), the value of v is confirmed by the value of w, ψ, p and b. In order to analyse the change laws of the value of v following w, ψ, p and b further, we can know the influence which various kinds of factors produced on the condition of **CPU**. Parametric example method is adopted to analyse the influence factors of **CPU** as follows.

In order to analyse above problem further, we assume the value of f (p) and g (p) first. We know from (1), f (p) and g (p) are the strictly monotony declining function of "p". f (p) and g (p) are larger, which separate represent, the rate of purchasing cost and stock cost in of u_2s **TAC** structure is heavier than u_1 's. Through real investigation and analysis, we know that *the percentage of stock cost* (t) *is widely different*. The scale of t can be from 9% to 50%, which is averaged to 25%(Brosocs and Kelose, 1999). Possibly, the scale of t can be from 8.5% to 45%, which is averaged to 20% (Anderson and Joseph Pine II, 1997). These data are statistic data from some enterprises. Otherwise, different kinds of components/parts have different t (relating to the quantity of demand). Anderson and Pine II (1997) have done qualitative analysis about this. g $(\cdot)$ has the similar characteristic with f $(\cdot)$. We can suppose as follows for consulting with t of enterprises.

$$f(p) = g(p) = 1 + \frac{t}{p} \; , \quad 0 < p < 1$$

$$f(p) = g(p)1, \qquad p \geq 1 \qquad\qquad\qquad (11)$$

Figure 1 demonstrates changes of f (p) when t was fetched different values (0, 0.03, 0.05, 0.08, 0.1). While $t = 0$, then f $(p) = 1$, which represents that overhead costs of u_1 and u_2 were proportioned equally according to the quantity of demand (traditional method of proportion). The bigger t is, the greater the change velocity of f (p) relatively is. That is to say, following the demand quantity lessening, the unit overhead cost proportion is larger, and it's change velocity will be larger following the change velocity of the demand quantity. The size of t is involved in production mode, management method and level of components/parts procurement, transport, and stock, etc. Generally, the more advanced the management method is, and the higher the management level is, then the smaller the value of t is. Contrarily, the greater the value of t is on the contrary.

The influence of other factors on relative price difference rate $1 - v$ permitted is analysed, under the price discount (viz. $b=1$) cannot be enjoyed by means of **CPU** of u_1 and u_2. See example 4.1 and example 4.2.

Example 4.1: Influence analysis of f (p) and g (p) changing on $1 - v$

Suppose $w = 1000$, separately insert the value of varies t (0、0.03、0.05、0.08、0.1) into f (p) and g (p), and discuss the influence of t changes on $1 - v$. Figure 2 has shown the change laws of $1 - v$ with the functional relation of p in the difference value of t. When w is steady, we can get the following findings.

Finding 1: The smaller p is, the greater the value of $1 - v$ is. Viz., the smaller the demand quantity of u_2 relative to u_1 is, the larger the unification permitted scale of t is. That is

to say, the components/parts of smaller demand quantity should be replaced by the components/parts of larger demand quantity so as to reducing the overhead cost, thus achieve the goal of economizing **TAC**.

Finding 2: The bigger t is, the smaller p is, and the faster the value of $1-v$ increasing speed is. That is to say, the permission scale of relative price difference rate is influenced by cost proportion method of components/parts. The permission scale of relative price difference rate which gets through traditional overhead costs proportion method, is smaller than the rate which gets through the overhead costs proportion method according to demand quantity which is closer to practical situation. However, even if the traditional overhead cost proportion method has already adopted, u_2 also should be replaced by u_1 (so long as relative price difference rate is not over 8%) when the demand quantity of u_2 is very small to u_1 (for example under 10%), Whether or when the percentage of overhead cost in **TAC** is less (for example, $w \leqslant 1000$). See Figure 2.
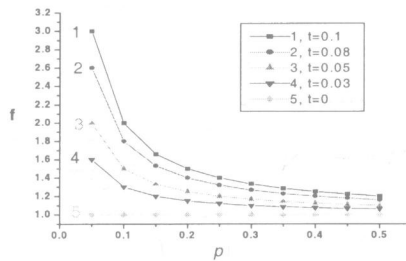
Figure 1 Various kinds of overhead cost proportion method changed with demand
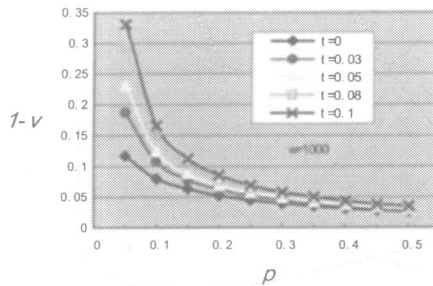
Figure 2 The influence of value of p on $1-v$

Example4.2 Influence analysis of w on $1-v$.

Figure 3 shows,

(1) No matter how to choose the value of t by changing f (p) or g (p), the value of w will have obvious influence on $1-v$.

(2) The smaller w is, the bigger $1-v$ is, that means relative price differential rate scale which unification permits is larger. Combining the definition of w, we can draw the following finding.

Finding 3: The smaller w is (viz. the total purchasing cost (Dxm_1) is smaller, also viz. the percentage of overhead cost in **TAC** (k_1xr_1) is bigger.), the larger the relative price differential rate scale which unification permits is. For example, when $w=100$, suppose the annul demand quantity of u_2 is the 30% of u_1, then u_2 also should be replaced by u_1 so long as the relative price difference rate is not over 13.3%(when f=g=1.1, viz. $t=0.03$) （See Figure3②）, or the relative price difference rate is not over 16.8% （when f=g=1.33, viz. $t=0.1$）(See Figure3④).

① ②

③ ④
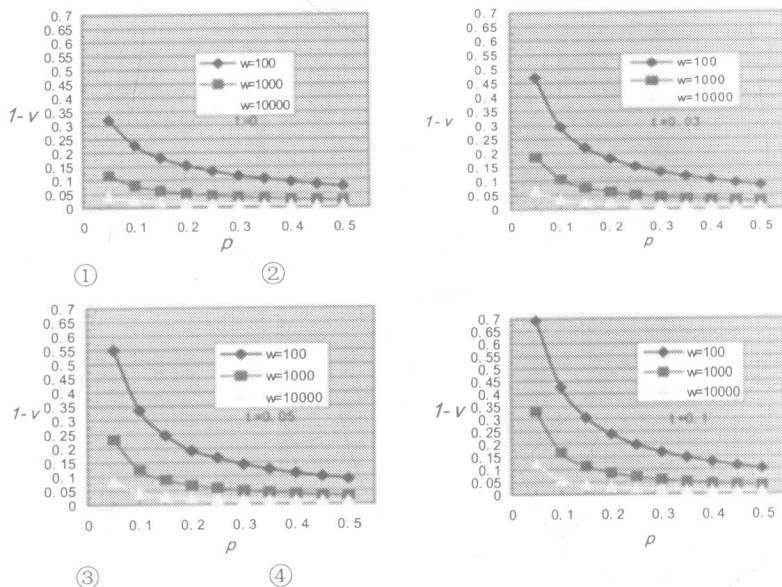
Figure 3 Influence of the **TAC** parameter w on $1-v$

Example 4.3 The influence analysis of p on $1-v$

After components/parts u_1 and u_2 are unified into one type of component/part u_1, the annual demand quantity of u_1 will increase. If u_1 is regarded as benchmark, the annual demand quantity of u_1 before **CPU** is D_1 and the unit-purchasing price is m_1. The annual demand quantity of u_1 after unification is $D=（1+p）x D_1$. With the demand quantity increasing, the amount of purchase each time will increase too. So certain price discount can be enjoyed. According to the assumption above, the unit purchasing price of u_1 after **CPU** is $m=bxm_1$. Then according to the analysed above, we can suppose, $b=1-zxp$.

Figure 4 shows that the influence of the purchasing price discount proportion $1-b$ after unification on the value of $1-v$ （$w=1000$, $t=0.05$）. z = 0 means the purchasing price discount by means of **CPU** cannot be enjoyed. z> 0 means the purchasing price discount by means of **CPU** can be enjoyed. With the demand quantity proportion "p" which u_2 relative to

u_1 increasing, the relative price difference degree which u_1 and u_2 should be unified is smaller and smaller. However, the reducing degree of the relative price difference while $z>0$ is much smaller than on $z=0$. The larger the discount rate of purchasing price enjoyed by means of **CPU** is, the smaller the relative price difference reduced is.

Therefore, findings can be drawn as follows.

Finding 4: When the purchasing price discount by means of **CPU** can be enjoyed, two types of components/parts of smaller relative demand difference can be unified. And by means of unification, the larger the discount rate of purchasing price enjoyed is, the bigger the rate scale of relative price difference $1-v$ is (See figure 4).

Finding 5: The rate scale of relative price difference $1-v$ that is permitted by CPU is far larger than the purchasing price discount $1-b$ that can be enjoyed by means of **CPU**. Namely, there is a magnified effect of $1-b$ to $1-v$.

For example, when $w=1000$, $t=0.05$ and $p=0.7$, if $z=0.05$, the relative price difference which can be enjoyed by means of **CPU** is less. Namely, if $1-v$ is not over 6.94%, u_2 should be replaced by u_1 and the purchasing price will drop 3.5%.
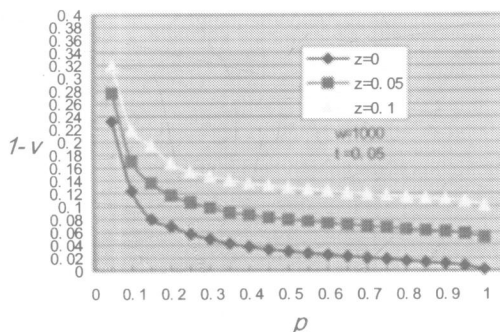
Figure 4 The influence of the value of b on $1-v$

5. MANAGERIAL IMPLICATION FOR CPU

Based on above analysis findings, reducing the types of components/parts by means of **CPU** (or called components/parts standardization) is not merely the urgent need of enterprises tackling challenge of the components/parts types increasing sharply, furthermore, from **TAC** view, not only some same kind of components/parts should be unified, but also the space of **CPU** (viz. the rate scale of relative price difference of components/part unit price) is quite big. Not only the similar components/parts of low value (viz. the unit purchasing price is lower, or viz. the annual total purchasing cost "Dxm" is lower) can be unified, but also the similar components/parts of high value (annual total purchasing cost "Dxm" higher) can be unified too, just slightly smaller the scale of relative price difference rate **CPU** permitted is (than low value components/parts).

From above analysis findings, we can know that the following factors may impact on **CPU** decision-making.

(1) Relative demand quantity of u_2 to u_1
(2) Relative unit purchasing price difference rate
(3) Total demand quantity of u_1 and u_2
(4) Proportion method to overhead cost (including purchasing cost and stock cost)
(5) **TAC** structure of u_1 (percentages of direct cost and overhead cost in **TAC**), Viz. the structure parameter of **TAC** "w"
(6) Purchasing price discount enjoyed by means of **CPU**.

So, we can propose suggestion to the management of enterprise **CPU** as follows.

(1) Enterprises can't only just consider the relative unit purchasing price difference. The more important thing is that other influence factors of **TAC** should be considered synthetically about the problem of **CPU**.
(2) Base on the premise of two components/parts be functional interchangeable, the two types of components/parts should be unified into one type of component/part if one of the following conditions is valid.

①The relative demand quantity rate of u_2 to u_1 is small (for example, under 15%), and the relative price difference rate of u_2 to u_1 is not too heavier (for example, not over 10%).

②The total demand quantity of u_1 and u_2 is much little. Meanwhile, the relative demand quantity rate is not too heavy (for example, not over 50%), and the relative price difference rate is also not too heavy (for example, not over 10%).

③Overhead cost (including purchasing cost and stock cost) is very heavier, or the percentage of overhead cost of u_1 in **TAC** structure is very big (viz. structure parameter w of **TAC** is very small, such as $w \leq 100$), and the relative price difference rate of u_2 to u_1 is not too large (for example, not over 10%).

④Certain purchasing price discount can be enjoyed by means of **CPU**, and the relative price difference rate of u_2 to u_1 is not too large (for example, not over 10%).

Certainly, the components/parts cannot be unified unrestrictedly. With the reducing of the components/parts types and the increasing of demand quantity, the relative price difference rate, which **CPU** permitted, will be smaller and smaller. Finally one best component/part type count will be achieved (when the relative price difference rate is down to zero).

6. CONCLUSION

This paper has discussed a mathematical model of components/parts unification to assist the procurement engineers to minimize the total acquisition cost. The model, however, is limited to consider only two components that are functionally interchangeable but with differing unit price and quality characteristics. The model can be extended for considering a more general situation where multiple functionally interchangeable components are involved. In addition, the safe stock level and suppliers' capacities have not been considered in the model. Extensions may also be made in the direction to include these factors in the model.

7. ACKNOWLEDGMENTS

8. REFERENCES

Bovet,D., Martha,J., 2000, Value Nets: Breaking the Supply Chain to Unlock Hidden Profits, John Wiley & Sons, Inc. (Chinese version)

Anderson, D.M., Joseph Pine II,B., Agile Product Development for Mass Customization: How to Develop and Products for Mass Customization, Niche Markets, JIT, Build-to-Order and Flexible Manufacturing, McGraw-Hill Companies, Inc. 1997.

Borsocs,D. J., Kelose,D.J., 1999, Material Flow—Integrated of Supply Chain Process, Peking: Mechanism Industry Press (Chinese version)

Chase,R.B., Aquilano,N.J., Jacobs,F.R., 1998, Production and Operations Management: Manufacturing and Services-8th ed, McGraw-Hill Companies, Inc.

Trent.,R.J., 1998, Purchasing and Supply Management: Trends and Changes Throughout the 1990s. , International Journal of Purchasing and Materials Management, Fall, pp.2-11.

Bakos,J.Y., Brynjolfsson,E., 1993, Information Technology, Incentives and the Optimal Number of Suppliers. Journal of Management Information Systems, Fall, pp.1-22.

Kolisch,R., 2000, Integration of Assembly and Fabrication for Make-to-Order Production. International Journal of Production Economics Vol.68, pp.287-306.

The cutting stock problem in make-to-order small/medium enterprises

F CONNOLLY
ManOPT Systems Limited, Limerick, Ireland
C SHEAHAN
Department of Operations Research, University of Limerick, Ireland

ABSTRACT

Increased costs in raw material along with greater competition and a more demanding consumer are leading industries to evaluate their manufacturing processes and practices in order to reduce raw material waste.

This paper addresses the issue of waste reduction in industry by presenting an integrated material optimisation system implemented and under continuing development in a SME with the combined aim of minimising waste while reducing lead times. It outlines the principles involved in the research and demonstrates the approach adopted in providing the operator with an efficient solution to the material optimisation issue in a holistic enterprise-wide manner.

The research is based on extensive experience in the component manufacturing industry and concepts developed in applied research undertaken in two make-to-order, facilities manufacturing precision-engineered work surfaces and panels.

The paper documents the issues faced in the design and implementation of an Integrated Material Optimisation and System for Make–to-Order Enterprises.

This research is an element of on-going research into the development of an Integrated Automatic Process Planning, Production Control, Scheduling and Waste minimisation system.

KEY WORDS:
Material Optimisation; Manufacturing Integration, Real-Time

INTRODUCTION

The Cutting Stock Problem (CSP) is one of the oldest problems in Operations Research. First discussed in the pioneering work of Paull (1956) and later by Gilmore and Gomory (1961), it is described as the search to find the best arrangement of shapes on rectangles to minimize waste or the number of rectangles. The majority of industrial packing tasks involve irregular shapes and regular stock material. In the leather industry the problem is a little more complex, where the raw material is also of irregular shape, and there are sections within the overall area outline that must be avoided because of defects.

Unlike the metals industry, where there is a better structure for waste reuse and recycling, the wood industry is faced with rising raw material costs. The industry also has to contend with more demanding customers, shortening lead-times, greater competition, and more complex product designs for an expanding international market.

The fundamental objective of industrial planning, in today's' manufacturing facilities, is to arrive at the best possible decision in a given set of circumstances for a given set of variables in the fastest possible time. Of course, many situations arise where the best is unattainable for a whole variety of reasons. The key objective though across the spectrum is to minimise costs and maximise profits. Since the effort required or the benefit desired in any practical situation can be expressed as a function of certain decision variables, optimisation has therefore been defined by Kerrigan (1988) as the process of finding the conditions that give the maximum or minimum value of a function.

CUTTING STOCK PROBLEM IN ACADEMIA

The research domain encompassing the CSP, the cutting and packing problem, has been very popular since the early 1960's there has been considerable interest and many publications. It is such that there have been a number of anthologies in an attempt to keep record recent developments in this area. There have also been attempts at normalising convention in this area in attempt to reduce ambiguity and misconception. The most popular literature review was developed by Dyckhoff and Finke (1992), who developed a classification method. They used this classification method to define clearly the differentiating factors between concrete and abstract packing problems.

Another attempt to address the many publications on the issue was conducted by Sweeney and Paternoster (1992). Unlike Dyckhoff and Finke (1992), however, Sweeney and Paternoster (1992) worked in reverse and evaluated the solutions returned from the differing problems posed. They grouped the problems into three categories based on the approached used to solve the problem. 1. Sequential assignment heuristics (a set of rules determining the order and the orientation of the items), 2. Single-pattern generating procedures (dynamic-programming based algorithms, which attempt to reapply a single 'optimal' pattern configuration) and 3. Multi-pattern generating procedures (linear programming based algorithms, which consider interactions between pattern; the solutions are approximated, hence heuristic). There have been many other works in the area, most notably, Golden (1976), Hinxman (1980), Rayward-Smith and Shing (1983), Sarin (1983), Coffman et al. (1984), Dowsland (1985), Coffman and Shor (1990), Haessler and Sweeney (1991), Dowsland (1991), Dowsland and Dowsland (1992), Whelan and Batchelor (1993), Dowsland and Dowsland (1995), Hopper and Turton (1997). Combinatorial problems, such as the cutting stock or packing problem can be stated as decision problems where a solution

corresponds to a correct yes or no response. An optimisation problem is converted by posing the question of whether of not there exists a feasible solution which has an objective function value equal or superior to a specified threshold (Garey and Johnson, 1979; Gary Parker, 1995). The rectangular packing problem or rather its decision analogue has been shown to be NP-complete (Fowler et al., 1981).

The guillotine cutting method has been approached by a number of researchers (Hwang et al. (1994), Kröger (1995), Rahmani and Ono (1995), András et al. (1996)). The guillotine cutting problem is a Rectangular Regular 2D problem, according to the Dyckhoff classification (Dyckhoff, 1990). Exact algorithms will determine problems of only limited complexity as they search for the exact solution. Heuristics are used to solve quite complex combinatorial problems. General heuristics methods like, genetic algorithms, simulated annealing and tabu search are used for many cutting problems.

ACADEMIA AND THE OPTIMAL SOLUTION

Academia has almost every case searched for the 'optimal' (greatest degree attained or attainable under implied or specified conditions) solution. This, the author argues, can be and is often, from the industries position, a pointless exercise (Figure 1). Industry demands an immediate solution, make-to-order especially, thrives on short lead times and accurate planning. In many instances the creation of an optimal solution will consume so much computing power and time that the time needed to produce a solution takes too long and the time/money lost waiting on the solution is more costly than it would be to produce whatever planned, manually. The 'Best' solution, in an industrial context, is an accurate near-optimal solution that is presented in the fastest possible time, with suitable consideration to all the relevant factors currently affecting the facility at that moment and for the duration of the planned production. Often, the best strategy is to propose suitable solutions in a decision support framework allowing the user to estimate the possible solution outcomes.
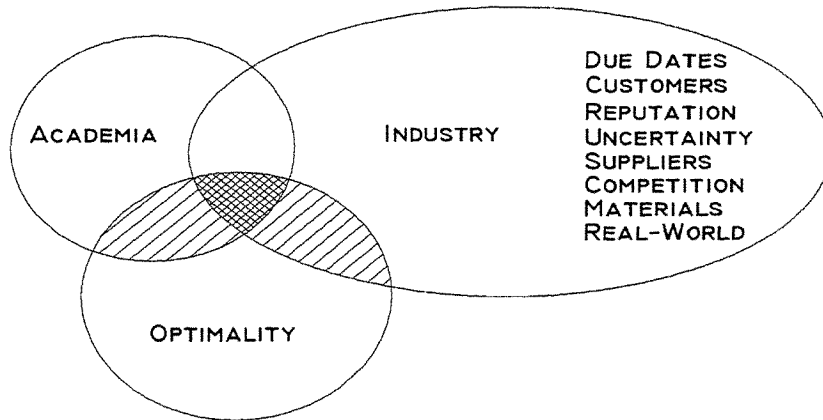
Figure 1

CUTTING STOCK PROBLEM IN INDUSTRY

The CSP in the wood industry is influenced by a number of key factors. The nature of the cutting operation is a major factor. The Cutting operation is influenced as much by the

material characteristics as the machining ability and technology available. The items and their characteristics affect the layout on the stock and these items are invariably complex and not regular in shape.

In both Case Study Firms, the primary raw material, or stock, is manufactured boards, such as Medium Density Fibreboard (MDF) and High Density Fibreboard (HDF). The material is compacted board of fine wood particles bound with a strong adhesive, generally with a fine laminated sheet of natural wood giving the board a decorative finish. The boards are a firm board purchased in regular sizes. While the finishes (the decorative laminated veneer) are similar across suppliers, they can be ordered from suppliers in a number of sizes. The boards can vary from 2655 x 2100 x 18 mm, 2620 x 2070 x 18 mm, 2440 x 1220 x 18 mm, 2440 x 1524 x 18 mm, etc. The boards cannot be bent or moved in any manner. Waste cannot be easily recycled or reprocessed for use unless it is of substantial size. One key factor with regard to raw wood materials, is the nature of the finish and its' grain direction or pattern. This closely linked to the product design and the grain direction required for the product design.

FACTORS AFFECTING THE CUTTING STOCK PROBLEM

Material Character:

The layout of the items ordered on the sheets is also influenced by a number of factors. Firstly, the grain direction of the products ordered must align with the grain direction of the stock it is to be made from, automatically determining the orientation of the piece. This is usually determined by the expected use of the piece. For example wood grain on a door generally runs vertically or parallel to the direction of the longest length, but on a drawer front this rule is the opposite with the grain direction running parallel to the shorter side (Figure 2). These generalisations can be over-ruled in exceptional cases.
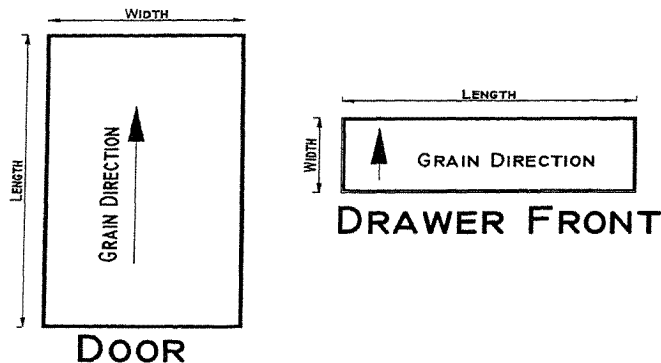
Figure 2

Cutting Technologies:

The machining techniques used to process the parts influence the layout rules and the order in which they are processed. In the first instance, the material type and characteristics as well as material size, decide the machinery available. Secondly the machine technology decides the cutting techniques available. With newer, but more expensive technologies, routing, stamping

and some laser technologies, it is possible to process very complex shapes and designs in one continuous cutting movement. With other older technologies, as is generally the case with wood technologies, the cutting technologies available only consider a guillotine cutting method. A guillotine method means that the cutting operation is:

1. A one-dimensional linear cutting motion operation.
2. The cutting action must start and finish at a piece edge, completely parting the material into two distinct pieces.

Obviously, the guillotine layout method is a more complex method of spatial layout than simple optimal layout, which must give consideration to the cutting procedures available. Another important factor in material cutting is what is referred to as the bridge width. This is the term given to the potential irregular edging feature found on material after a cutting action. For example, on metals a burr edge, caused by material slippage, often occurs after stamping action. To overcome this, a certain bordering margin is added to all pieces to prevent the irregularity affecting the actual finished pieces. One other factor often overlooked but pertinent, is the total distance the cutting tool must travel to complete the cutting action. This distance combined with the material concerned can affect accuracy, wear, tool life and safety and must be limited in some cases.

Manufacturing Techniques:

The manner in which pieces are produced or manufactured have a large influence on the cutting procedures and manufacturing techniques followed during processing. Generally pieces are grouped and entered into the material optimiser. The result influenced by the material type, cost of the products, due dates or any other business criteria. Scheduling of parts is one of the most important aspects of manufacturing. The sequence that the parts are cut from the stock directly affects the speed and order at which the items reach the next operation. The sequence of the parts may also be important for packaging and shipping. Geometrical or weight constraints may require the parts to be packed in a certain order and therefore the order in which they arrive can assist greatly in the speed of packing.

PROBLEM DEFINITION

It is the authors position that, certainly with regard to industrial concerns, the actual solution to a CSP is now almost of secondary interest. The prime concern for any manufacturing facility is that a suitable or good solution can be determined in a suitable time-frame. The actual method/heuristic/technique used to solve the Cutting Stock solution is not important. While the above statements are quite controversial from an academic point of view, the reality is that, no manufacturing facility is prepared to wait for an 'optimal' solution if they know that waiting will delay the product due-date or order-completion date. Tardiness has an economic value and in almost every instance this value is greater than that of the material savings that waiting for the optimal solution can provide. It is not only material costs, but also intrinsic costs, such as in today's highly competitive market, late order delivery is an associated trait which all suppliers want to avoid.

This creates a number of extremely pertinent questions, for example, what is a good solution? How long are manufacturing facilities prepared / allowed to wait for a solution? Is there any such thing as a good solution? None of the answers to these questions are routine, they are all dependent on the current situation and status of the manufacturing facility at that time. This is best evaluated from real-time data collected from the shop floor, economic data based on previous sales and customer priority.

For the most part, the pieces entered into the system are of average size and the complexity of the problems is not of a serious nature. The solutions derived are straight-forward and the time needed to derive the solutions is not an issue, very often under 420 seconds. This is quite satisfactory for the facility. However, when the problem space is larger, there are more stock options, or more pieces of varying size to be placed, the time needed can be too long. The resources needed to solve the problem are also now an issue. Any application will consume all available computing resources regardless and therefore limit the resources available to other critical operations on that machine. If, as in some cases, the problem needs 7 hours to complete, the machine on which the computation is being conducted on is virtually rendered useless for the duration of the problem.

RESOLUTION

The key element in any decision making involved in material optimisation is real-time data. Any manufacturing system must have a detailed data supply structure available to provide data to the Shop Floor Controller or the decision maker. This data must consist of the shop floor status capacity, especially the status of the machine on which the cutting operation is to be conducted, this in many cases is the first operation/machine. It must also provide the urgency and priority of the line items as well as the quantity needed. This data will determine the urgency of a cutting list or plan and therefore the time available to the shop floor controller to solve the solution. In most cases the line items to be placed on stock will consist of a number of material/stock types and therefore have to be solved sequentially.

The material optimisation module is part of a larger integrated manufacturing solution system being developed in the Case Study Firm. The function of the material optimiser module is to gather the orders delivered to it and to propose the best pattern of cutting in a guillotine manner (Figures 5 & 6).

The system developed in the Case Study Firm (Figure 3) has included a time limiter set by the user. The first stage in the material optimiser operation is that an order horizon is selected by the operator. The operator can select day 'd' from 0-days to 12 days (2.5 working weeks). This means that when the optimiser is activated the system will include all orders with a due date falling in the next 'd' days from the current date. The optimiser will then attempt to optimise the pieces using the stock available. The optimiser first groups all the orders to be made from each stock type and then attempts to solve each one in turn. The time limit set will determine how long the optimiser searches for a solution for each stock type. If the time limit expires and there has been no solution found the optimiser returns a message stating that the time allowable for that material type has expired and it continues on, attempting to solve the next material type.

CASE STUDY FIRM II
PROCESSING AND STOCK AREA LAYOUT
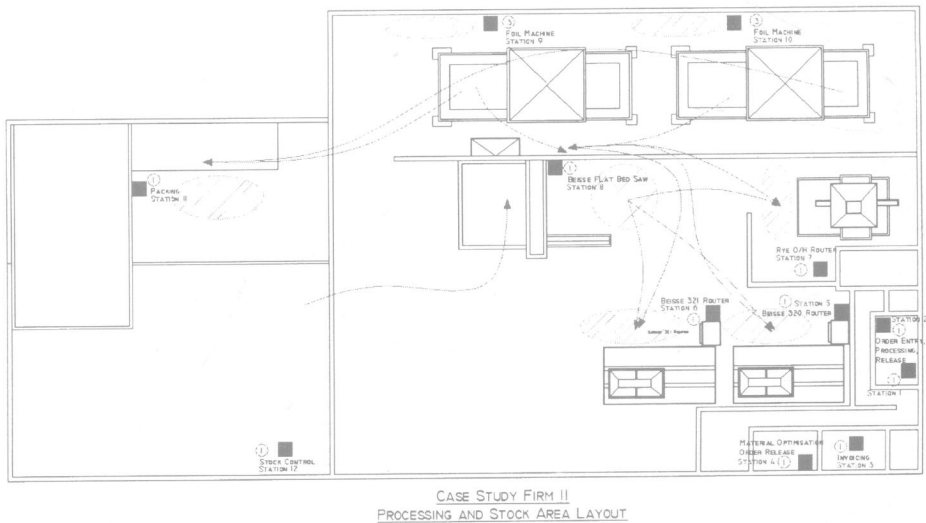
Figure 3

After attempts have been made to solve all the material types, solutions, to the materials where solutions were determined within the time allowed, can be viewed graphically. This graphical display outlines the pieces, their position, layout, orientation, dimensions, and individual piece identification code (Figures 7 & 8). The total tool path distance is also recorded as is the percentage waste and material usage. When the operator has viewed the solutions they can be accepted or rejected. If the solution is accepted the optimiser will provide printed graphical and textual displays of the pieces, position, layout, orientation, dimensions, etc. If rejected, the horizon can be widened/shortened in an attempt to include more parts and improve the solution.

The system also has an 'automatic' control (Figure 4). In this case, solutions for the stock types are attempted as usual. When all stock items are attempted, the solutions determined in the time-limit are printed to allow the operator start processing those particular order items. The other material stock pieces are selected and heuristics are used to attempt to solve the problem.

It has been observed from previous solutions that the problem complexity and solution quality is very much determined by the not only the number of pieces, but the size of the pieces relative to the stock available. For this reason a number of heuristics were developed.
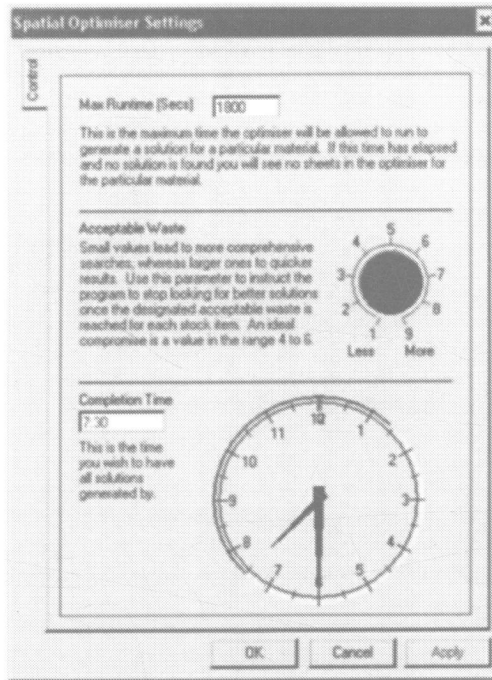
Figure 4

If the problem consists of a sizable number of small pieces to be placed on stock of a substantially larger size, the problem area is so large that the time expected can be unacceptable. In an instance such as this, the system determines if >80% of the pieces are >4.8 % of the average area of available stock. If this is true, the pieces to be placed are deemed small and order is split randomly into two problems and solved simultaneously, this time with a time limit of up to 7200 seconds.
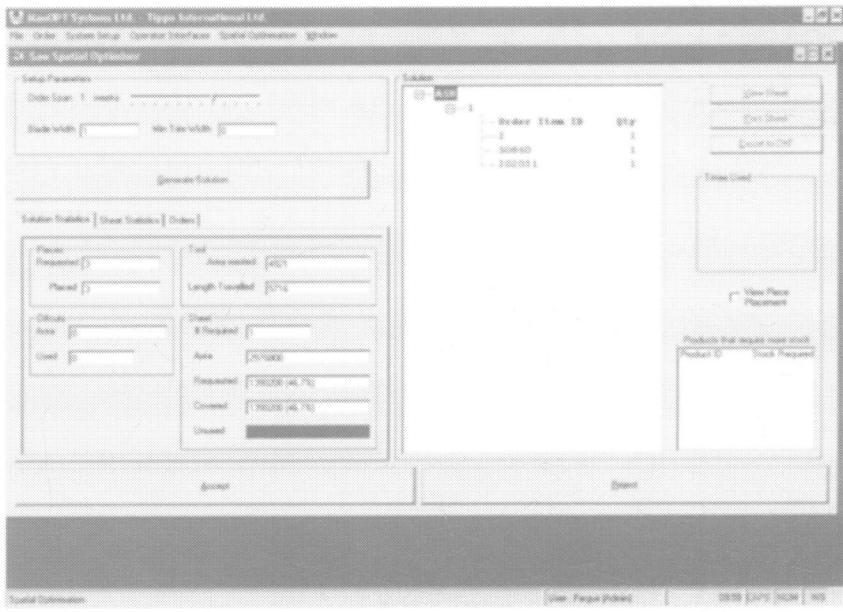
Figure 5
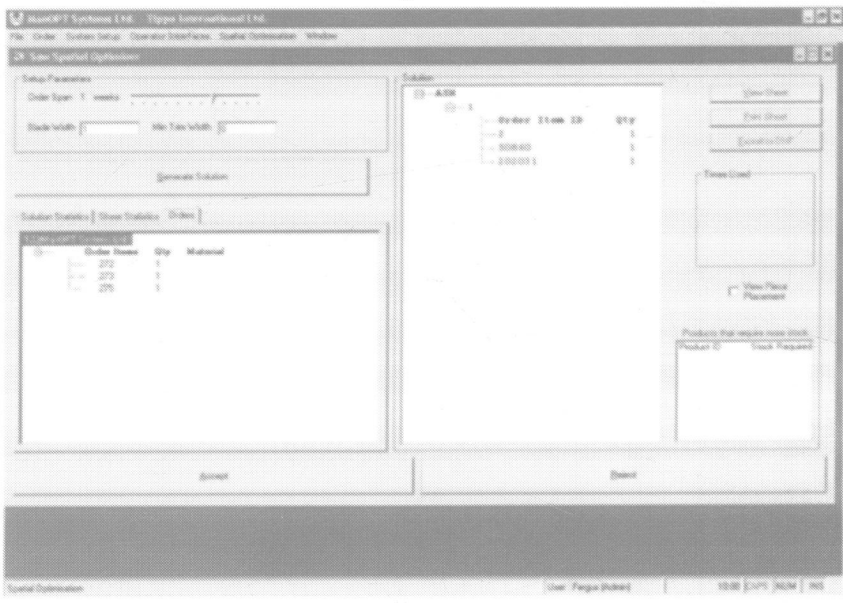
Figure 6

If the combined line items do not meet the above criteria, i.e. if <80% of the pieces are < 4.8 %, the order is deemed to be a 'mixed' order. The fastest manner to solve this problem is to solve the layout problem for the large pieces and attempt to fill the space left with smaller pieces. The smaller pieces not used are then solve in a separate solution. In this instance, the number of sheets needed to contain the larger pieces (those >4.8 % of the average area of available stock) is estimated. This number of sheets is rounded up and the area left over is determined, a number of smaller pieces, <4.8 % of the average area of available stock, are included and this is solved as one problem. Any pieces not used are then solved separately. Both problems are solved sequentially and again this time limit is 7200 seconds.

An additional heuristic introduced to assist in the manufacturing of the parts is to group the small pieces into groupings of similar widths. The advantage of this is that if the pieces are very similar in width the optimiser would find it easier to group the pieces as having the same dimension and fit them into strips of that measurement to be cut from the one board. This heuristic is particularly useful as it mimics the actual processing techniques and human decision making process that would be adopted normally. The grouping of such orders might be some what more costly in terms of material; however it is certainly faster and more practical in terms of processing as it reduces the number of cuts and processing time required.
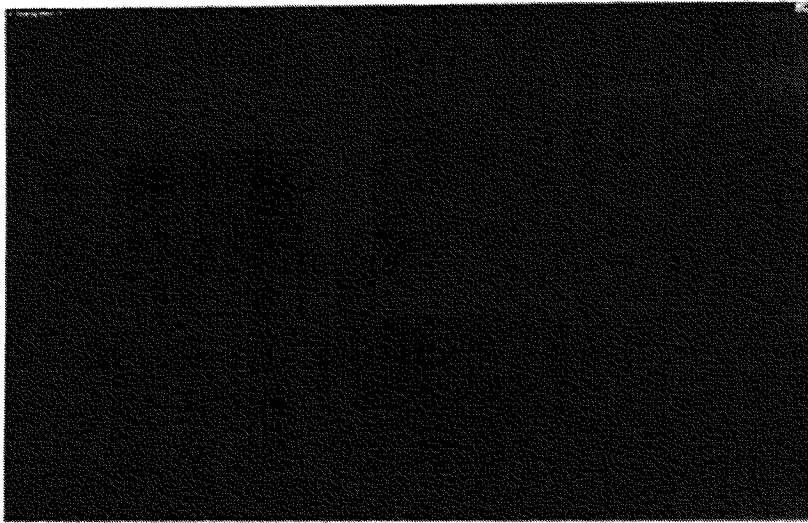
Figure 7

CONCLUSION

The time need to solve the problems presented to the Case Study Firm was reduced dramatically. The software developed reduced the total solution time required for certain line item combinations from 39600 seconds down to 1680 seconds. The orders used were combination orders taken exactly from industry and had no predetermined pattern. All experimentation was conducted on actual real-world data so as to test the application of the solution properly. In general though the heuristics applied in combination reduced the time needed to propose a solution by approximately 1/12th.
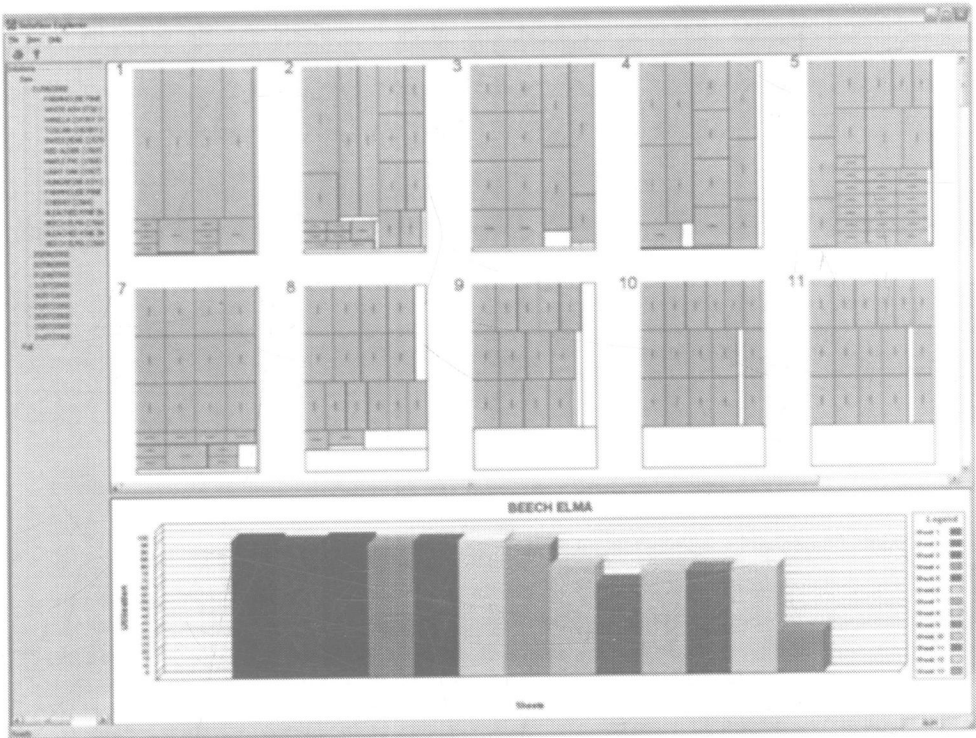
Figure 8

The total processing time improvement though is a lot harder to estimate as the time taken to generate the solution must be considered in light of the reduced solution time, but also in terms of the processing time now required. The system operates currently by optimising all the line item materials ordered. The user sets a time limit for each material and the solutions determined are presented. If the automatic setting is on, the optimiser will automatically start to re-optimise the materials not solved in the preset time. The user selects a time the following day at which they require a solution. The optimiser will divide the time left by the number of solutions to be solved and allow equal time to resolve each solution. This ensures that the solution has been resolved by the time selected the following morning. This research is on-going and is part of a larger project based on the development of an integrated material optimisation and scheduling system.

REFERENCES

András P., András, A. and Zsuzsa, S., 1996. A genetic solution for the Cutting Stock Problem. In: Proceedings of the First On-line Workshop on Soft Computing, Aug. 1996, Nagoya University, pp. 87-92.

Coffman E. G. and Shor P. W., 1990. Average-case analysis of cutting and packing in two dimensions. European Journal of Operational Research 44, 134-144.

Dowsland W. B., 1991. Three-dimensional packing-solution approaches and heuristic development. International Journal of Production Research 29, 1673-1685.

Dowsland K. A. and Dowsland W. B., 1992. Packing problems. European Journal of Operational Research 56, 2-14.

Dowsland K. A. and Dowsland W. B., 1995. Solution approaches to irregular nesting problems. European Journal of Operational Research 84, 506-521.

Dyckhoff H. and Finke U., 1992. Cutting and Packing in Production and Distribution. Springer Verlag, Berlin.

Dyckhoff H., 1990. Typology of cutting and packing problems. European Journal of Operational Research 44, 145-159.

Garey, M. R. and Johnson D. S., 1979. Computers and Interactability: A guide to the theory of NPcompleteness. W.H. Freeman and Company, San Francisco.

Gary Parker R., 1995. Deterministic Scheduling Theory. Chapman Hall.

Gilmore, P.C., and Gomory, R.E., (1961), A Linear Programming Approach to the Cutting Stock Problem, Operations Research, 9, 849-859.

Golden, B., 1976. Approaches to the Cutting Stock Problem. AIIE Transactions 8, 265-274.

Hinxman A. I., 1980. The trim loss and assortment problems. European Journal of Operational Research 5, 8-18.

Hopper E. and Turton B. C. H., 1997. Application of Genetic Algorithms to Packing Problems – A Review. In: P. K. Chawdry, R. Roy and R. K. Kant (eds.), Proceedings of the 2nd On-line World Conference on Soft Computing in Engineering Design and Manufacturing, Springer Verlag, London, 279-288.

Haessler R. W. and Sweeney, P. E., 1991. Cutting Stock Problems and solution procedures. European Journal of Operational Research 54, 141-150.

Hwang S. M., Cheng Y. K. and Horng J. T., 1994. On solving rectangle bin packing problems using genetic algorithms. In: Proceedings of the 1994 IEEE International Conference on Systems, Man and Cybernetics. Part 2 (of 3), San Antonio, TX, USA, 1583-1590.

Kerrigan, F., (1998), Optimisation by Simulated Annealing, Thesis, University of Limerick, Ireland.

Kröger B., 1995. Guillontineable bin-packing: A genetic approach. European Journal of Operations Research 84, 645-661.

Paull, A.E., (1956), Linear Programming: A Key to Optimum Newsprint Production, Pulp. Pap. Mag. Cam., 57, 146-150.

Rahmani A. T. and Ono N., 1995. An evolutionary approach to two-dimensional guillotine cutting problem. In: Proceedings of the IEEE Conference on Evolutionary Computation, pp. 148-151.

Rayward-Smith V. J. and Shing M. T., 1983. Bin packing. Bulletin of the Institute of Mathematics and its Applications 19, 142-146.

Sweeney P. E. and Paternoster E., 1992. Cutting and packing problems: a categorised, application orientated research bibliography. Journal of the Operational Research Society 43, 691-706.

Whelan P. F. and Batchelor B. G., 1993. Automated packing systems: Review of industrial implementations, SPIE, Machine Vision Architectures, Integration and Applications 2064, 358-369.

Rapid Product Development

MEM technology in making human skull-absent substitutes

G-X TANG, R ZHANG, and Y YAN
Mechanical Engineering of Tsinghua University, Beijing, China

ABSTRACT: The design and manufacture of a skull-absent substitute is a complicated clinical problem. By now every method that is adopted to repair a skull contains many defects under operation. They take a long time and are hard to ensure the accurate shape. Using MEM (Melted Extrusion Manufacture) rapid prototyping technique, which is based on the scattering/piling-up principle, a patient's skull and the skull-absent substitute solid models can be manufactured from CT scanning data. The substitute made by MEM technique fits the original skull very well. In this paper we introduce the theory of the MEM technique and describe the manufacturing process of the skull and skull-absent substitute. We also discuss the problems in making skull-absent substitutes using the MEM technique, which need to be solved in the future.

KEYWORDS: MEM technique; rapid prototyping; skull-absent substitute; surgical intervention

1 INTRODUCTION

The skull is one of the most important skeletons in the human body. Once the skull is injured, a patient would suffer from severe agonies. There are great quantities of skull repair operations every year all over the country. For example, the amount of skull repair operations was as high as 2,850 in 2001 in the neuro-surgery branch of Beijing Tian-tan Hospital. There is an urgent demand for developing a rapid and accurate repair forming technique. By now all the repair methods of skull-absent substitutes used are time consuming. Doctors need to shape the skull substitute repeatedly during operations. It increases the difficulty and is hard to make the skull substitute fit the original skull well. In order to solve the forming problem of a skull substitute, the MEM rapid prototyping technique has been utilized. This technique not only can produce a skull physical model according to the CT scanning data to help surgeon to plan and simulate an operation before surgery, but also can realize the rapid forming of a skull-absent substitute, which possesses the complicated inner and outer structures and fits the original skull very well. Once the promising technique is further used in the clinic, the operation time will be heavily reduced and the operation accuracy will be greatly enhanced.

2 EXPERIMENT PARTS

2.1 Material and equipment

The filament B107 with a diameter of Φ 0.36 mm is used as the forming material. It possesses high intensity, high rigidity, low contractibility and rapid solidification property, which are suitable for making a skull-absent substitute. The experimental equipment is the MEM-250-II, a rapid forming machine developed by Tsinghua University (see Fig.1). The dimension of forming chamber is 250 mm×250 mm×250 mm. The layer thickness can be treated within 0.05~0.6 mm. The forming machine has integrated the hardware, software and NC digital controlling system, and consists of a forming chamber, a squeezing nozzle and a movable lifting platform. It is controlled by a computer.

2.2 The principle of MEM technique

Fig.2 shows the principle sketch of the MEM rapid prototyping technique. It is a digital forming technique based on the scattering/piling-up principle. It makes the parts through piling up materials layer by layer on the forming platform. First, the materials are heated to become fusion or semi-fusion. Then, they are squeezed out continuously through the nozzle at a low velocity. Driven by the scanning system, the nozzle completes two dimensional scanning. When the squeezing and scanning are at the same step, the squeezed materials are accumulated to form the layer. After one layer is accumulated, the forming platform decreases by the height of one layer to continue the accumulation of the next layer till the whole part is manufactured.
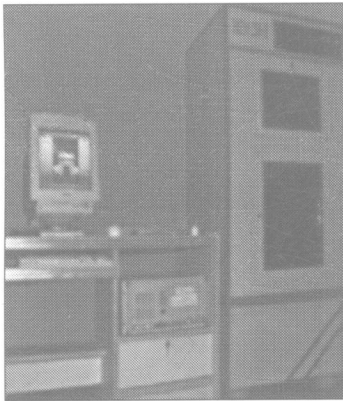
Fig.1   Profile of MEM-250-II machine
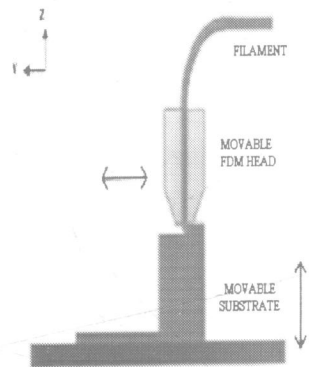
Fig.2   Principle of MEM technology

2.3 Data processing

The raw data of a patient's skull are in the form of cross-sectional CT-images. Each of the CT images contains the geometric information of a cross section of the measured part. An image segmentation algorithm is first applied to extract the geometric information from CT-image as

a set of 2D contours. These 2D contours are further converted into 3D contours by using a special software named MIMICS. The 3D raw CT-contours with the known scanning distance between the slices are then be used in further processing. A STL format file, which can be recognized and converted into CAD physical model data by a CAD system, is exported after the treatments. A Daphne slicing software developed by Tsinghua University is employed to segment the CAD solid model data layer by layer. Finally a file with the CLI format is created for the forming machine to use. The whole data processing flow chart was shown in fig3. In the data processing, interpolation is necessary between the planes. The volume data must be interpolated to adapt to the resolution of the model building process. The extraction of the interest surface of the object within the data set is more important. Usually threshold approaches are applied for segmentation, which is always critical. However, it is difficult to select an optimal threshold for an anatomical structure. Thus it depends on the surgeons to enhance accuracy and to ensure quality.

| CT image information | → | STL Format file | → | CAD Physical model | → | CLI Format file | → | skull Solid model |
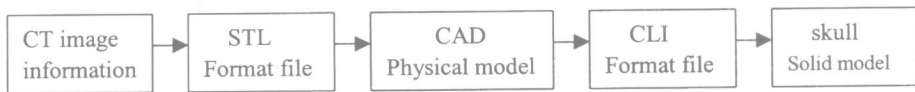
Fig3   The whole data processing flow chart

The original data in our experiment are provided by Beijing Tian-tan Hospital. The size of the skull-absent part of the patient is 60 mm×80 mm. The distance between the CT scanning layers is 2 mm. When the proper forming parameters, transmitting voltage of 6V, grid velocity of 30 m/s, supporting velocity of 22 m/s, forming chamber temperature of 55℃ and thickness of 0.15 mm, are set, the skull solid model with absent is rapidly manufactured by using the MEM forming method. Fig4 shows the physical model of the patient's skull.
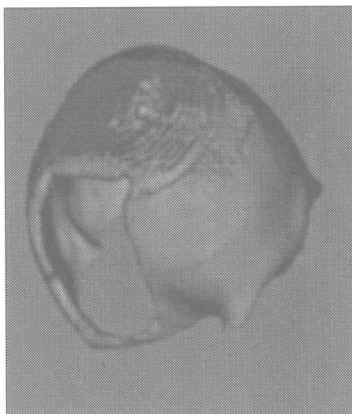
Fig4. Physical model of the patient'skull
With absent dimension of 60mm×80mm
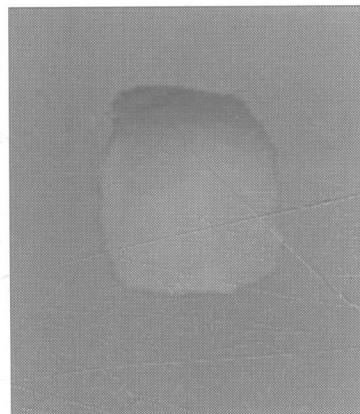
Fig5.   Physical model of the
skull-absent part

When fabricating the skull-absent substitute, more complicated work must be done. In order to obtain the skull-absent information data, we utilize the editing tools such as drawing, rotating and mirror toolbars in the MIMICS software to gain the information of the skull-absent part. After a STL format data is exported, we use the same MEM rapid prototyping method as mentioned above to construct a physical model of the skull-absent part. Fig. 5 shows the physical model of the skull-absent part.

3. Experimental results

By using the MEM rapid prototyping technique, we have made a solid model of the patient's skull and its matching part as shown in Fig.6 and 7. Fig. 8 shows that there is an appropriate transition between the skull-absent substitute and the original skull solid model, and the skull-absent substitute with a smooth surface fits the skull solid model very well. In conclusion, we can realize the rapid manufacturing of skull-absent substitute by using the MEM rapid prototyping technique.
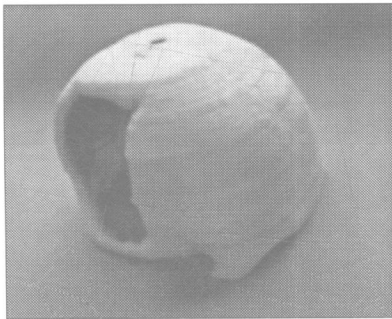
Fig6.   Solid model of the patient'skull
made by MEM technique
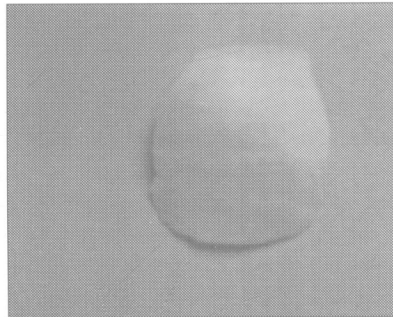
Fig7. Skull-absent substitute
made by MEM technique
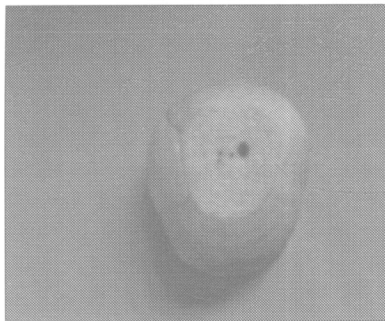
Fig8.   Well fitness between the substitute
and the original skull
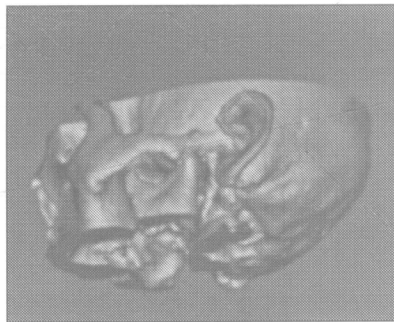
Fig9.   Colored MEM skull models distinguish
anatomical structures.

4 DISCUSSIONS

4.1 The advantage of MEM

The manufacturing of a skull-absent substitute is individualized. The size of skull-absent parts from different patients is not the same. It demands the manufacturing technique to be rapid and flexible. The traditional methods such as tool cutting, laser cutting, injection molding and die molding can not make the forming processes suitable to the arbitrarily complicated topological geometrical figure. Whereas the MEM rapid forming technique based on the scattering/piling-up principle can implement it. It is beyond the complicacy limitation of the parts and does not need special machining facilities. The technique can obtain the maximal flexibility and rapidness, and meet the demand of rapid manufacturing of the skull-absent substitute in clinic. Furthermore, there is a choice of materials and even different colors. This allows the reproduction of more than one anatomical structure in one model distinguished by color (see Fig9). By using colored MEM skull models, several vulnerable structures like nerves can be avoided in the placement of the fixation pins and screws. So it makes surgical operations more precise.

4.2 Data acquisition of the skull-absent part

The original data contain the cross-sectional geometric information of the skull. After vectorization the data, a three-dimension physical model is produced for further use. To obtain the high quality and accurate physical model, the interpolation of physical data is needed. When choosing the area of image segmentation, an experienced doctor is needed to help ensure the smoothness of the substitute and meet the demand of a clinic. Moreover, many more efforts must be done, such as referring to the data of the healthy hemi-skull, using the edit toolbars in Mimics software to plot the skull-absent part layer by layer, and canceling the unwanted data of the skull in each layer. Then the data information of a skull-absent substitute can be acquired. However, the method is not perfect. Having more scanning layers by CT makes it easier to obtain a smooth surface but is also more time-consuming. Decreasing the number of CT layers saves processing time, but it will inevitably reduce the quality and cause a ladder-style surface of the final parts. So a more accurate and effective acquisition method of obtaining skull-absent part data is needed.

4.3 The further development of making skull substitute by MEM technique

The substitute, which meets the clinic demand from geometric shape and fits the original skull-absent part well, can be manufactured by the MEM rapid forming technique. But the kinds of repair material suitable to MEM are very few now. The ultimate goal of the skull-absent substitute is to implant it to a human body. The material must be compatible with the MEM rapid forming technique and also meet the demand of biologic compatibility. Therefore it becomes the research emphasis to develop an appropriate repair material. Moreover, depending on the complexity of the model, the number of slices and the resolution of the data set, the production of a model may last up to one day. To establish this technique for a wide range of medical applications, it is necessary to further reduce the building time.

5 CONCLUSIONS

1) The clinic significance to manufacture a skull-absent substitute is extensive. The substitute, which fits the original skull-absent part well, has been manufactured by the MEM rapid forming technique. It meets the clinic demand.

2) The CT data of a patient have been utilized directly to the reconstruction. A three-dimension digital model is created by a series of computer software. The skull-absent substitute possessing the complicated inner and outer structures is manufactured by the MEM forming machine. This opens up a new approach for clinic implantation.

3) Doctors can obtain the substitute that fits the skull-absent part well before operations, which greatly reduce the operation difficulty and time, and benefits the patients as well.

4) Further study is needed to find bio-materials which are suitable to MEM rapid forming technique and the acquisition of substitute data information.

REFERENCES
1. **Petzold R, Zeilhofer H.-F. Kalender W.A.** Rapid prototyping technology in medicine-basics and applications. Computerized Medical Imaging and Graphics. 1999;23:277-284

2 **Zhuo Xiong, Yongnian Yan, Renji Zhang.** Layered manufacturing of tissue engineering scaffolds via multi-nozzle. 5th Asian Symposium on Biomedical Materials'2001. 239~248

3 **Liangwei Wu, Yongnian Yan, Guodong Hong, et al.** The accuracy analysis of melted extrusion manufacturing. China Mechanical Engineering. 1997(5):23-34

4 **Yongnian Yan, Renji Zhang, et al.** Research on the bonding of material paths in melted extrusion modeling, Materials and Design, 21(2000), p.93-99

5 **Liangwei Wu, Guodong Hong, et al.** Material issues in melted extrusion manufacturing. International Conference on Rapid Prototyping&Manufacturing'98. 159~165

Corresponding author: Tang Guang-xin

Telephone: (86-10)62782988 (O) 、(86-10)62778936 (H) ,
fax:010-62788675
e-mail: tgx01@mails.tsinghua.edu.cn
Post code: 100084
Address: Department of Mechanical Engineering, Tsinghua University, Beijing, P.R.China.

Application of rapid prototyping to fabrication of casting mould

Y SHI, X LU, N HUANG, and S HUANG
Rapid Manufacturing Center, Huazhong University of Science and Technology, Wuhan, People's Republic of China

ABSTRACT

It will take much time and expenditure to manufacture casting moulds or parts using traditional casting technologies, which can not meet the requirements of the today's market. For this reason, new methods and flow routes of rapid manufacturing casting moulds or parts based on Laminated Object Manufacturing(LOM) and Selective Laser Sintering(SLS) are presented in this paper. Using these methods and flow routes, the time and cost of developing new production can be reduced greatly, which has been proven by practical application.

1 PREFACE

Nowadays competition becomes more and more cruel. New products will be taken the place soon. It's estimated by a motor company in China that 6.5 million more dollars will be brought to the company and its related plants if the new products come into the market a month early. After the entrance of WTO, whether she could react to market in time is important for Chinese manufacturing industries to survive. Rapid prototyping (RP) , as an advanced fabrication technology, is an available method to solve the above-mentioned problems. For example, a complicated gear case was fabricated by Volkswagen Automotive Company Ltd. with RP technology in two weeks rather than in eight weeks, and 2000 casts were fabricated by the famous American Pratt & Whitney Lab using RP technology, which saved 90 percent of cost and 70 percent of time compared with its using traditional methods. Thus, it is very necessary to study and spread RP technology.

Rapid Manufacturing Center of Huazhong University of Science and Technology(HUST) has made great progress in RP and rapid tooling(RT) in China, and has gained the following achievements until recently:

(1) RP Systems

LOM and SLS were mainly studied, and a series of achievements have been attained in RP machine, materials, software, control system, laser and so on.

A. Nine patents have been obtained, such as tensionless LOM delivery device and so on, which made LOM and SLS machines run reliably. The precision of LOM green parts reaches ±0.10mm, and the precision of SLS green parts ±0.20mm. The cost of RP machine decreases largely.

B. LOM materials were investigated, and the green parts with the thickness reaching 1mm and local thickness even reaching 0.3mm were made. Therefore, it enlarges the applying scope of LOM. Five low-cost SLS materials were developed successfully.

C. RP application software of self-determination patent right was developed.

D. Two multi-task、large-data、multi-axis、high-speed and real-time control systems based on net communication and integrated circuit were developed successfully.

E. A series of RP systems of HRP(Figure 1) and HRPS(Figure 2) were developed, whose security and technical index reach international level and cost is 1/4-1/3 of the same other products(1).
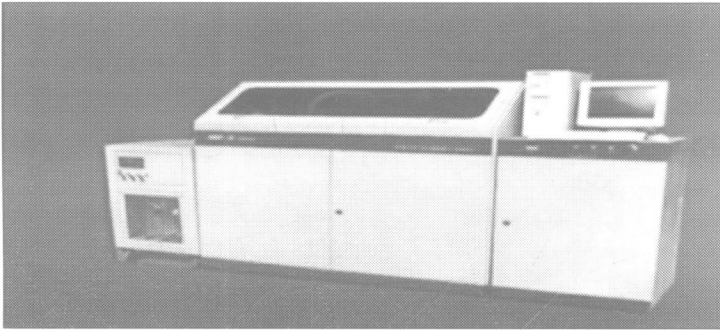
Figure 1.   LOM system of HRP-IIIA developed by HUST
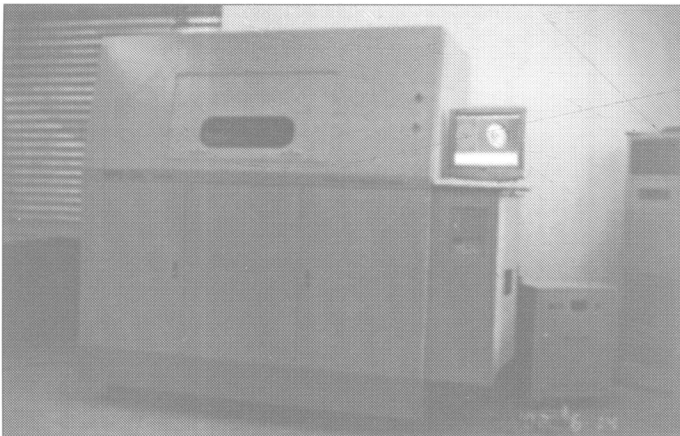
Figure 2.   SLS system of HRPS-III developed by HUST

(2) The study of RT. After a few years of study and practice, the flow routes in Table 1 were developed successfully.

Table 1. The flow routes of RT based on RP

RP Systems	The flow routes of RT
HRP **(LOM)**	(1) LOM prototype—silicon rubber mould—plastic parts (2) LOM prototype—plaster prototype(ceramic prototype)—casts (3) LOM prototype—silicon rubber mould—plaster prototype(ceramic prototype)—casts (4) LOM prototype—transfer coating—resin sand or soluble glass sand mould (core)—casts (5) LOM prototype—wood mould for casting
HRPS **(SLS)**	(1) SLS investment mould—ceramic(plaster) mould—casts (2) SLS coated sand mould(core)—casts (3) Direct SLS of plastic parts (4) Direct SLS of injection mould

In this paper, rapid casting methods, flow routes and some examples based on LOM and SLS are introduced.

2 RAPID MANUFACTURING OF CAST BASED ON LOM

2.1 Direct manufacturing plaster mould through LOM prototype, and then casting aluminum alloy casts.

2.1.1 Design of 3D model

3D model is drawn using prototype software(such as UG, etc.) and turned to data model accepted by RP system. Then slices of 3D model are produced using the RP application software in order to fabricate green part layer by layer.

For example, 3D foam model of a bar grate mould shown in Figure 3a is designed bisectionally based on 3D model. The dimension and thickness of the mould are $340 \times 240 \times 25mm^3$ and 11mm, and its foam mould is shown in Figure 3b.
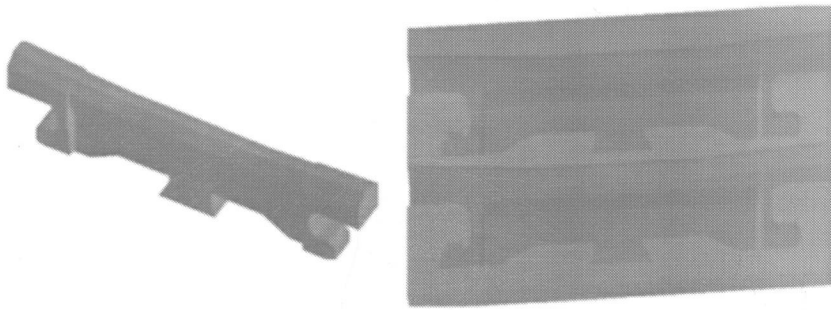
(a) 3D foam model of a bar grate   (b) the foam mould of a bar grate

Figure 3.   Foam model and foam mould of a bar grate

2.1.2   *Rapid Manufacture of mould prototype*

Paper mould prototype is fabricated using LOM system of HRP-III. LOM has a series of advantages of rapid manufacture and low-cost. It takes eight hours to fabricate the bar grate mould prototype. After polished, the surface of the paper mould prototype is smooth and plane. The size of the prototype is 100 ± 0.2mm which meets the demand of the foam mould. Figure 4 is a photo of the fabricated mould prototype.
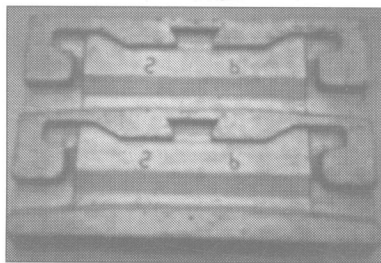
Figure 4.   The paper mould prototype of the foam mould

2.1.3   *The casting process based on mould prototype*

The foam mould cavity of the bar grate is very shallow, so it is easily removed from the mould. Therefore, the precision casting of plaster prototype is selected directly to meet the high demand of roughness for the cavity face(Ra6.3-3.2 μ m). The roughness requirement of the external mould surface is not very high, so resin sand may be coated directly on the paper mould prototype to model.

Aluminum alloy casts are cast in a constant pressure and often appear such defects as air hole, contraction cavities and loose because the gas-penetrability and the heat conduction performance of the plaster mould are very low. To attain aluminum alloy mould with close-grained structure, casting process under a vacuum condition is adopted: the plaster prototype and resin sand prototype are assembled and put in a sand case, then dry sand of 20-40 mesh is filled in the sand case and a layer of thin plastic film is coated on the sand case, finally the bottom of the sand case is vacuumized when casting in order to produce negative pressure. The aims of casting under the negative pressure are as follows :

(1) Help the plaster prototype exhaust and prevent the mould producing air hole.

(2) Enhance the filling performance of aluminum alloy liquid in the complicated cavity.

After casting, let in air immediately to compensate the shrinkage of casts through the casting sprue which helps to eliminate the loose structure of aluminum alloy.

2.1.4 Fabrication of foam mould

Compared with sand prototype casting, aluminum alloy mould of plaster prototype casting has high size precision and doesn't need to be processed. After smoothed carefully by hand, the roughness of aluminum alloy mould can reach Ra3.2 µ m. PS grains are injected into the assembled mould which is put in a casting steamer, and then the foam mould may be fabricated.

It is optimum to fabricate a few middle and small foam moulds with the casting steamer. The foam mould of the bar grate and its foam part are shown in Figure 5.
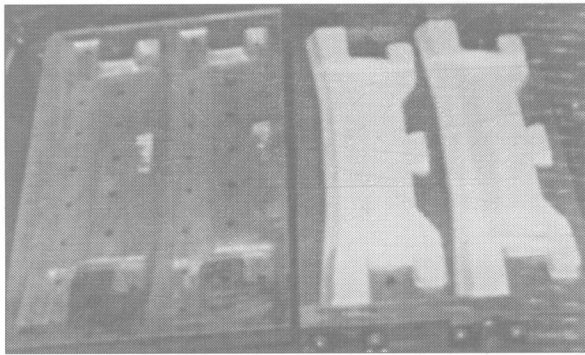
Figure 5. The foam mould and its foam part of the bar grate

The time of the above-mentioned whole process decreases by 1/3 of the traditional process from the 3D model design of the bar grate product to the mould prototype and from the aluminum alloy casting to the fabrication of the foam mould. The size precision of the aluminum alloy mould can reach CT7 or higher. The roughness of the cavity surface can reach Ra3.2 µ m. The green parts of the aluminum alloy mould are smoothed, and then the foam parts could be produced with them(2).

2.2 Using LOM part to manufacture plaster mould through a silicon rubber as the transitional mould and then casting aluminum alloy casts.

For some casts with deep cavity and difficultly drawing mould, a plaster mould is manufactured through a silicon rubber as the transitional mould and then aluminum alloy casts are cast. The aluminum alloy mould shown in Figure 6 is a metal cover mould (one mould for making two parts) of railway electrization made by means of the plaster mould which is made by the silicon rubber transitional mould. When the metal cover mould is manufactured, a LOM mould is made into a protrudent one according to 3D design and the

concave mould of the silicon rubber is made using the LOM mould in order to produce the protrudent plaster mould with which the aluminum alloy mould may be cast.
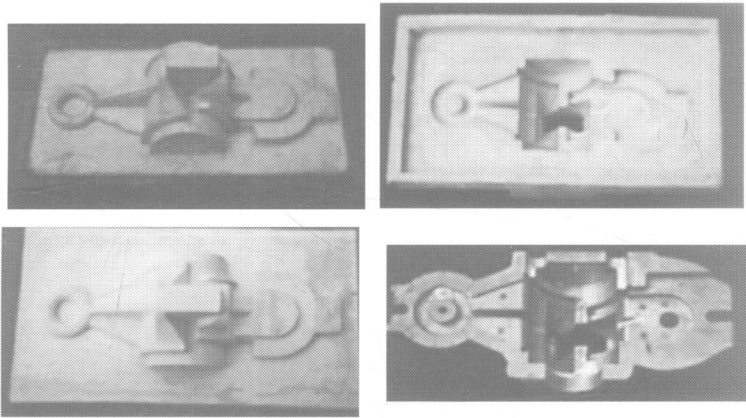
Figure 6. **An aluminum alloy mould manufacturing process through transferring LOM prototype to silicon rubber mould**.

2.3 Rapid manufacturing of casts based on LOM prototype and transfer coating.

For manufacturing process of some complicated moulds, it is difficult to guarantee the quality of the mould using the conventional sand casting process. For example, it will take much not only time but also cost to use both NC and electric spark machining. It will be economical and rapid to manufacture the complicated moulds if the precision process based on LOM prototype and transfer coating is adopted. Figure 7 shows the rapid manufacturing flow chart of the precision process.
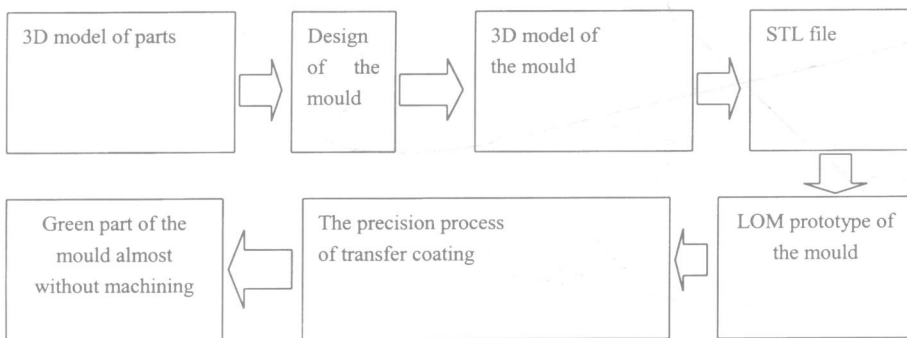
Figure 7. **Rapid manufacturing flow chart of metal moulds based on LOM**

Figure 8 shows 3D model of a turbine shell which can be changed into STL files to make prototype mould using HRP−III LOM machine(figure 9). Its precision reaches 100 ± 0.2mm and the roughness is Ra1.6 μ m.
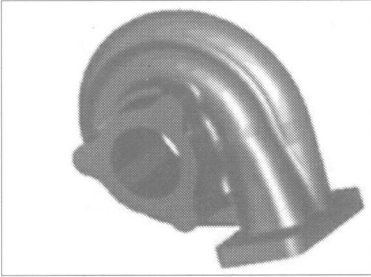
Figure 8.   3D model of a turbine shell   **Figure 9.   LOM prototype mould**
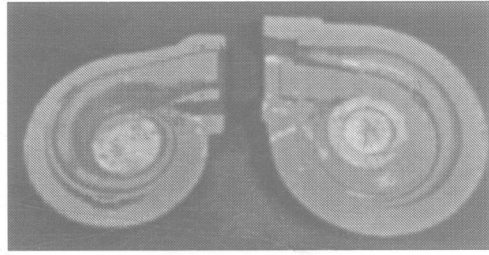
Transfer coating is a kind of coating which is directly coated on LOM green moulds in which the sand is filled. After the sand with coating ossifies, the shape of the LOM prototype moulds will transfer to the surface of sand mould or core. A cast with precise size and smooth surface can be cast because the coating doesn't influence the size of the moulds (3).

The process of the above-mentioned manufacturing the mould is as follows: First, the LOM mould is placed on a piece of wood covered a layer of plastic film and sprayed with a water-based stripper. After the water-based stripper vaporizes, the seen surface of the LOM mould becomes smooth. Second, a self-ossified transfer coating is brushed onto the seen surface of the LOM mould to make a film with the thickness of 0.5-1.5mm. Third, a water-glass sand is filled in the LOM mould, which is blown with CO_2 to ossify the water-glass sand and stayed for 2-4 hours. Fourth, after the LOM mould filled water-glass sand is turned over, the above-mentioned procedures are repeated on the other surface to get the other half of mould cavity. Finally, a special device for drawing mould is used to steadily draw mould and to assure the coating layer shifting to the mould cavity safely, through which a complete and precise casting mould is attained.

The casting mould is roasted 10 hours under the condition of 100℃-240℃, and then it may be used to cast.

ZL108 alloy is selected as the material of the casting mould. The casting temperature is 700℃. After the casting mould is cooled, its casting system and overlap are gotten rid of, and it is smoothed locally and polished, and then a green mould without machining is gotten. After some exhaust plugs and related parts being added, the final mould is fabricated.

Figure 10.   The foam moulds of turbine shell and its cast

Figure 10 shows the whole foam moulds of turbine shell and its cast. The testing result shows that the size precision of cast can meet the demand of design absolutely.

3   RAPID MANUFACTURING OF METAL PARTS BASED ON SLS

3.1 Precision casting of investment mould based on SLS

Using the HRPS-III SLS machine, many sorts of complicated polymer green parts are fabricated. Then the green parts are made into a ceramic prototype shell by the process of precision casting of the investment mould. Thus many metal parts with thin thickness could be cast using a vacuum differential pressure casting machine. Figure 11 shows the flow chart of the process.
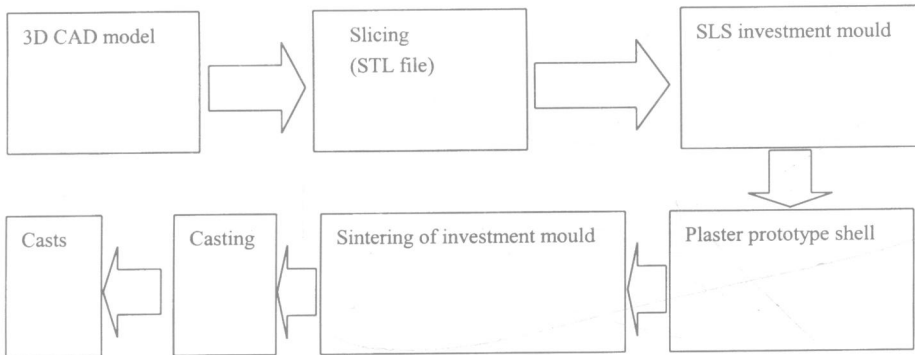
Figure 11.   The flow chart of rapid manufacturing process of metal parts based on SLS

3.2 Sand prototype casting based on SLS

Casting mould and core could directly be made with coated sand by SLS, as long as its 3D CAD model is designed. Figure 12 shows the casting mould, core and metal cast made by a HRPS series of SLS machine developed by HUST.

Figure 13 shows some aluminum alloy casts with thin thickness manufactured through the SLS

investment mould. The aluminum alloy casts with thin thickness by precision casting of the vacuum differential pressure are shown in Figure 14.
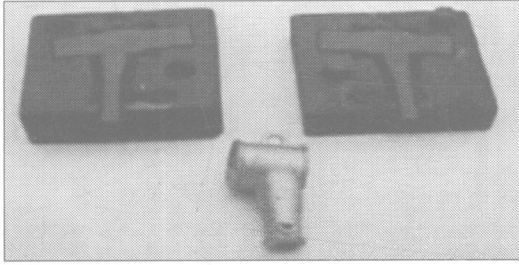
Figure 12.   The casting mould, core and metal cast
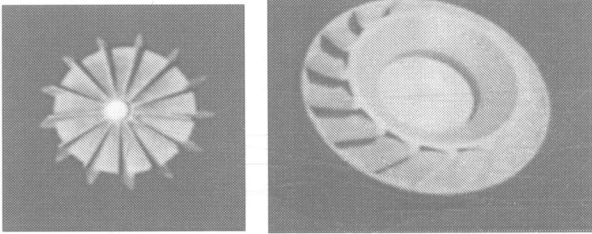
Left: an impeller,   minimum thickness 2.0mm , maximum thickness 10mm   size：$\phi 120 \times 45 (mm^2)$

Right: a turbine of supercharger,   minimum thickness 1.5mm , maximum thickness 20mm   size：$\phi 220 \times 80 (mm^2)$

Figure 13.   Classical aluminum alloy casts with inhomogeneous thickness
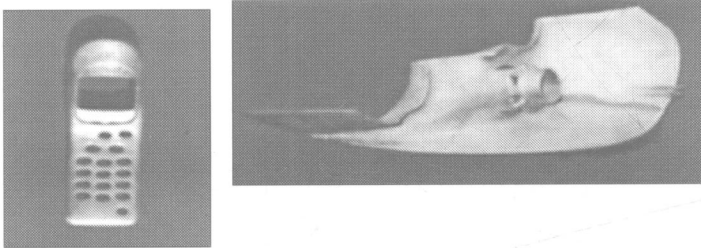
Left: mobile shell, average thickness 1.0mm, minimum thickness:0.7mm , size:145x53.4x6.5(mm^3)

Right: direction lamp shell for car, average thickness 2.6 mm, minimum thickness1.5mm , size: 350x150x100(mm^3)

Figure 14.   Classical thin thickness aluminum alloy casts

4   DEVELOPMENT AND TRENDS

(1)   Direct manufacturing metal moulds or parts with SLS.
(2)   Manufacturing complicated and thin thickness magnesium alloy parts through precision

casting with SLS investment mould.

5 CONCLUSIONS

Compared with traditional fabrication technologies of casting mould, RT has the following features:
(1) Manufacturing time is shorter from design of moulds to mould parts. It takes a few hours to fabricate a mould part with its CAD model by RT technology. But if using traditional methods, it will take a few months to make the mould. Thus the developing period of new products is reduced greatly.
(2) There is no limitation in design of the mould structure. It is much easy to design moulds with RT. Some complicated moulds, which are not fabricated by means of any other technologies, could be made with RT. So RT is available for design of the mould without considering the difficulties during fabrication.
(3) The whole RT process is automatic except the post processing by hand.
(4) The cost of developing new products is reduced greatly. It is more economical to use RT technology than traditional machining methods.

REFERENCES

1 **Yusheng Shi, Shuhuai Huang, Zude Zhou, Guoqing Chen, Jianguo Yang.** (2000) Design and development of low-cost selective laser sintering equipment. Mechanical Engineering of China. Vol.10, pp.1123-1125, 1135.

2 **Shengping Ye, Zhichao Wu, Hanbing Fei.** (1996) Rapidly Making Practice of Foam Molding Tools for Lost Foam Process. Special Casting & Nonferrous Alloys. Vol.6, pp.13-15.

3 **Shenping Ye, Zhichao Wu, Wenfeng E.** (2001) Metal mold manufacturing based on rapid prototyping in directly transferred coating process. Special Casting & Nonferrous Alloys. Vol.5, pp.41-44.

The Author: Yusheng Shi, male, Ph. Doctor, was born in June, 1962. And he is a professor of Huazhong University of Science and Technology and majors on research and development of RP/RT . He has written more than 40 research papers.

Address: Rapid Manufacturing Center, School of Material Science and Engineering, Huazhong University of Science and Technology   Wuhan   430074   P. R. China

Tel: 027-87544167   **E-mail:**shiyusheng@263.net

The research of a self-adapting delaminating algorithm based on profile loop and its application to rapid prototyping system

D CAI, Y SHI, and S HUANG
Rapid Manufacturing Center, Huazhong University of Science and Technology, Wuhan, People's Republic of China

ABSTRACT

Along with development of the rapid prototyping (RP) technology, the contradiction between fabricating speed and precision become more and more acute. In order to solve the contradiction, an effective self-adapting delaminating algorithm, which can judge accurately whether two layers break, is put forward in this paper. The algorithm only utilizes the profile loops information obtained from slicing CAD model, so its operation efficiency is high. The self-adapting delaminating method can increase fabricating speed and reduce running cost under the condition of ensuring fabricating precision. Usually, the fabricating time can be save from 20 to 80 percent according to the different parts.
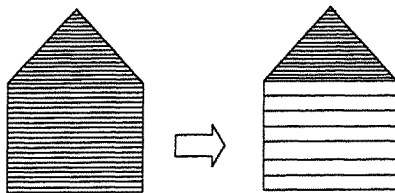
Fig. 1 Self-accommodating delaminating
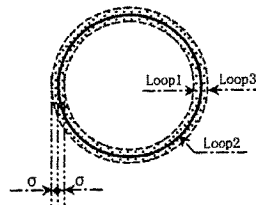
Fig. 2 Containing area of profile loop

1 INTRODUCTION

During the process of fabrication in rapid prototyping system, the operator often hopes to adopt the different layer thickness at the different height of CAD model in order to save fabricating time. Namely, the operator must take thin layer only to avoid emerging large sidestep at some height and can adopt thick layer at other height for saving fabricating time. For example, at the cone of the model shown in the fig.1, the thin layer is taken. However, at the cylinder of the model, taking thick

layer can save fabricating time. The self-adapting delaminating technology can meet with this demand. The method, which can adjust automatically the layer thickness at the different height of the model according to CAD file information, is called as the self-adapting delaminating technology. Of course, some technological parameters should be in harmony with fabricating layer thickness. For example, when the layer thickness is increased, the laser power should be enhanced.

2 ANALYZE SEVERAL SELF-ADAPTING DELAMINATING METHODS

The self-adapting delaminating technology is one way to solve the contradiction between fabricating speed and precision. For this reason, many scholars and experts try to search the proper self-adapting delaminating algorithm. Several self-adapting delaminating methods had been put forward. But these methods have some serious shortcomings more or less. Furthermore, the all methods just are being at theoretical degree. Nowadays, the self-adapting delaminating methods include mainly the curvature numeration and the acreage numeration. In addition, Dr. Rendong Wu, a doctor in Tinghua University, had studied a self-adapting delaminating method based on Wavelet analysis(1) at one time. It is regretted that this methods cannot be applied to practice.

2.1 The comparing curvature method

The comparing curvature method determines the current layer thickness according to the normal vector, which stands for the curvature of surface, of mini-triangle facets in STL model. The brief ideas of the method are as follows: firstly suppose the current fabricating layer thickness, which is relation to the RP equipments and equal to the maximal layer thickness, and compare the curvatures of all mini-triangle facets at maximal layer thickness height of STL model. If the curvature of every mini-triangle facet exceeds the pre-enacted criterion that be confirmed by the fabricating precision of RP part, then the current fabricating layer thickness should be less than maximal layer thickness. Then reduce the presupposed current fabricating layer thickness a mini-thickness, and repeat the above process until the presupposed current layer thickness reaches the minimal fabricating layer thickness. If the normal vectors of all mini- triangle facets at one layer thickness are less than the pre-enacted criterion, then the layer thickness should be considered as the proper current manufacturing layer thickness. The main strong points of the method are clear principle and little oddness case, and its serious shortcomings are excessive numeration and much repeated calculation.

2.2 The comparing area method

The comparing area method determines the current layer thickness according to the area error between the comparing layer and the base layer. The comparing area method is similar to the comparing curvature method. The difference is that the comparing area method is to compare area and the comparing curvature method is to compare the curvatures of all mini-triangle facets. The main strong point of this self-adapting delaminating method is small repeated calculation, and its serious shortcoming is that many oddness cases are not solved. Although some oddness cases by acreage aggregation operating may be eliminated, apparently it will increase calculation quantity largely.

2.3 The self-adapting delaminating method based on Wavelet analyzing

The method comes from the view of signal analysis. Analyze the shape profile from the character, which the wavelet commutation can combine flexibly the time information with the frequency information, in order to obtain all frequency components of RP parts. The delaminating principle of the method comes from analyzing frequency information. Namely, the current fabricating layer thickness is thin at the height, which its components of high frequency are more, and the current fabricating layer thickness is thick at the height, which its components of low frequency are less. This method has little oddness case, but it needs to put up a complicated math model. Furthermore, the frequency component standard is hard to be confirmed.

3 THE PRINCIPLE OF ADAPTING DELAMINATING ALGORITHM BASED ON LOOP CONTAINING

A group profile loops can be obtained from slicing STL model at one height, and the error information between two layers can be acquired from comparing and analyzing the profile loops. Before analyzed detailedly, it is necessary to explain a few technical definitions such as the corresponding profile loops, the plane error criterion, the jiggling layer thickness, etc.

The corresponding profile loops come from two adjacent layers obtained from slicing STL model. Under the condition that the error between the two adjacent layers(layer0 and layer1), is imperceptibility, when all dots of a loop on layer0 move upwards along the surface of model to the height that the layer1 locates at it, then all dots should locate in a loop on layer1, in this way, name the two loops as the two corresponding profile loops. The every loop between layer0 and layer1 must be corresponding relation if the error between the two layers is imperceptibility. Fig.3 shows the corresponding relations of loops between two layers. It is clear that the all-corresponding loops are similar in shape and size. Name similarly such two dots, which locate respectively at the same azimuth of two corresponding profile loops, the corresponding dots.
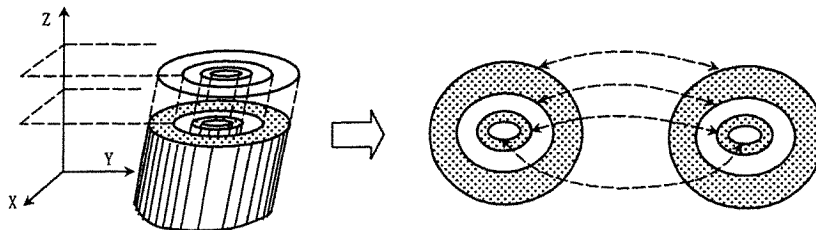

Fig. 3 The corresponding relation

The plane error criterion is to judge whether the distance of two corresponding dots between two corresponding profile loop exceeds acceptable size standard. If the distance between two corresponding dots exceeds the plane error criterion, means that the size error between two layers is too large, in other words, the current fabricating layer thickness is too large.

The jiggling layer thickness is a very thin mini-thickness, when the error between the current analyzing layer and the previous fabricating layer is less than the plane error criterion, in order to

farther analyze whether can adopt more layer thickness to fabricate, need to increase a mini-thickness to the current fabricating layer thickness. The increasing mini-thickness is named as the jiggling layer thickness.

Now, the jiggling layer thickness and the plane error criterion are enacted as " δ " and " σ ". In this paper, the " σ " is assumed to be equal to the minimal fabricating layer thickness. And the" σ " should be determined according to the factual applying case. Here, the " δ " is 0.05 mm, it is enough fine to the normal rapid prototyping system. Now obtain the base profile loops by slicing the STL model at the start fabricating height, add a minimal layer thickness " δ ", this height is the current analyzing layer height. And then the STL model at the current analyzing layer height is sliced to obtain a group profile loops. Firstly compare the number of profile loops between two layers, if they aren't equality, then the error between two layers is too large and the current analyzing layer thickness is too thick, namely the current sintering layer thickness should be equal to the current analyzing layer thickness minus the " δ ". On condition that the number of loops between two layers is equal, put up the one-to-one corresponding relation of the profile loops between two layers, and then compare the corresponding profile loops in turn, if the size error between any two corresponding profile loops exceed the enacted plane error criterion " σ ", it means that the error between two layers is too large and the current analyzing layer thickness is too thick. Only when the errors between all corresponding profile loops are less than the enacted plane error criterion " σ ", consider that the error between two layers are within the acceptable degree to the self-adapting delaminating method. Then add a " δ " to the current analyzing layer thickness, and circulate the above comparing until the current analyzing layer thickness reach the maximal layer thickness.

Putting up the corresponding relation is to compare only the corresponding loops in turn and not need to compare the all profile loops, in this way, a mass of calculating can be save. Now how to judge whether the error of two corresponding loops exceed the enacted plane error criterion? As shown in the fig.2, the loop loop2 is one of the two corresponding loops, magnify and deflate the loop loop2 according to the plane warp criterion " σ ", and that the loop loop1 and loop3 can be obtained. If the all dots in the other loop of the two corresponding loops locate in the shadow region surrounded by the loop loop1 and loop3, then, to the adapting-delaminating method, consider that the error of two corresponding loop is within the acceptable error area.

4 ACCOMPLISH THE ADAPTING DELAMINATING ALGORITHM BASED ON LOOP CONTAINING

The self-adapting algorithm based on profile loop containing consists of some branch-algorithms, and that the every branch-algorithms could judge partly whether the error of two layers exceed the acceptable scope. Now initializes some parameters· the height of the base layer is "h", the jiggling layer thickness is " δ " ,the minimal layer thickness is " σ ", and the plane error criterion also should be equal to the " σ " The process of detailed algorithm is as follows:

4.1 Compare the number of profile loops obtained by slicing STL file

Slice respectively STL file at the height "h" and "h+ σ + δ ", the two groups profile loops be

marked with layer0 and layer1. Compare the number of the layer layer0 and the layer layer1, if they are not equal, apparently the error of the two layers exceed the acceptable degree for the self-adapting delaminating method, then the current fabricating layer thickness should be "h+ σ ". When the number of loops between the layer layer0 and layer1 are equal, continue to next process.

4.2 Ascertain the one-to-one corresponding relation of loops between two layers

Putting up the one-to-one corresponding relation of profile loops is to reduce the repeated calculation. Suppose the all errors between two layers are less than the error criterion, and then the one-to-one corresponding relation of profile loops can be put up. And then analyze the correctness of the supposition by comparing the corresponding loops one by one, if the error of all corresponding loops are less than the enacted error criterion, then it shows that the above supposition is right, or else it is false. Now, for putting up the one-to-one corresponding loops between two layers, need to find out the maximum and minimum of the X-coordinates and Y-coordinates of all dots in a loop, and regard them as the maximum and minimum of the X-coordinates and Y-coordinates of the loop. After it, arrange the sequence of loops on the layer layer0 according to the minimum of the X-coordinates and Y-coordinates of the loop. The principle of arranging sequence is as follows: A loop should be arranged in front if its X-minimum is smaller, but if the error of X-minimum of two loops is between − σ and + σ , need to rearrange the sequence of two loops according to Y-minimum of them, namely the loop should be arranged in front if its Y-minimum is smaller.

In the same way, arrange the sequence of all loops on the layer layer1. After have arranged sequence of loops, then the two loops with same number on their layer are the corresponding loops. In this way, the one-to-one corresponding relation loops between layer0 and layer1 have been put up in sequence. Arranging the sequence of loops according to not only the X-minimum but also the Y-minimum is for avoiding arrange falsely the sequence when coming up against the oddness shown in fig.4, the X-minimum and Y-minimum of the loop R00 and R01, two profile loops on the layer layer0, are XminR00、 Ymin0、 XminR01 and Ymin1, and the X-minimum and the Y-minimum of the loop R10 and R11, two profile loops on the layer layer1, are XminR10、 Ymin0、 XminR11 and Ymin1, they meet with the equation.(1).

$$\begin{cases} X_{minR00} > X_{minR01} \\ X_{minR11} > X_{minR10} \\ X_{minR00} - X_{minR01} < \sigma \\ X_{minR11} - X_{minR10} < \sigma \end{cases} \quad (1)$$

If arrange the sequence of loops only according to the X-minimum, the loop R01 will be arranged in the front of the loop R00 because of XminR00 being more than XminR01. When the corresponding relation of loops between the layer layer0 and layer1 is put up, the loop R01 will correspond to the loop R10 and the loop R00 will correspond to the loop R11. In this way, the error between the layer layer0 and layer1 will be consider mistakenly to exceed the error criterion. However, apparently the loop R00 corresponds to the loop R10 and the loop R01 corresponds to the loop R11, furthermore, the error between the layer layer0 and layer1 does not exceed the error criterion. In addition, if the error of X-minimum of two loops is less than - σ or more than + σ , then the error between the layer layer0 and layer1 exceeds the error criterion, so, even though arrange falsely the sequence,

because the two conclusions are both same, namely the error between the layer layer0 and layer1 exceed the error criterion, therefore, this case can be overlooked. As shown in fig.4, if arrange sequence of loops according to both the X-minimum and the Y-minimum, then loop R00 will correspond to loop R10 and the loop R01 will correspond to loop R11, it matches up to the fact.

4.3 Compare the maximum and minimum of the X-coordinates and Y-coordinates of the corresponding loops

Compare respectively the X-maximum, Y-maximum, X-minimum and Y-minimum of the two corresponding loops each other, if the anyone error is less than - σ or more than + σ , then consider the error between two profile loops exceed the plane error criterion.
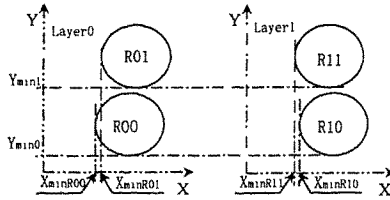
Fig. 4 Arrange sequence
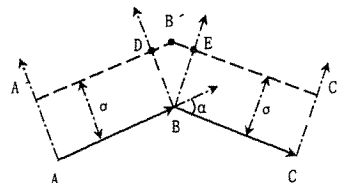
Fig. 5 Magnifying loop

4.4 Magnify and reduce the profile loop

Arrange the sequence of dots in the loop in one direction (clockwise or anti-clockwise). In this paper, adopt the clockwise to study. And for deflating loop, only need to arrange sequence of dots with anti-clockwise, then the method of deflating loop is same as the method of magnifying loop. Magnifying the profile loop is namely to move flatly the dots outside along the radial direction. In this paper, only illuminate how to move levelly two adjoining straight segment. Fig.5 marked dot A, B and C, the straight segment AB and BC move flatly with " σ " distant and reach the straight segment A´B´ and B´C´, the dot D and E are two dots on straight segment A´B´ and B´C´, the straight segment BD is vertical with straight segment A´B´ and AB, the straight segment BE is vertical with straight segment B´C´ and BC. Now only need to obtain the X-coordinates and Y-coordinates of the dot B´ , then the all dots in magnifying loop can be obtain with the same method. Enact the X-coordinates and Y-coordinates the dot A, B and C are respectively XA, YA, XB, YB, XC and YC, then the direction vector(2) of the straight segment AB and BC are respective the vector（XB－XA, YB－YA） and （XC－XB,YC－YB）, the outside normal vector of the straight segment AB and BC can obtain from circumrotating anti-clockwise the vector （XB－XA, YB－YA）and the vector（XC－XB,YC－YB）. The outside normal vector of the straight segment AB is :

$$[X_B-X_A \quad Y_B-Y_A] \begin{bmatrix} \cos(-90°) & \sin(-90°) \\ -\sin(-90°) & \cos(-90°) \end{bmatrix} = [(Y_B-Y_A) \quad -(X_B-X_A)]$$

The outside normal vector of the straight segment BC is :

$$[X_C - X_B \quad Y_C - Y_B] \begin{bmatrix} \cos(-90°) & \sin(-90°) \\ -\sin(-90°) & \cos(-90°) \end{bmatrix} = [(Y_C - Y_B) \quad -(X_C - X_B)]$$

Enact the coordinates of the dot D as coordinates (XD、YD),then,

$$(YD - YB)/(XD-XB) = (XA - XB)/(YB - YA) \qquad ------------- (2)$$

because the plane error criterion that is equal to the " σ ", the length of the straight segment BD should be equal to the " σ ", therefore the " σ " should match with equation.(3).

$$\sigma^2 = (YD - YB)^2 + (XD - XB)^2 \qquad ---------------------- (3)$$

Enact the math expression (YB − YA)/(XA − XB) being equal to "k", then the allied solution to the Eq. (2) and the equation. (3) is :

$$\begin{cases} X_D - X_B = \dfrac{\pm \sigma}{\sqrt{k^2 + 1}} \\[3mm] Y_D - Y_B = \dfrac{\pm k \sigma}{\sqrt{k^2 + 1}} \end{cases}$$

Because the vector (XD-XB, YD − YB) stands for the direction of the straight segment BD, and it is vertical with the straight segment AB, if XB is more than XA and YB is more than YA, then it is known from geometrical relation(3) that XD is less than XB and YD is more than YB, with the same reason, if XB is less than XA and YB is more than YA, then XD is less than XB and YD is less than YB, if XB is less than XA and YB is less than YA, then XD is more than XB and YD is less than YB, if XB is more than XA and YB is less than YA, then XD is more than XB and YD is more than YB. In this way, the mathematic sign of allied solution to the Equation. (2) and (3) can be ensured. In the same way, the coordinates of the dot E can be obtain. If the coordinates of the dot E is the coordinates (XE、YE). Because a straight-line can be confirmed from a dot and a vector, then the equation of the straight-line A´B´ and B´C´ are respectively the equation. (4) and (5).

$$y-YD = (YB - YA)/(XB - XA)(x-XD) \qquad ------------(4)$$
$$y-YE = (YC - YB)/(XC - XB)(x-XE) \qquad ------------(5)$$

If two straight-line equations are A1x+B1y+C1=0 and A2x+B2y+C2=0, then the coordinates of their intersecting dot will be as follows:

$$x0 = (B1 \times C2 - C1 \times B2)/(A1 \times B2 - B1 \times A2) \qquad ------------(6)$$
$$y0 = (C1 \times A2 - A1 \times C2)/(A1 \times B2 - B1 \times A2) \qquad ------------(7)$$

Enact the coordinates of the dot B´ as XB´, if XB´ is equal to x0 and YB´ is equal to y0, then the coefficient A1, B1, C1, A2, B2 and C2 match with equation.(8).

$$\begin{cases} A_1 = (Y_B - Y_A) / (X_B - X_A) \\ B_1 = -1 \\ C_1 = Y_D - X_D \times (Y_B - Y_A) / (X_B - X_A) \\ A_2 = (Y_C - Y_B) / (X_C - X_B) \\ B_2 = -1 \\ C_2 = Y_E - X_E \times (Y_C - Y_B) / (X_C - X_B) \end{cases} \qquad (8)$$

Now the coordinates of the dot B´ can be obtained from allying the Equation. (6), (7) and (8). In this way, the coordinates of other dots in the magnifying loop also can be obtained. In addition, if arrange the sequence of dots in the loop with anti-clockwise, then the coordinates of all dots in the deflating loop also can be obtained from the same method.

4.5 Compare the error of two corresponding loops with the plane error criterion

Magnify and deflate one of the two corresponding loops, and then analyze whether the all dots in the other one loop of two corresponding loops locate inside the region surrounded by magnifying loop and deflating loop. Generally adopt the method of checking up intersecting dot to judge whether a dot locates inside the region. Now to illuminate the method of checking up intersecting dot: Suppose that there is a dot, and its coordinates are the coordinates (x_0, y_0), make a horizontal shooting line that starts from (x_0, y_0) and its direction is same as the X-axes, then the equation of the shooting line is as follows.

$$\begin{cases} x=x_0+u & (u \geq 0) \\ y=y_0 \end{cases}$$

If the number of the intersecting dots of the shooting line with the loop is odd number, then the dot locates inside the loop, or else locates outside the loop. As shown in fig.6(a), the shooting line (a) intersects with the loop at two dots, and the shooting line (c) intersects with the loop four dots, then the dot A and C locate outside the loop. And the shooting line (b) intersects with the loop a dot, the shooting line (d) intersects with the loop three dots, then the dot B and D locate inside the loop. However, also need to deal especially with the shooting line that it passes through an apex dot of the loop. As shown in fig.6(b), the shooting line (f) passes through an apex dot of the loop and intersects with the straight segment (6) and (7), if the number of intersecting dots is counted up to two, then the dot F will be considered falsely to be outside the loop. To the dot E, if the number of intersecting dots is counted up to only one, then it will be considered falsely to be inside the loop. The right method is to analyze the relation to the shooting line and the two adjoining straight segment. When two adjoining straight segment locate on the same side of the shooting line, the number of intersecting dots should be counted up to two, or else should be counted up to one. Namely, if Y-coordinates of the other two apex dot, which are aside from the apex dot intersecting with the shooting line, on the two adjoining straight segment both are more (or less) than y0, then the two straight segment should be above (or under) the shooting line, and the number of intersecting dot should be counted up to two, or else it should be counted up to one.
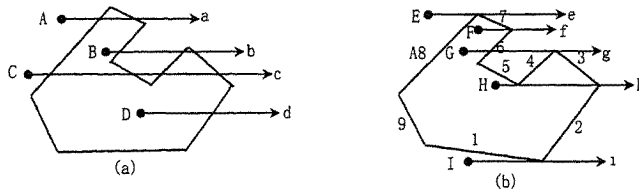

Fig. 6 Checking up the intersecting dot

In the same way, circulate comparing and analyzing the all other corresponding loops between two layers, if the error of anyone group corresponding loops exceeds the plane error criterion, then the error of the two layers will exceed the error criterion. Namely the current analyzing layer thickness is too thick. When the errors of all corresponding profile loops are less than the plane error criterion, add a jiggling layer thickness " δ " to the current analyzing layer thickness, and then circulate the above comparing and analyzing until the current analyzing layer thickness reach the maximal

thickness.

5 APPLYING

The self-adapting delaminating algorithm had been applied to SLS rapid prototyping system developed by Huazhong University of Science and Technology(HUST). Generally the 20~80 percent of the fabricating time can be save, its fabricating effect is not same to the different parts. Here give an example to show its fabricating effect, as shown in fig.7, it is a box, its length is 228. 5mm, its width is 214.4mm and its height is 425.5mm. Fabricate respectively the part with self-adapting delaminating method and traditional uniform delaminating method. In addition, fabricate it with different layer thickness when fabricating it with traditional uniform delaminating method. The fabricating data are shown at Table 1. It can be known

Fig 7 box

from the Table 1 that the self-adapting delaminating algorithm can save largely fabricating time, and the quality of part is the same as the traditional uniform delaminating with thin layer. Although the traditional uniform delaminating method can also reduce fabricating time by increasing layer thickness, but the quality of part is bad, namely will produce large staircase. This method is not able to ensure the precision of part.
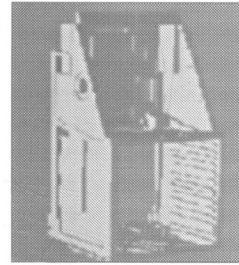
Table 1. Fabricating data

Delaminating style	Scanning speed (mm/s)	Laser power (%)	Layer thickness (mm)	Fabricating time (hour)	Quality of part
Uniform delaminating	2000	15	0.1	50	good
Uniform delaminating	2000	30	0.3	17	bad
Self-adapting delaminating	2000	Automatic adjust	Max: 0.4 / Min: 0.1	18	good

6 CONCLUSION

The self-adapting delaminating algorithm based on Profile Loop can not only be applied successfully to rapid prototyping system but also be fit for other layer manufacturing system, its principle are clarity, its oddness case is little and its operation efficiency is high.

ACKNOWLEDGEMENTS

REFERENCES

1 **Rendong Wu, Qingping Lu, Wei Lu, Yongnian Yan,** 1998, "Self-Adapting Delaminating Algorithm Based on Wavelet Analysis on Rapid Prototyping", China mechanical engineering, Vol. 11,No.11, pp.56 ~ 60

2 **Jiaguang Sun,** 1998, Computer Graphics, Tsinghua University Publishing, pp. 366 ~ 418, Sep.

3 **Xinxiang Sheng** , 1985 Handbook of Modern Mathematics in Science and Engineering, Huazhong University of Science and Technology Publishing, pp. 115 ~ 124, Jan.

Research and implementation of framework for selective laser sintering system of low cost

J XIE, Y SHI, S HUANG, and Z DUAN
Rapid Manufacturing Center, Huazhong University of Science and Technology, Wuhan, People's Republic of China

SYNOPSIS

At present, the cost of SLS system is higher, which embarrasses the development and application of SLS. For this reason, a low cost framework of SLS system, which is developed by Rapid Manufacturing Center of Huazhong University of Science and Technology (HUST), is presented in the paper. This SLS system has commercialized, and the practical application indicates that it can meet the requirements of industry.

1 INTRODUCTION

RP (Rapid Prototyping) has the following characteristics:

● The fabrication time is short. It only takes several hours transforming a CAD (Computer Aided Design) model to an entity. Compared with conventional manufacturing ways, RP can reduce the period of design and manufacturing and play an important role in developing new product.

● Design and manufacturing is a whole. Before, behind CAPP (Computer Aided Processing Plan) is a bottleneck in design and manufacturing. As to RP, CAD and CAM (Computer Aided Manufacture) can perfectly be combined. From design to manufacturing, RP is a closed-loop.

● Flexibility is high. Only CAD models are varied and parameters are renewed, different parts can be manufactured.

● Material is universality. Polymer, ceramic, metal, paper and their composite material can all be used to RP.

RP is novelty thought, combining computer technology, control technology, laser technology, material technology and mechanical technology. RP has developed quickly since 1980s. Until now, there are several kinds of commercial RP, such as LOM (Laminated Object Manufacturing), SLS (Selective Laser Sintering), SLA (Stereo Lithography Apparatus) and FDM (Fused Deposition Modeling).

Compared with the other types, the outstanding merit of SLS is universality of material. Theoretically, any powder materials that can be felted after heating up may act as SLS materials. Under the irradiation of laser, SLS material is melt and solidified. Furthermore, powder that isn't sintered can be used again.

Though there are many merits, the application of SLS isn't popularity because of its high cost. In order to drive the application of SLS, the Rapid Manufacturing Center of HUST successfully has developed a low cost SLS system named HRPS-III.

2 PRINCIPLE OF SLS SYSTEM

The principle of SLS system is shown in Figure 1. Firstly, SLS powder material is spread on the work platform of SLS system. Secondly, SLS material is warmed up to below the temperature of melting. Then, laser beam begins scanning the powder of entity according to the information of the contour under computer control. The temperature of powder in the sintering area rises to melting point because of absorbing in heat energy of the laser beam. Finally, the grain of powder begins melting, and adhering each other, contour of one layer is formed. While powder in the non-sintering area is still relax，acting as support of next layer. After one layer is finished, the work platform descends a certain altitude, and the above process is repeated until 3D part is formed.
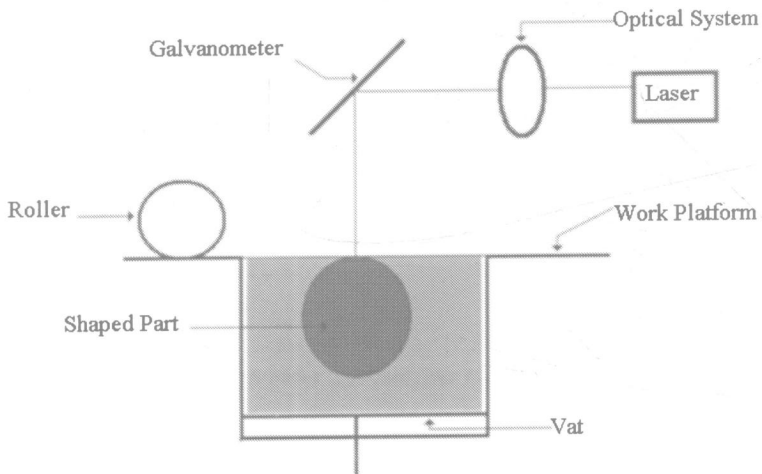

Figure 1. Principle of SLS system

3 FRAMEWORK OF SYSTEM

In order to reduce the cost of SLS system, an optimized framework is designed. A CNC (Computer Numerical Control) based on Windows without computer inside is developed for SLS system made by HUST.

3.1 A CNC without Computer inside

Logically speaking, the functions of hardware and software are same. That is to say, same function can be finished by the software or hardware. The commonness of software and hardware is higher complexity and fault. But, in the normal operation, the fault of hardware maintains a certain lever. When hardware reaches its life, the rate of fault will ascend. Unlike hardware, the rate of software fault goes zero with the conquering bug in the design. So, hardware decides the reliability of CNC. If the scale of hardware is compressed and the proportion of software is increased, the reliability of CNC will be improved accordingly.

In order to improve the reliability of SLS control system and to simplify its hardware configuration, a CNC without computer inside is designed (1). A CNC without computer inside doesn't mean having no a computer, it refers to the whole system only uses one computer.

The typical characteristic for CNC without computer inside is that every part has no absolute computer, no centralized control, software modules with independence are derived by data flow. After event or data flow is ready, the correlative software module outputs result without considering another data flow and modules. In this kind of configuration, software is the core of the control system and fulfills most functions, while hardware is compressed to the least scale, only I/O ports, such as on-off port and A/D port, are left. EPP (Enhancement Parallel Port) acts as I/O Bus. There is no relationship between the hardware and software.

Compared with multi–computers control mode, the framework of CNC without computer inside can realize the whole optimization, while multi–computers control only realizes local optimization.

The practice application indicates that HRPS-III, adopting the configuration of without Computer inside, has such merits as simple configuration, high reliability and quick respond.

3.2 Software IC

As above mentioned, software is the core of SLS system. At present, different developers develop different SLS system software because of no uniform standard. What is worst, all kinds of software have no interconvertibility and compatibility. The repeat of design wastes manpower and money, it isn't good to generalize RP obviously.

So, the thought of Software IC (Integrated Circuit) is brought forward (2). Software IC is also

called software module, it is an entity designed through the technology of modern soft engineering, such as category, abstract, encapsulation and inheritance. Software IC materializes the thought of reuse of software.

After analyzing the CNC of RP, several categories are divided. Same or similar parts belong to a correlative category, like hardware IC. Every piece of software IC possesses independency, transplant, group and extension, and all of software IC attribute to a library of CNC. Only taking out correlative software IC from the library, a new CNC of RP is built, such as a CNC of SLS, without developing system from the start.

Summing up the general characteristics of current RP, a software IC library of RP is shown in Figure 2.

- Software IC for Man-machine interface. The IC answers for debugging, monitoring and setting parameters.

- Software ICs for data processing. The part includes STL (STereoLithography Interface Specification) data processing software IC and direct slicing software IC. On the one hand, the format of STL, which seems a kind of standard in RP field, is known well by customers and applied in many fields. On the other hand, STL simulates raw model through triangle, so there is error between the raw model and STL model. Direct slicing format, which can get over the above-mentioned error (3), may improve the precision of work piece, decrease the workload, and reduce the cost, so it is also applied in HRPS-III.

- Software ICs for RP process optimization (4). The ICs include the fabrication direction, scanning route, fabrication parameters (such as laser power, scanning speed, scanning spacing, layer thickness of powder, laser spot size and so on). Specially, in order to improve the efficiency, a new subarea direction change scanning mode is put forward and applied in HRPS-III. The new scanning mode can avoid inner hole and groove of a section through plotting the scanning line. The practical application indicates that the new scanning mode can improve the fabrication efficiency and decrease the distortion.

- Software IC for format transform. This part answers for realizing the transforms between different formats and produces different tracks according to the demands of RP process.

- Software ICs for control. This part includes laser power control, scanning control and temperature control, which control hardware through the port. The laser scanning IC answers for 3D Galvanometer's scanning system, which makes the laser beam change direction. The laser control IC answers for adjusting the laser power. The temperature control IC is used to monitoring the pre-heating temperature during SLS.

The practical application indicates that the control system that is built with the thought of software IC does not only possess the advantages of easily transplanting, grouping and extending, but also the advantage of introducing the third side software, which may reduce
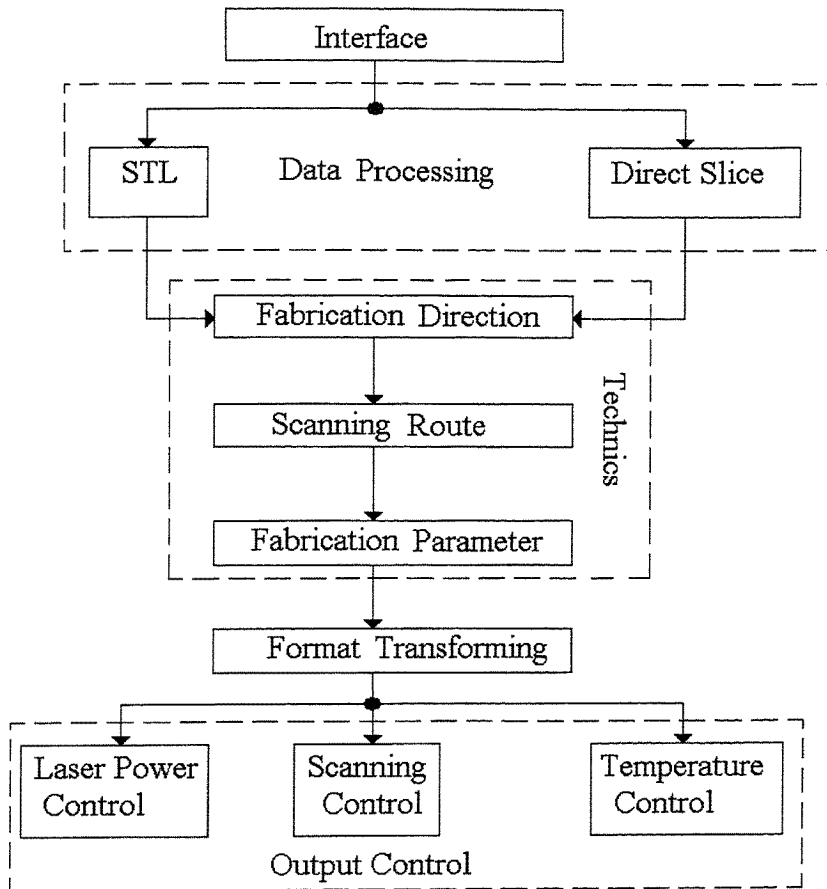
cost and avoid re-development.

Figure 2.   A software IC library of RP

3.3 The realization of CNC based on Windows

Windows 9x is applied in many fields because of its abundant GUI (Graphics User Interface) and multi–tasks operation.

Compared with CNC based on DOS, CNC based on Windows 9x has higher efficiency and

friendly interface. However, Windows isn't a real operation system. Furthermore, in order to ensure system's safe, Microsoft Company adopts the strategy of shielding bottom to prohibit customers to directly use hardware interrupt and to visit memory. How to get over this problem is a key to develop CNC based on Windows 9x.

Through developing VxD (Virtual Device Driver), the problem on efficiently driving hardware is solved. VxD is a tache to link application program and bottom facility, is a 32-bit drive program with protection mode, manages hardware device and software source, and lets multi- application programs simultaneity employ the source. Through VxD, operation system is a virtual physics device, intercepts requests of application program to hardware, and provides application program with a virtual circumstance connected with hardware. Under the multi-task circumstance, different applications share the same physics sources and realize multi-thread simultaneously running

In a sense, VxD can finish anything. VxD doesn't occupancy conventional memory because of its running under protection mode, Moreover, there is no mode shift derived from real mode. Generally speaking, the speed of VxD is higher than DOS's. What is more important, VxD based on Windows 9x is dynamically loaded and unloaded, and realizes plug and play.

In HRPS-III, VToolsD for Windows 9x developed by Vireo Company is used, The VToolsD possesses a interface with C++, encapsulates VMM (Virtual Machine Manager) and VPICD (Virtual Programmable Interrupt Controller Device), provides categories and clear structure flow.

CNC based on Windows has the following advantages:

- The rate of performance and value can be improved. Windows possesses friendly GUI, perfect memory management, prominent 32-bit software circumstance and flexible compatibility of hardware. Developer uses it not only as a visual system but also as a multi-function platform to realize image processing, the assignment and assembly of hardware and software.

- Remote control can be realized. Based on Windows, it is easy to realize remote manufacturing and communication among devices and enterprises utilizing network. Customers apply advanced manufacturing technology in time.

- An open framework of CNC can be established. Windows is regarded as a normal application plat, which has widely accepted any interfaces and protocols and helps customer to establish more open framework of CNC.

- The close state of CNC can be changed. Compared with special hardware or based on DOS, CNC based on Windows is more flexible and extensible. At the same time, productive enterprises, software enterprises and developers are easy to be combined to learn from strong points to offset weaknesses one another.

4 CONCLUSIONS

Because of adopting classic framework and novel technologies, the rate of performance and value of HRPS-III is best of all in the similar products.

HRPS—III has the following specialties:

- Low cost. Under the precondition of realizing good performance and high reliability, the price of HRPS-III is low because of adopting a CNC based on Windows without computer inside.

- Bigger build chamber dimensions. The workspace of HRPS-III is up to 400 W × 400 D × 500 H mm^3.

- Smaller system dimensions. The outline dimensions of HRPS-III (double powder cartridges delivering from the top) is only 1080 W × 1270 D × 1850 H mm^3.

- Higher fabrication efficiency. Because double powder cartridges delivering from the top (self-determination patent right) are used, HRPS-III only takes 4s to spread a layer of SLS powder. In addition, the distortion of SLS part is decreased through adopting the subarea direction change scanning mode.

The outline dimensions of HRPS-III are shown Figure 3, and the main parameters of HRPS-III are shown in Table 1.

The dimensions of testing part are shown in Figure 4, and the test results of testing part are shown in Table 2.

HRPS-III has been put in to market for two year, its success proves adopting advanced technologies can realize high performance and low cost of SLS system.

REFERENCES

1 **Qing Zhong, Ji Li, Xuebin Chen,Shuhui Huang**. (1999) A CNC System without Computer inside for Laser Rapid Prototyping Machine. J. Huazhong Univ. of Sci. &Tech. Vol.27, No.12, pp. 63-65.

2 **Lenz ManFred.** (1987) IEEE SOFTWARE.Vol.4, No.4, pp.34-42.

3 **Ailin Liu, Yuejia Xiao, Han Ming, Lichao Zhang, Shuhuai Huang**. (2000) Description of CAD Solid Model in RPM and Error-Tolerant Slicing. China Mechanical Engineering.

Vol.11, No.4, pp.427-429.

4 Ian Gibson, Dongping Shi. (1997) Material properties and fabrication parameters in selective laser sintering process. Rapid Prototyping Journal. Vol.3, No.4, pp.129-136.

ACKNOWLEDGEMENTS

APPENDIX
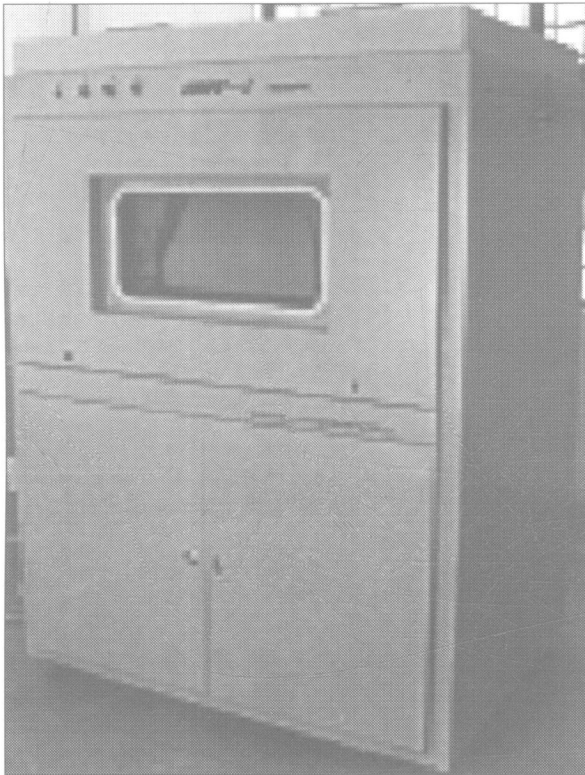
Figure 3. Outline of HRPS—III type SLS system

Table 1. Main parameters of HRPS—III

Chamber dimensions (mm)	400 W × 400 D × 500 H
Laser	CO_2
Wave length (um)	10.6
Power (w)	50
Scanning system	Dynamic focusing scanning
Spot size (mm)	< 0.35
Scanning speed (m/s)	4
Layer thickness (mm)	0.075 ~ 0.3
Spreading powder time (s)	4
Pointing accuracy (um)	12
Warm-up power (w)	0 ~ 3000 (adjust)
Application software	HRPS-STL and HRPS-DSlice
CAD format	STL and PIC
Outline dimensions (mm)	1080 W × 1270 D × 1850 H
Circumstance	$10\,^{o}C \sim 28\,^{o}C$, RH < 60%
Power supply	220 VAC, 7 KVA, 50 KHZ
Delivering powder mode	Double powder cartridges delivering from the top
Precision (mm)	± 0.25
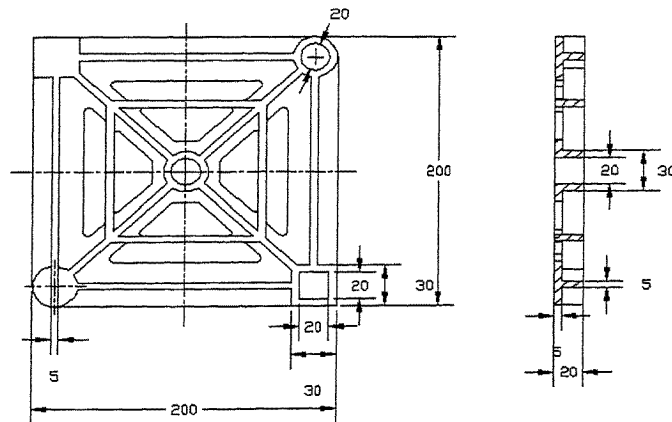
Figure 4. Testing part

Table 2. Test results of testing part
(precision: ±0.25mm)

Title	Size (mm)	Direction	Results (mm)			Average	Average error
			First	Second	Third		
Length	100	X	100.01	99.94	99.96	99.97	0.03
Width	100	Y	100.06	99.96	100.02	100.01	0.01
Height	10	Z	10.05	10.06	10.06	10.06	0.06
Cliff thickness	2.5		2.54	2.52	2.40	2.49	0.01
			2.38	2.54	2.30	2.41	0.09
Inner diameter of corner circle	10		10.10	10.06	10.14	10.10	0.10
Inner diameter of center circle	10		10.06	10.14	10.12	10.12	0.12
Outer diameter of center circle	15		15.06	15.08	14.98	15.04	0.04
Inner side of corner square hole	10	X	10.02	10.04	9.96	10.01	0.01
		Y	10.10	10.08	9.96	10.05	0.05
Outer diameter of corner square hole	15	X	14.96	14.98	15.02	14.99	0.01
		Y	15.00	14.98	14.98	14.99	0.01
Soleplate thickness	2.5	Z	2.56	2.48	2.52	2.52	0.02

The research of the SLS process optimization based on the hybrid of neural network and expert system

Y SHI, J LIU, D CAI, and **S HUANG**
Rapid Manufacturing Center, Huazhong University of Science and Technology, Wuhan, People's Republic of China

ABSTRACT

In order to optimize SLS rapid prototyping process parameters, neural network and expert system are combined. According to the features of SLS process, a BP network model is chosen for SLS process parameters optimization, and the detailed design and discussion of its structure as well as related parameters choice are carried out The successfully developed Expert System based on it has been used in the automatic optimization of shrinkage compensation coefficient of SLS and has received satisfying result.

1 INTRODUCTION

The SLS may directly produces three-dimension parts with solid powder, which is not confined by complexity. It does not need any die&mould and directly changes the CAD model into prototyping parts. Theoretically, any powder which can melt after being heated can be used as the SLS material, including polymer, ceramic, metal powder and their mixture. Because of the unique advantages in the diversification of the SLS material, such as saving material, wide use, fabricating without support, high intensity and rapid fabricating metal parts, the SLS can be widely used and attracts attention from all walks of industry.

The SLS process has not only physical effects but also chemical changes. So if satisfying parts are wanted, the best prototyping parameters must be needed. However, There are still not a systematical theory and scientific method for selecting SLS parameters which are chosen mainly through experience at the present time, which doesn't meet the needs of SLS powder material fabrication. So, in order to change the present situation, new theory and method must be found to make sure that the high precision and high intensity parts can be produced directly in SLS process.

Artificial neural network, which has characters of human intelligence and does not need any precise mathematical model, is suitable for the characters of SLS (strongly depending on experience and lacking information). Artificial neural network after successfully training can rapidly reason and associate, and filtrate the inputted noise. But its operation is in the dark and cannot explain the process of reasoning Expert system has specific knowledge to denote and can explain the process of reasoning But it cannot learn from examples and hard to obtain knowledge Therefore, the neural network and expert system are combined to optimize the SLS process parameters, which will make the problems difficultly to solve or even to be solved in traditional method be easily done away

2 THE PRINCIPLE OF NEURAL NETWORK EXPERT SYSTEM FOR SLS PARAMETERS OPTIMIZATION

The neural network expert system for SLS parameters optimization is mainly composed of neural network reasoning organ, S/D convertor, D/S convertor, explaining mechanism and repository(1). As is showed in Figure 1, neural network reasoning organ is the core of the system. Compared with the traditional expert system, it is the repository which stores and manages knowledge and is the reasoning organ which can reason too. S/D convertor is the input convention. It changes the information inputted by users and experts into numerical value model which can be identified by neural network.
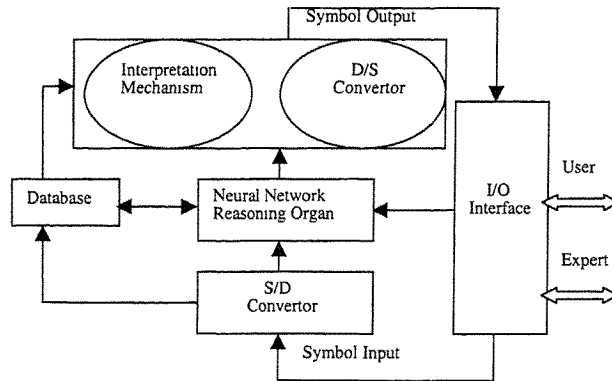
Fig. 1 The structure of the neural network expert system

The database includes neural network algorithms database, sample database and interpretation database. The neural network algorithms database supports the neural network reasoning organ. The sample database is used to store training samples. Interpretation database is the interpretation language that experts dynamically produce for users according to neural network training condition. The system operation is divided into two phases: representative sample training and system application. The former one is operated by experts. First, objects are analyzed, and the I/O mode of neural network, the number of I/O nerve cells and its network structure are decided. The optimized parameters are output mode. Known information which can be easily decided (such as requirements for technique and material) is input mode. Second, the data from experience are used to create training sample database and the neural network training then is carried out. During the training, its weights and thresholds are adjusted according to the errors between the actual output mode and the ideal one of the neural network, until it gets less than the expected minimum and the system becomes stable.
Then the matrix contains the knowledge from the experts. Last, the actual output mode is analyzed and then interpretation database is established to offer users interpretation language. After the training of the representative sample, users can use the system.

The input values are provided by users. After being computed and reasoned, the output values (optimized process parameters) and corresponding interpretation language are gotten. When the input values provided by the users are the same as the those from the training samples, the output values received after reasoning by neural network will be equal to the those from the

training samples. When the input values are different from the those from all the training samples, the users can usually also get the output values, which are equal to the those from the closest training samples. When the input values and all the input values from training samples are very different, the output values will not be gotten. Then the sample is trained as a new training sample and the neural network studies through increment, which will make the weights and thresholds of the neural network be changed and it get more knowledge which will be stored into the neural network. Thereby, the ability of the system processing knowledge can be enhanced. The system operation flow is shown in figure 2.
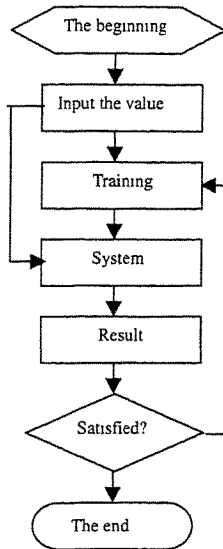
Fig. 2 The flow of the neural

3 NEURAL NETWORK EXPERT SYSTEM FOR OPTINIZING SLS PROCESS PARAMETERS AND ITS REALIZATION

3.1 The structure of neural network expert system
3.1.1 Database module
The databases include neural network function database and algorithms database. The real time established databases are: neural model database, weight database, threshold database and interpretation database(1). They are shown in figure 3.

Network model database

| The layers of network | The number of nerve cells in each layer | Nerve cells transfer function | Arithmetic parameters |

Weight database

| Each layer's grade in the network | Each layer's weight | Change of the weight in each layer |

Training sample database

| The number of samples | The input value of each sample | The ideal output value of each sample | The actual output value of each sample | The extremum of the input value |

Interpretation database

| The regulations from experts | The Interpretation language |

Testing sample database

| The number of samples | The input value of each sample | The output Value of each sample |

Threshold database

| Each layer's grade in the network | Each layer's threshold | Change of the threshold in each layer |

Fig. 3 The structures of the real-time databases

The key of the system is obtaining weight database and threshold database, which can also be stored as the inhere database of the system in order to study through increment or to be applied. The neural network function database includes liner function and S function. The algorithms include normal BP arithmetic, BP arithmetic improved by "addition momentum", BP arithmetic improved by adapted learning rate and their combination.

3.1.2 The neural network design module
It decides the parameters, the number of layers, the number of never cells in each layer and the function of output layer.

3.1.3 The neural network training module
An interface of the training sample database and training sample data with the neural network is set up. According to the network trained by the training sample database, the data from weight database and threshold database can be real-time updated and tracked during training. When the error varies strongly or remains unchanged, the training can be stopped immediately. When the training is finished, the relation between the ideal output values of the sample and the actual output values will be analyzed and then the interpretation database will be created.

3.1.4 Application module
It provides the interface of testing data to neural network and uses the well-trained network to calculate the testing data. Then the output values of the network will be gotten and the testing

sample database will be created. The interpretation database uses the expert principles to reason, judge and interpret the calculation results. If the testing sample database is correct after testified by practicality, it can be used as training sample database and let neural network to carry out increment learning.

3.2 The realization of neural network expert system
3.2.1 The design of network
3.2.1.1 The design of network structure
First, the structure type of the network and its structure parameters will be chosen. Three-layer BP network structure is recommended. If it is not satisfying, four-layer one will be tried. The output layer transfer function includes S function and liner function. They are decided by the threshold range of the output values. If the output values are between 0 and 1, S function is chosen; if not, liner function is chosen(1).

3.2.1.2 The design of network parameters
The design of network parameters includes: the maximum cycle times of the arithmetic iterative operation, the goal error, the additional momentum gene, the learning rate, the error change rate, the learning rate increment, and the learning rate decrement. The computer program provides the beginning values of the BP network parameters, which can be changed by the users if needed. If the normal BP network needs to be created , the momentum gene should be 0, and both the learning rate increment and the learning decrement rate should be 1. If BP network with additional momentum is adopted, the momentum gene should not be 0, and both the learning rate increment and the learning rate decrement should be 1; if BP network with adaptive learning rate is adopted, the momentum gene should be 0, and both the learning rate increment and the learning rate decrement should not be 1. When the "enter" button on the interface of the computer is pressed, the system will call the corresponding function to automatically create the desired BP network according to the parameters chose by the users.

3.2.2 Network training
The existed database may be used or a new database is created by inputting new data(2). After inputting data, the network training can begin. When something abnormal happens, the training can be immediately stopped and the network will be redesigned. Why the training stops automatically? There are two reasons: one is the network output error has reached the goal error; the other is that it hasn't reached the goal error, but the maximum cycle time has been reached. If the reason is former, the training output values, weights and thresholds will be written into the raining sample database. Then the ideal outputs and the actual outputs will be analyzed to create the interpretation database. The limit of the input value is set by the system automatically without precondition and interpretation. If the reason is later, the training sample data will be inputted again to continue training.

3.2.3 Operation flow chart of neural network expert system
The operation process of the neutral network expert system is shown in Figure 4. Each database is independent and can be written and read at any moment. The system can work in sequence or can operate each module separately. The discontinuous network training can be realized. A well-trained network can be applied time after time.
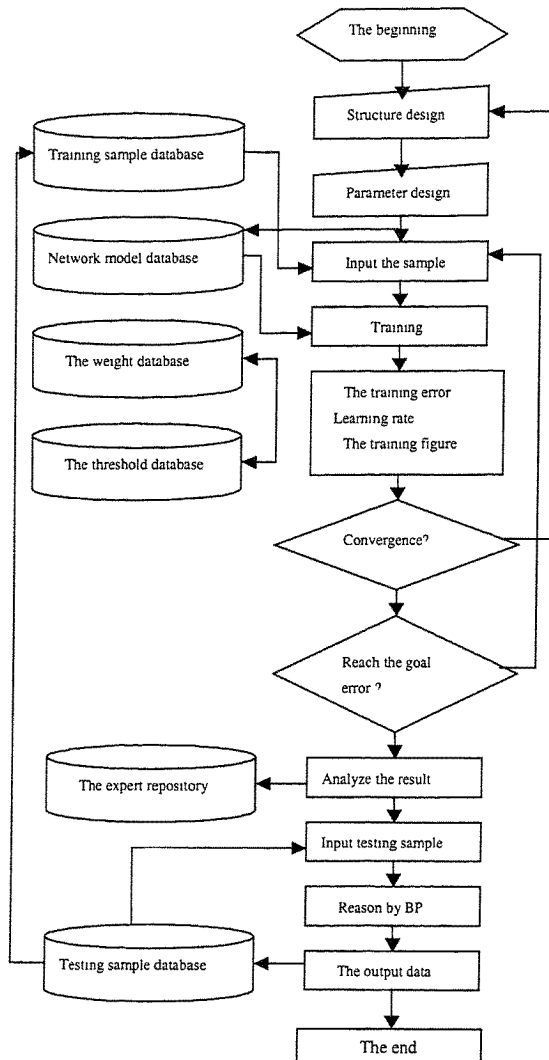
Fig. 4 The operation flow of neural network expert system

4   AN APPLICATUION SAMPLE – THE SELF-ADAPTING ADJUSTMENT OF COMPENSATION COEFFICIENT DURING SLS

During the SLS, the material has experienced melting, cooling, and solidification, which will result in phase change of SLS material. So the material will shrink or inflate, which has great influence on the precision of SLS parts. Therefore before SLS parts being fabricated, the academic values in X, Y, Z directions must be multiplied by the compensation coefficient to compensate the error caused by the above –mentioned shrinkage (compensation coefficient= relative shrinkage rate +1)(3)

According to the results of experimental research, the factors which affect the shrinkage and inflation of the SLS parts are the choice of SLS parameters and the effect from the material performance If the SLS material is decided, the compensation coefficients are mainly influenced by the laser power, the scanning speed, the scanning width and so on Through experiments, the relative shrinkage rate of the standard testing parts can be obtained then the training sample database to train the neural network and the neural network expert system can be created After the SLS material is selected, the scanning power, the scanning speed and the scanning width are also decided, the compensation coefficient will be modified automatically by neural network expert system instead of experts Thereby, the compensation coefficient can be optimized automatically

4.1 Creating of training sample database
The input values of the network are the scanning power, the scanning speed and the scanning width The output value is relative shrinkage rate S function is chosen to enhance tolerance of the network Samples from experiments are used to create training sample database Some training samples and the output values of the neural network are shown in Table 1

Table 1.   Some training samples and their neural network outputs

Serial number	Input mode			Ideal output	Actual output	Maximum error of the ideal output and the actual output	Compensation coefficient calculated by neural network expert system
	Laser power (w)	Scanning speed (m/s)	Scanning width (mm)	Comparative Shrinkage ratex10	Comparative Shrinkage ratex10		
1	12 5	1 5	0 1	0 244	0 24556	0 00487	1 024556
2	13	1 5	0 1	0 252	0 2518		1 02518
3	14	1 5	0 1	0 267	0 26462		1 026462
4	15	1 5	0 1	0 28	0 27795		1 027795
5	16	1 5	0 1	0 295	0 29181		1 029181
6	17	1 5	0 1	0 31	0 3062		1 03062
7	18	1 5	0 1	0 322	0 32103		1 032103
8	19	1 5	0 1	0 338	0 33613		1 033613
9	20	1 5	0 1	0 35	0 35127		1 035127
10	21	1 5	0 1	0 366	0 3661		1 03661
11	22	1 5	0 1	0 381	0 38024		1 038024
12	22 5	1 5	0 1	0 387	0 38693		1 038693
13	15	0 9	0 1	0 3183	0 31343		1 031343
14	15	0 95	0 1	0 3124	0 30952		1 030952
15	15	1	0 1	0 307	0 30569		1 030569
16	15	1 05	0 1	0 3021	0 30197		1 030197
17	15	1 1	0 1	0 2977	0 29841		1 029841
18	15	1 15	0 1	0 2938	0 29504		1 029504
19	15	1 2	0 1	0 2904	0 29186		1 029186
20	15	1 25	0 1	0 2875	0 28891		1 028891
21	15	1 3	0 1	0 2851	0 28621		1 028621
22	15	1 35	0 1	0 2832	0 28375		1 028375
23	15	1 4	0 1	0 2818	0 28155		1 028155
24	15	1 45	0 1	0 2809	0 27962		1 027962
25	15	1 5	0 05	0 1	0 10052		1 010052
26	15	1 5	0 06	0 14	0 1377		1 01377
27	15	1 5	0 07	0 183	0 1824		1 01824
28	15	1 5	0 08	0 225	0 2262		1 02262
29	15	1 5	0 09	0 259	0 25973		1 025973
30	15	1 5	0 11	0 283	0 28116		1 028116
31	15	1 5	0 12	0 273	0 27282		1 027282
32	15	1 5	0 13	0 256	0 25724		1 025724
33	15	1 5	0 14	0 237	0 23826		1 023826
34	15	1 5	0 15	0 22	0 2187		1 02187

4.2 Design and training of the neural network

There are three variables in the input mode and there is one variable in the output mode, so the number of nerve cells in the input mode and the output mode is 3 and 1 separately. First a hidden layer is selected and then the number of nerve cell in it is decided. S function is chosen as the output layer function. The training times of the network are shown in Table 2 under different nerve cell numbers in hidden layer when the learning rate is 0.01 (the momentum gene, the learning rate increment, the learning rate decrement, the error change rate are separately these beginning values: 0.98, 1.05, 070, 1.04) and the goal error is 0.0001. It may be known from Table 2 that the training cycle times are less than the others when the nerve cell numbers in the hidden layer are 3,7,or 5. When the nerve cell numbers in the hidden layer are 3, the less training time is needed. However, 2 nodes are added to enhance tolerance of the network. So the nerve cell numbers in the hidden layer become 5.In order to increase the precision of the network, the goal error is reduced to 0.00005.When the training time reaches 41801, it takes 3 minutes. Then the training is finished and the goal error is reached. After training, the absolute maximum error between the ideal compensation coefficient and the actual one is 0.000568.The knowledge from the expert is stored in the trained network model database, weight database (see in Table 3) and threshold database (see in Table 4).Both the ideal output values and the actual ones is between 0.1 and 0.4.When the actual output values exceed the above-mentioned bound, it means that the chosen process parameters are incorrect, according to which an interpretation database will be created.

4.3 Application

For the HB1 type polymer powder material developed by HUST, when the laser power is 15w, the scanning width d_{sp} is 0.1mm and the scanning speed is 1.5m/s, the compensation coefficients reasoned and calculated by the neural network expert system in X and Y directions are all 1.027795, and the compensation coefficient in Z direction is 1. According the above-mentioned compensation coefficients, the SLS testing part successfully fabricated in HRPS-III type SLS machine developed by HUST is shown in figure 6. The practical testing results show that its absolute precision is ±0.23mm.When the users want to use the new material to fabricate SLS parts, they only need to put the data into the training sample database and read the stored network model database, the weight database and the threshold database to carry through increment learning and to train again.

Table 2. Results of various neural network design plans

Number of nerve cells in hidden layer	Learning rate		Training times
	Original value	Final value	
3	0.01	1.00	5452
4	0 01	1.05	9092
5	0.01	0.84	4427
6	0.01	0.58	11624
7	0.01	0.60	5628
8	0.01	0.5	23903
12	0.01	1.07	47437

Table 3. Weight matrix

		Hidden layer				
		1	2	3	4	5
Input layer	1	1.400479	-1.0231	2.163521	0.88093	-0.53337
	2	-0.00813	-0.01336	-1.27966	-0.11663	-0.24153
	3	2.052011	-2.29291	-1.20839	-4.88407	2.594516
Output layer		1.048048	-0.98311	1.241064	-3.16552	-2.28741

Table 4. Threshold matrix

Hidden layer					Output layer
1	2	3	4	5	
-0.35183	1.008248	0.022014	-0.36869	1.542074	0.82820

5. CONCLUSION

It can be seen from the above-mentioned research and practical application, the hybrid of neural network and expert system gains success and shows its unique advantage. For expert system, it is easy to obtain knowledge through the learning of neural network. And for neural network, its pure numerical value handling is clearer through expert system. The technique and system mentioned in this paper not only are suitable for the automatic optimization of the SLS parameters but also can be extended to other related fields. So it has the characters of universality, practicability and potential. The neural network expert system mentioned in this paper has the following characters: A. The neural network expert system combines expert system and neural network, so it can learn from other's strong points to offset one's weakness. B. When optimizing the SLS parameters, it does not need any mathematical models and has no strict requirements for the input parameters, which can offset the weakness such as lacking information and uncertainty during the SLS process. C. The time spent in experiment and feeling for the parameters can be remarkably reduced. In addition, the manufacture cost can also be reduced and the efficiency can be raised.
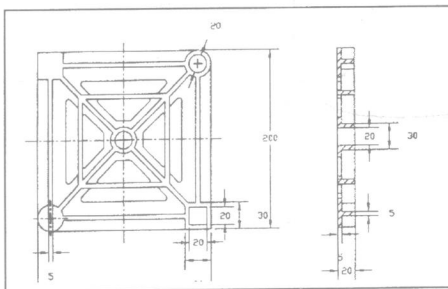
Fig. 5 The size of the SLS testing part
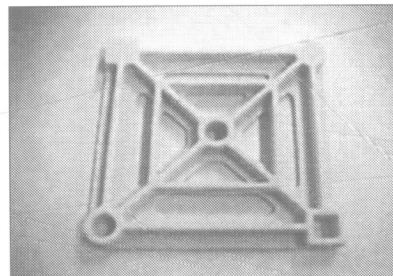
Fig. 6 The SLS testing part fabricated through optimizing the compensation coefficient

REFERENCES

1 **Chuanyou Pan.** The research of the SLS process parameters optimization based on the hybrid of neural network and expert system.[thesis for master degree],Wuhan,HUST,2001.6
2 **Xiangsheng Li.** The research of the SLS key technique[thesis for doctor degree],Wuhan,HUST,2001.7
3 **Yusheng Shi.** Modeling method for weight-on-bit optimizing model based on artificial neural network: EARTH SCIENCE—JOURNAL OF CHINA UNIVERSITY OF GEOSCIENCES, 1999, 4（24）431~435.

The research of the biomaterials' rapid forming machines

Y YAN, R WU, L CHEN, and W ZHENG
The Department of Mechanical Engineering, Tsinghua University, Beijing, China

SYNOPSIS

A rapid forming machine was developed to form materials with bio- activity and bio-degradability into scaffolds based on low temperature jetting/extrusion deposition process. The machine used a PMAC NC card and servomotors to realize the point-to-point and contour controls in the X/Y plane, and an icebox and 4 nozzles were integrated into the machine to attain the layer-accumulated manufacturing. The 4 nozzles all have their own material feeding system respectively, and jet/extrude different kind of materials with different methods. The samples of large segmental bones produced by the machine were practiced on dogs and rabbits, repaired their bone damnification.
Keywords: Rapid Forming, Biomaterials, Tissue Engineering, Numeric Control, jetting/extrusion technique

1 INTRODUCTION

Tissue engineering is the leading field of replacement and repair science of human organs. Based on application engineering and life science, tissue engineering develops biology replacements which are used to repair, maintain and promote the function of the damnified organs and apparatus. The replacement has three essentials, including target cell, scaffold and growth factors. In these essentials the scaffold has biomimetic live structure which has complex material gradient and pore gradient, and formed by biomaterials. This structure can be obtained by the method of manufacture process. The principle of rapid prototyping is adding material forming method based on dispersion-accumulation, so this technique can make multi-materials scaffold with intricate interior structure, and at the same time can achieve high-degree customized forming process according to every individual patient. Tsinghua University has developed a new rapid forming based technique to make large segmental bone. This paper talks about the design and planning of the biomaterial RP forming equipment based on low-temperature jetting/extrusion adding forming. This machine is used as research platform for actual application.

2 CELL SCAFFOLD BIOMATERIAL AND ITS RAPID FORMING PROCESS

Cell scaffold has many requirements for biomaterial, including biocompability, biodegradation, innocuity, pore rate, mechanical capability, control-release performance, and so on. Currently the typical scaffold materials include abiomaterial, bio-macromolecular material, and bio/abio composite material, etc. These materials have diverse usage, some used as scaffold material to provide intensity, some used as bio-activity material to derive the growth of organ, some used as growth factor to promote the growth of cell, some used as control-release material to control the release of growth factor, and some used to meet the special requirement of certain forming process. All these materials have their own live temperature scale and phrase transition characteristics; especially in liquid condition they have different viscosity. So in the forming process multi-nozzles must be used, and different material use different nozzle to carry out accumulation on the matrix.

In the process of jetting/extrusion accumulation forming technique the jetting/extrusion units and the method of slicing-adding to forming parts are similar with that of MEM technique. The principle process is:

First use CAD software to design the structure of tissue engineering scaffold, get the slice forming data by slicing software; next choice the needed nozzle, under the control of computer, the forming machine do XY plane scan movement, and at the same time the material is extruded and accumulated on the matrix. This process can integrate multi-materials together, and keep the forming environment under the certain low temperature, so the material can freeze simultaneously when it is jetted/extruded onto the matrix or latest slice, and stick to the matrix or the latest slice, slice by slice, the extruded materials add to be formed as the needed structure of scaffold. Last after finishing the forming process, the structure formed under low temperature need some disposal such as freeze-desiccation, then we can get solid scaffold which has certain structure and certain intensity under common temperature. After biochemical disposal, the scaffold can be used in animal test. .

3 SYSTEM DESIGN OF FORMING MACHINE

The forming machine can realize the forming process of accumulating biomaterial. The Requirement of process to equipment can be included as following several points:

(1) For different materials different jetting/extrusion methods are needed to enable the export of materials. Usually the structure forming materials are liquid with high viscosity before extrusion, so screw pump nozzle which can offer high pressure is needed in order to ensure stable extrusion and instant switch response; the functional assistant materials typically are liquid with low viscosity, then electromagnetism valve or piezoelectricity valve with low pressure can be used to realize jetting. The whole enable system is assembled to be a multi-nozzle system.

(2) The requirement of numeric control: the accumulating process of biomaterials is the same as that of MEM technique. The differences between the two processes meanly focus on control mode. When jetting biomaterials, since there are several materials, so during forming one slice, contour scan mode and point-to-point control mode are necessary at the same time. The two modes can realize the forming of the structure materials and functional assistant materials;

(3) The requirement of environment temperature. In the forming process of low temperature jetting/extrusion, it is essential to freeze and felt the materials together immediately after been extruded/jetted from nozzles, so there need to keep certain low temperature within the

forming environment. During the forming process, it need not only assure the nozzles cannot be jammed by the solidification of material, but also keep the bio-activity of the materials unchanged under the temperature of the nozzles, so we must control the temperature of jetting/extrusion system strictly.

(4) The requirement of lustration. In order to ensure the biochemical characters such as bio-activity and biocompatibility, it is essential to keep the parts contacting with material directly lustrating. And at the same time, these parts should be airproofed and can be easily unpicked and washed and replaced.
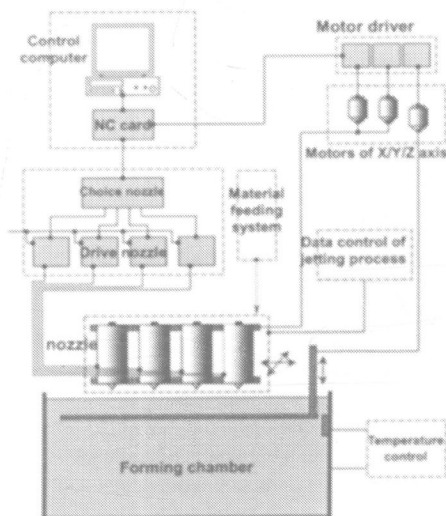
Figure 1 the design of biomaterial RP forming machine system

The system design of the forming machine is shown as above figure 1. The module of numeric control and path scan translate the control information into plane scan movement, and match the movement with the switch of nozzle, harmonize the producing and accumulating of slices; The module of jetting/extrusion carries out the producing process of slices。 This module includes material container, enable mechanism of implementing jetting/extrusion, temperature controller, pressure generation device, motherboard for installing multi-nozzles, nozzle drive device, choice circuit, and so on. The module of forming environment ensure the process parameters be required during the forming course. This module includes forming platform, icebox, recycle container for protecting liquid; The module of mechanical structure includes the mechanics of forming machine. The design of mechanics module can follow that of MEM forming machine, and the same as other accessories such as ventilation and illumination.

4 DESIGN AND REALIZATION OF MAIN MODULE

(1) Numeric control module. The industrial control computer and PMAC numeric control card compose the kernel of control system. For X and Y axis the servo motors are used since the plane scan movement require the motion of X and Y axis simultaneously and in phase, and with high precision. The movement speed of Z-axes is comparatively low, so it uses stepper motor. The movement system uses linear-motion guide and ball screw. The function is

realized by the following step: first carry out sequence control of circulation course of "plane manufacturing-descend Z-axes" , and implement accumulating forming process; next provide the two control mode of X and Y axis movement, which are couture scan control mode(the mode for forming structure materials and point-to-point control mode(for adding growth factor); then output the choice real time (or beyond) signal for choosing multi-nozzles during plane scan motion course; Last the control system should get offset value for every nozzle by means of testing relative position of every nozzle before beginning forming process. Swapping nozzles during forming process causes the offset.

After assembling the forming machine, we tested the motion system. The motion speed can be up to 70mm/s, which meet not only current but also future requirement of the process.

(2) Enable module. The module can assemble four nozzles together, including two nozzles based on screw pump which can provide high pressure and extrude high viscosity materials continuously and two nozzles based on electromagnetism valve and piezoelectricity crystal which jet low viscosity liquid by point-to-point mode. Every nozzle has its own close material container, material feeding channel, pressure generation device and temperature controller. And at the same time every nozzle is provided with its own drive module, and can implement jetting/extrusion action following the choice signals output by NC module. The method to control the multi-nozzles and their scan motion is shown as figure 2.
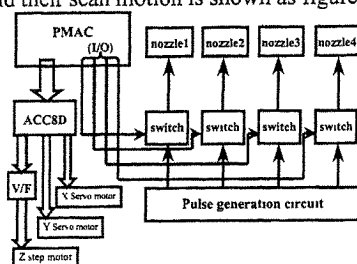
Figure 2 the principle of controlling multi-nozzles

The same requirement of nozzles is export materials by time, quantificational and directional way, and the nozzles should have well response character. The means exporting materials are different from each other because the materials have different characters. The nozzle 1 and nozzle 2 adopt screw pump mode to extrude materials. In this way, filament materials, granule materials and liquid materials can be used. This kind of nozzle is mainly used to extrude high viscosity materials. The nozzle 3 bases on electromagnetism principle, and mainly fit to jet low viscosity liquid, such as water, etc. The nozzle 4 bases on piezoelectricity principle, and serve as important role for jetting costly low viscosity liquid materials, such as growth factors(for example, bone morphogenetic protein). The nozzle system is shown as figure 3.

1: ball screw and LM guide;
2: groupware of forming platform
3: material feeding part;
4: nozzles
5: the scaffold to be formed
6: icebox.
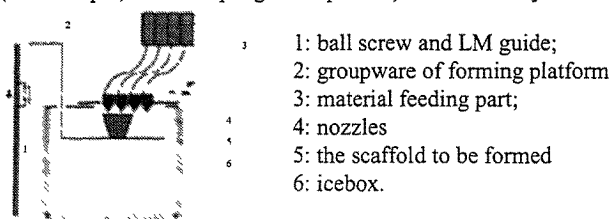
Figure 3 The nozzle system of TissForm machine

(3) Forming environment module. The main aim of this module is to keep the low temperature and lustration needed by forming process. The applied scheme is to integrate icebox into the

forming machine, and make the icebox as forming chamber. The top of icebox is open, forming platform appear as L shape, and extend into the icebox from the open top of icebox together with nozzles. The open top is insulated by adiabatic soft membrane. The merit is that the temperature can be controlled, and easily adjusted, and need comparatively less man work. The temperature on the forming platform should be fallen down to -30℃, which is needed by the process of forming bio-activity materials.

We had made a TissForm machine shown as figure 3, and developed control software according to the machine.

Figure 3 TissForm biomaterial forming machine

Table 1 shows TissForm machine's parameters.

name	TissForm		
Forming material	Biocompatible materials		
Number and type of nozzles	Screw pump	Electromagnetism valve	Piezoelectricity crystal
	2	1	1
NC card	American Del ton company Pmac NC card		
Environment	-30℃—30℃		
Forming space	200*200*200 mm^3		
Scan speed	70 mm/s		

5 THE CONCLUSION OF USING THE TISSFORM MACHINE

We had formed large segmental bone cell scaffold by the TissForm machine, and the samples are shown as following figure 4. The pictures show that the sample had got favorable pore rate and needed pore structure.

（TCP+PGA）   （HA+PLLA）
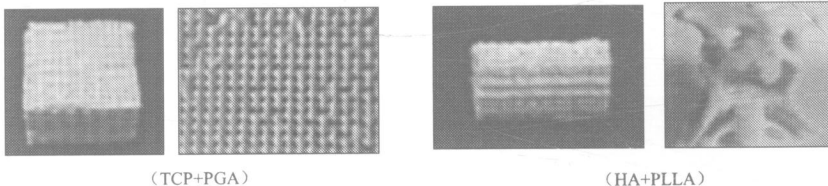
Figure 4 large segmental bone sample formed by TissForm machine

Figure 5 shows X pictures of the dog's radio, which had been planted biomimetic large segmental bones make by TissForm. The pictures indicate that after 24 weeks of planted the large segmental bone the dog recovered completely.
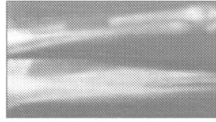
Figure 5 the X pictures of dog's radio before and 24 weeks after planted bio-activity large segmental bone

6 SUMMARY

This paper designed a kind of biomaterial RP forming machine based on the manufacturing idea of dispersal/accumulation, and the machine can be used to form bio-activity materials such as tissue engineering cell scaffold. Those characters, such as four jetting/extrusion nozzles, low temperature forming chamber, high scan speed, and so on, not only meet the process requirements of forming large segmental bone, but also leave comparatively large space for the development of RP forming biomaterials. The Tissform machine realized the automation and celerity of forming biomaterials, and at the same time it is new development of RP machines.

REFERENCES

1 Xiongzhuo, Yongnian Yan,, etc. The two new kinds of RP techniques for forming bone tissue engineering cell scaffold Chinese mechanical engineering. 2001, 12(5): 515-518
2 Wangli , Yongnian Yan, etc. analysis and realization of bone tissue engineering cell scaffold, material guide. 2001, 15(11): 49-51

Numerical simulation of direct metal laser sintering process

W JIANG
School of Mechanical Engineering, Dalian University of Technology, People's Republic of China
K W DALGARNO and **T H C CHILDS**
School of Mechanical Engineering, University of Leeds, UK

ABSTRACT

Direct metal laser sintering process (DMLS) uses high power laser beam to selectively fuse fully metallic powder into metal components or tooling. It has potential to become an indispensable industrial tool. However, the characteristic of local energy input will inevitably cause inhomogeneous temperature distribution, residual stresses and deformations. Knowledge of thermal history of DMLS is critical to develop control schemes for acceptable geometry and properties of components. In this research, the governing principle of DMLS is modeled as a non-linear heat conduction problem with a moving heat source. Temperature-dependent thermal properties and latent heat are taken into account. The combination of FEM and temperature recovery method is applied to solve the problem. The simulation results reveal uneven transient temperature distribution, density profile and asymmetric geometry of sintered parts in DMLS.

NOTATION

C	specific heat (J/kg K)	X	solid function
h	heat transfer coefficient	$\varepsilon, \varepsilon_{powder}$	porosity of sintered part and powder bed respectively
k	thermal conductivity (variable) (W/m K)		
k_s	thermal conductivity of solid material (W/mK)	$\rho, \rho_{powder}, \rho_s$	density of material (variable), powder bed and solid (kg/m^3)
L	latent heat (kJ/kg)	$\rho_{average}$	average density in sintered layer i
q	laser heat flux (W/mm^2)		
T_a	ambient temperature (°C)	$\Delta h_i, dh_i$	thickness of powder layer i before and after sintering (mm)
T_b	powder bed preheating temperature (°C)		
v	equivalent laser blade scan speed (mm/s)	$[C], [K]$	capacitance, conductance matrix
V	unit volume in material	$\{F\}$	heat flux vector

1   INTRODUCTION

Direct metal laser sintering process (DMLS) uses a high power laser beam to directly fuse fully metallic powders to produce metal components or tooling. It has capability of reducing cycle time in product development, lowering cost, and building very complex geometries, and has potential to become an indispensable industrial tool in rapid fabrication of low-volume production of metal components or tooling[1,2].

However, the characteristic of local energy input will inevitably cause inhomogeneous temperature distribution, which will result in thermal stresses, deformations and cracking. Understanding temperature history of DMLS is critical for determination of quantitative relationships between process parameters and geometrical and mechanical properties of sintered parts, and for development of intelligent process control.

Currently modeling work mainly concentrated on amorphous polymers[3-7]. This paper reports numerical modeling of direct metal laser sintering. The governing principle is modeled as a non-linear heat conduction problem with a moving heat source. Temperature dependent thermal properties and latent heat due to phase change are taken into account. The combination of finite element method and temperature recovery method is used to solve the problem. The model is applied to predict important process properties that may affect quality of 316 stainless steel sintering, such as temperature, density and geometry. The simulation results reveal uneven transient temperature distribution, density distribution and asymmetric geometry of sintered parts. These results will be used as input to mechanical model to predict residual stresses and deformations. The purpose of this work is to develop a process simulation to optimize DMLS and to control it intelligently.

2   SIMULATION METHOD

Direct metal laser sintering is a very complicated process. Firstly, metallic powders undergo a phase change from solid to liquid and then back to solid, which is associated with latent heat. Secondly, thermal properties, such as thermal conductivity and specific heat are assumed to be temperature dependent. Thirdly, boundary conditions are very complex. All these make the problem highly non-linear. The combination of finite element method and temperature recovery method is applied to solve the problem.

2.1 Heat conduction equation

During DMLS process, the penetration of laser heat flux into porously packed powder bed leads to heat flows, which can be assumed to follow heat conduction. Fig.1(a) shows schematic views of DMLS. The laser beam of power P, diameter d, scans at speed u in ±X direction, over current width w of layer, and rasters with spacing s in +Y direction. In model application, the above heat flux is described as a blade aligned with XY axes, d wide and w long, moving along +Y direction with speed v, with Gaussian distribution. The axes are fixed in the powder bed. The general heat conduction equation with a moving heat source for

temperature dependent thermal properties is given as [8,9]

$$\rho C \frac{\partial T}{\partial t} = k \left(\frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + \frac{\partial k}{\partial T} \left\{ \left(\frac{\partial T}{\partial y} \right)^2 + \left(\frac{\partial T}{\partial z} \right)^2 \right\} + \left(\frac{\partial k}{\partial y} \frac{\partial T}{\partial y} \right) + \left(\frac{\partial k}{\partial z} \frac{\partial T}{\partial z} \right) - \rho C v \frac{\partial T}{\partial y} \quad (1)$$

(a)   (b)
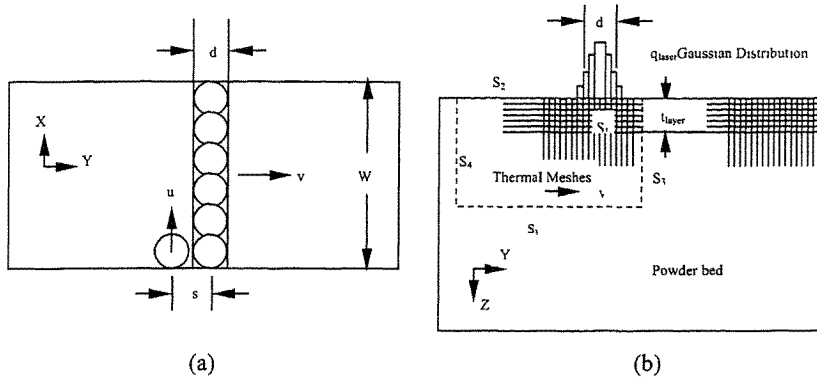
Fig.1 Schematic views of DMLS process (a) and finite element mesh (b)

2.2 Finite element model

2.2.1 Meshes

Two kinds of meshes are used in the model, thermal mesh and powder bed mesh. The thermal mesh is discrete and composed of four-node brick elements, as shown in Fig.1(b). Each four-node brick is regarded as a pair of three-node triangles. In the neighborhood of heat source q, thermal mesh is relatively fine: its y-spacing is 1/6 of beam diameter d and its z-spacing is 1/6 of powder layer thickness t_{layer}. It coarsens with distance from the heat source in multiples of this fine mesh size until in the corner diagonally opposite the heat source, one element of thermal mesh is 40 beam diameters and three powder layers in y and z respectively. Powder bed is also divided into a mesh structure. The powder bed is the same size as the fine thermal mesh, namely d/6 in y and $t_{layer}/6$ in z. Linear interpolations are used to map between the powder bed and the thermal meshes. The thermal meshes superimpose on the powder bed, moving at v in Y direction by d/6 in each time step.

2.2.2 Material properties

Thermal conductivity, specific heat, and density are needed to determine global temperature field from Eq.(1). Specific heat C and thermal conductivity k_s of solid 316 stainless steel are assumed to vary with temperature, as shown in Fig.2. Thermal conductivity variable k is a function of temperature and local bed porosity ε. The relationship of thermal conductivity variable k with density (or porosity ε) is shown in Fig.3. The fact that thermal conductivity increases with decreasing ε is very important to DMLS because it permits energy to be relatively easy to be conducted through fully sintered material to raise temperature of liquid-powder interface, thereby promoting rapid sintering at interface.
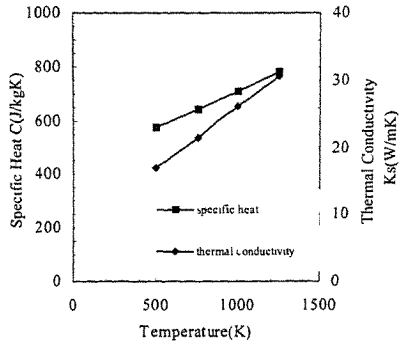
Fig.2 Temperature dependent
thermal properties[10]
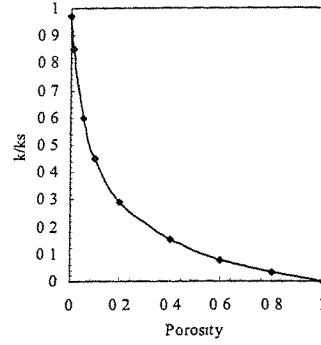
Fig.3 Density dependent
thermal conductivity[11]

2.2.3 Boundary conditions

The boundaries of thermal region are divided into four types, S_1, S_2, S_3, S_4, as shown in Fig.1(b). Either temperature or heat flow rates q_s are controlled on these boundaries.

$$k\left(\frac{\partial T}{\partial y}l_y + \frac{\partial T}{\partial z}l_z\right) + q_s = 0 \qquad (2)$$

here l_y and l_z are direction cosines of outward normal to the surface. On surface S_1, q_s follows Gaussian distribution. And on surface S_2, q_s is due to cooling by combination of radiation and convection as laser beam passes, follows $q_s = -h(T - T_a)$. The boundary temperature on surface S_3 is T_b, while on surface S_4 the boundary condition follows $\frac{\partial T}{\partial y} = 0$.

2.3 Densification and shrinkage

The liquid-phase sintering law used to describe sintering rate for metal is as follows[11]:

$$1 - \frac{\varepsilon}{\varepsilon_{powder}} = \left[\tanh(5 \times (1 - X))\right]^4 \qquad (3)$$

while $\qquad \varepsilon = \frac{\rho_s - \rho}{\rho_s} \qquad\qquad \varepsilon_{powder} = \frac{\rho_s - \rho_{powder}}{\rho_s}$

X change from 1 to 0, linearly with temperature as temperature changes from solidus to liquidus temperature.

The powder bed, which includes unsintered powder, a liquid pool and sintered region, has an irregular shape since the upper surface of the powder bed recedes due to shrinkage that occurs in the sintering process. For simplification, mass conservation principle is employed to deal with shrinkage calculation[7]. The shrinkage that occurs in the sintered part is assumed to occur only in the Z direction, as follows

$$\rho_{average}dh_i = \rho_{powder}\Delta h_i \qquad (4)$$

2.4 Solution strategy

The governing equation of heat conduction problem in matrix form can be obtained through usual finite element discretization procedure,

$$[K]\{T\}+[C]\left\{\overset{\cdot}{T}\right\}=\{F\} \qquad (5)$$

$\{T\}$, $\{\overset{\cdot}{T}\}$ are nodal temperature vectors and time derivative of temperature vector respectively. It has been noted that Eq.(5) is highly non-linear since the matrices [K], [C] and $\{F\}$ are strongly dependent on temperature, especially when several different phases are involved at same time. Therefore, time-march scheme is of great importance in terms of stability, accuracy and efficiency. By time-marching scheme, we obtained

$$\left[\frac{1}{\Delta t}[\overline{C}]+\theta[\overline{K}]\right]\{T\}_{t+\Delta t}=\left[\frac{1}{\Delta t}[\overline{C}]-(1-\theta)[\overline{K}]\right]\{T\}_t+\{F\} \qquad (6)$$

where $[\overline{K}]=(1-\theta)[K]_t+\theta[K]_{t+\Delta t}$; $[\overline{C}]=(1-\theta)[C]_t+\theta[C]_{t+\Delta t}$

When temperature of material crosses melting point, latent heat due to melting or solidifying is needed. In order to take account of latent heat, temperature recovery method is used[12]. The calculated nodal temperature $T_{t+\Delta t}$ from Eq.(6) is modified in the phase change process by

$$\int_{T_t}^{T_{t+\Delta t}}\rho CVdT=\int_{T_t}^{T_{t+\Delta t}}\rho CVdT-\int_{\xi_t}^{\xi_{t+\Delta t}}\rho LVd\xi \qquad (7)$$

Where $T'_{t+\Delta t}$ is modified temperature, ξ_t and $\xi_{t+\Delta t}$ are volume fraction of solid phase at time t and t+Δt, respectively. The calculation iterates until temperature field converges.

2.5 Simulation parameters

Table 1. Simulation parameters applied in the numerical calculation

Laser power P(watt)	500	Latent heat L(kJ/kg)	308
Scan speed u(mm/s)	100	Liquidus temperature T_m(°C)	1380
Scan spacing s(mm)	0.5	Solidus temperature T_{solid}(°C)	1280
Laser spot diameter d(mm)	1.1	Powder bed temperature T_b(°C)	20
Energy absorptivity	0.92	Powder bed density ρ_b(kg/m^3)	4710
Layer thickness t_{layer}(mm)	0.5	Solid density ρ_{max}(kg/m^3)	7850
Vector length w(mm)	10	Weighted parameter of FEM θ	0.875
Layer length(y direction) (mm)	5.5		

Parameters used in the simulation are summarized in Table1. k is function of both temperature

and porosity, C is function of temperature, as shown in Fig.2, Fig.3. These parameters are recomputed at each time step of calculation.

3   RESULTS AND DISCUSSION

3.1 Temperature profile

(a) Temperature vs distance   (b) Temperature vs time
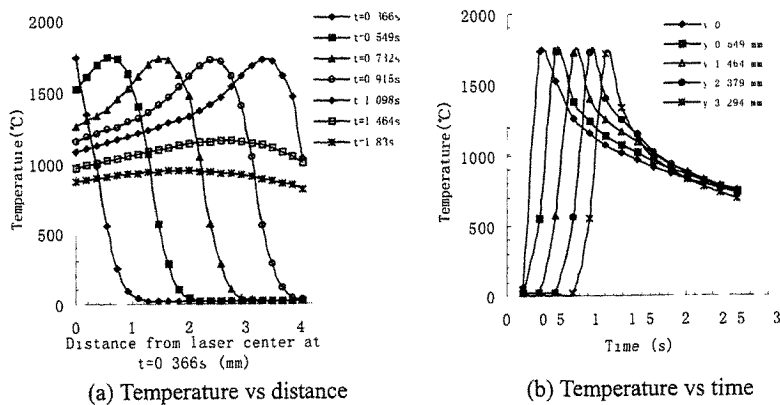
Fig.4 Temperature simulation results

Fig.4 shows transient temperature distribution at various positions and time. The temperature field is a function of both position and time. The severe temperature gradient during heating and significant values of temperatures reached in the course of analysis can clearly be noted. Also neighboring nodes undergo different heating or cooling histories and some nodes in the heating phase (expansion) are next to nodes that are in the cooling phase (contraction).   This fact determines a complicate origin of residual stresses.

3.2 Density

Fig.5 show a steep change in bed density from unsintered region to sintered region, and a very uniform high density in the sintered region. In the sintered region, i.e. in the first layer, within 5mm from the laser center of sintering start point, uniform density is $7849kg/m^3$, while in the region of 5~9mm, solid density reduce to powder density. This is because every position in the sintered region experiences similar temperature history, temperature rise almost instantly to melting point near laser center then gradually cool down as laser move forward. Metal powders undergo a phase change from solid to liquid and then back to solid, thus arriving at uniform high density. The laser finishes sintering at 5.5mm. Because of boundary effect and heat diffusion, metal powders are melted only partially, and a steep change in bed density from sintered region to unsintered region occurs. As less heat is conducted to the 2nd or 3rd layer, less powder are melted at these layers. A wider range of fully sintered part of the 2nd or 3rd layer can be obtained by optimizing powder properties, process parameters and material combination.

3.3 Geometry
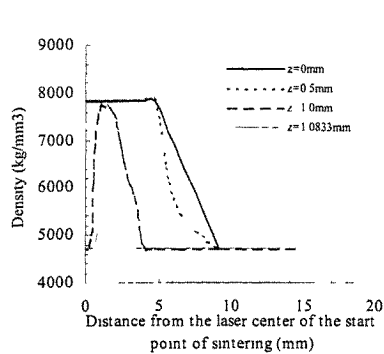
Fig. 5 Density distribution
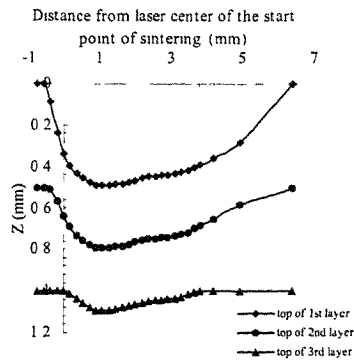
Fig.6 Sintered geometry

Fig.6 shows the geometric shape after sintering. The sintered geometry is asymmetric, which is approved by sintering experiment. In the first layer, at 1.1 mm from the laser center of sintering start point, the maximum sintered depth is 0.49 mm, sintered width is 6.8mm. At 2nd layer and onward, less heat is conducted, and smaller sintered region achieved. The fully sintered geometry can be used to establish a model for FE residual stresses analysis.

4   CONCLUSION

DMLS of 316 stainless steel has been investigated numerically for a non-linear heat conduction equation with a moving heat source. Temperature dependent thermal properties and latent heat due to phase change are taken into account. The combination of finite element method and temperature recovery method is used to solve the problem. With material properties and process parameters as input, the temperature and density distribution and sintered geometry are predicted. From simulation results, we can conclude:

1.   Temperature field in the sintered region is varied with position and time. Every position experiences similar temperature history. Also neighboring nodes undergo different heating or cooling histories and some nodes in the heating phase are next to nodes in the cooling phase.

2.   Since sintered region experiences similar temperature history, metal powders fully melt and solidify at different times. In most of the sintered region, uniform high density is achieved, while steep density change occurs near the boundary.

3.   Asymmetric geometric shape will be obtained after sintering. The sintered depth is greater near laser center of sintering start point, and smaller as the distance increases.

The simulation results reveal uneven transient temperature and density distribution, and asymmetric geometry of sintered part in direct metal laser sintering process. The model will be useful for building residual stress analysis model and for setting operating parameters, and especially useful for developing a fully automatic, non-feedback control system that could continuously control sintering properties of metal components.

REFERENCES

1 **Beaman, J.J., Barlow, J.W., Bourell, D.L., Crawford, R.H., Marcus, H.L., McAlea, K.P.** Solid freeform fabrication. 1997 Kluwer Academic Publishers, Dordrecht.

2 **Jacobs, P.F.** Stereolithography and other RP&M technologies. 1996 ASME Press, New York.

3 **Nelson, J.C., Xue, S.,Barlow, J.W.,Beaman, J.J., Marcus, H.L., Bourell, D.L.** (1993) Model of the selective laser sintering of bisphenol-A polycarbonate. *Ind Eng. Chem. Res.*1993,32, pp2305-2317

4 **Weissman, E.M. and Hsu, M.B.** (1991) A Finite Element Model of Multi-layer Laser Sintered part. In Solid Freeform Fabrication Symposium, pp.86-93.

5 **Sun, M-S.M. and Beaman, J.J.** (1991) A three dimensional model for selective laser sintering. In Solid Freeform Fabrication Symposium, pp.102-109.

6 **Bugeda,G., Cervera,M. and Lombera,G.**(1999) Numerical prediction of temperature and density distribution in selective laser sintering process. *Rapid Prototyping Journal*, vol.5, No.1, pp.21-26

7 **Childs, T.H.C., Berzins,M., Ryder,G.R. and Tontowi, A.E.** (1999) Selective laser sintering of an amorphous polymer-simulations and experiments. *Proc. IMechE,* 213(B) pp.333~349.

8 **Carslaw,H.S. and Jaeger, J.C.** Conduction of heat in Solids. 2nd Edition, Oxford University Press, New York, 1959

9 **Rosenthal,D.** (1946) The theory of moving sources of heat and its application to metal treatments. *Transactions of ASME*, Nov. pp.849~866.

10 Metals handbook, Volume 3, Properties and selection· stainless steel, tool materials and special-purpose metals, Metals Park, Ohio: American Society for Metals, 1961

11 **Childs,T.H.C., Hauser, C., Taylor, C.M. and Tontowi, A.E.**(2000) Simulation and experimental verification of crystalline polymer and direct metal selective laser sintering In Solid Freeform Fabrication Symposium, pp.100-109.

12 **Tszeng,T.C., Im,Y.T. and Kobayashi,S.**(1989) Thermal analysis of solidification by the temperature recovery method. *Int J Mach Tools Manufact.*,29,pp.107-120

A new fused deposition rapid prototyping machine

S ZHANG and **G LIU**
Institute of Advanced Manufacturing Technology, Tongji University, Shanghai, China

SYNOPSIS

A new concept of FDM with screw extrusion principle is discussed in this paper. By using hydrokinetics principle and experiment data, a new screw extrusion device has been designed and produced. The characteristics of fused deposition process, integrated influence of maximum press, press sensitivity and consistency, temperature adaptability, machine reliability and cost was considered during the new machine design stages. The productivity of the new FDM machine with screw extrusion was reached 5 times higher than existing FDM process.

1 INTRODUCTION

Fused deposition is one of the popular RP technologies. There are many types of fused deposition RP systems in the market, for example, FDM series (Stratasys Inc., USA), MEM series (Tsinghua University, China) and so on. These RP systems allow the use of non-toxic ABS polymer, polycarbonate and other materials to build strong and durable, fully functional prototypes and testing parts. But these systems are faced by a common problem of low flow quantity (usually less than $0.1 cm^3/s$) of material extruded from the nozzle, thereby the efficiency of building a part from them takes a relative long time, and it is improper to build middle or large-size parts by using them.

Due to this reason, many researchers are presently exploring the possibility for the solving the problem by using a screw extrusion device. A new rapid prototyping machine with screw extrusion process has been developed by the authors. The tested results of the new machine are remarkable: simple structure, easy operation, low cost and higher productivity.

2 PROCESS OF THE FUSED DEPOSITION TECHNOLOGY

The process of establishing parts by using fused deposition technology is to translate viscous state material into a hypostatic part composed of many small threadlike "roads" via automatic controlled extrusion equipment. These threadlike "roads" are described by a digital file created on computer. When the solid-state fused material is fed into the heater, it is heated and extruded from the nozzle. Here, the solid-state fused material has been turned into threads, and then the thread-like fused material is deposited on the floating soleplate as shown in Fig.1.
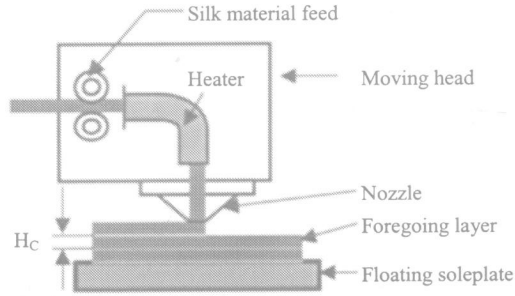
Fig. 1   Operational process of the deposition technology

A quantity of heat from the extruded thread-like fused material will be partly conducted to the foregoing layer. Between layers, diffused motion of molecules will come into being, and the fused material will be felted on the foregoing layer. Accordingly, the final part structure is holistic. The x, y, z directions' position of the nozzle in relation to the floating soleplate is controlled by computer, whereupon complicated-geometry prototypes can be built. Precision of prototype dimensions may reach ± 0.2mm.

2 OBJECTIVES OF NEW MACHINE DESIGN

Objectives of new fused deposition machine design are:

1)   To resolve conflicting requirements for improvement performance of FDM technology, increase the rate that material is extruded from a nozzle so that parts can be built more quickly;

2)   To enable an increase in the viscosity of the extrudate so as to produce physical parts with more desirable mechanical properties, decrease the cross-sectional area of the extrudate to achieve better feature resolution;

3)   The combination of high required pressure, high rates of material flow, and high rates of change of pressure are not adequately served by the existing art;

4)   To better utilize the innate capability of extrusion-based layered manufacturing in locally varying the composition of the material being deposited;

5)   To provide a system that performs layered manufacturing which functions without significant operator assistance in typical engineering environments. This includes the ability to monitor and modify processing characteristics, simple and automatic handling of the generated parts.

3 PERFORMANCE OF THE MACHINE

The design of an extruder for high speed layered manufacturing is prescribed by peak

pressure capability, pressure agility, pressure uniformity, temperature compatibility, reliability, and cost.

Peak Pressure could be calculated as follows: Viscosities (η) of interesting extrusion materials vary from 10^{-6} kg.s/cm^2 for waxes and liquid metals to about 0.01 kg.s/cm^2 for engineering polymers. Nozzle orifice inner diameters (d, r=d/2) of interest range from 0.08 mm to 0.8 mm. The maximum linear velocity (V) for the nozzle while it is depositing extrudate, with a cross-section approximately equal to the nozzle orifice, varies from 5 mm/s to about 500 mm/s. Assuming that the length (L) of the orifice is typically twice the orifice's inner diameter, Poisseuille's formula for pressure drop in a pipe can be rewritten to estimate the pressure drop associated with impelling an extrudate through a nozzle orifice:

$$\Delta p = \frac{8L\eta Q}{\pi r^4} = \frac{8 \times 4r \times \eta \times \pi r^2 V}{\pi r^4} = \frac{64\eta V}{d} \tag{1}$$

At the limit of low speed, low viscosity, and large orifice size, the required operating pressure is about 0.0004 kg/cm^2 above ambient; at the limit of high speed, high viscosity, and small orifice size, the required operating pressure is 4000 kg/cm^2 above ambient.

Pressure Agility: Mechanical constraints require that a gantry slow its velocity along straight line trajectories as it approaches non co-linear intersections of straight line trajectories. The time constraint that parts should be built as quickly as possible requires that the gantry speed up as it leaves such an intersection. As a result, the magnitude of the instantaneous vector velocity of the nozzle with respect to the supporting base changes rapidly.

Pressure Uniformity: Variations in the extrusion rate, apart from those required to build the part, cause observable fluctuations in the surface of a final part. Variability of 1 percent or less is typically required to make the extrusion appear uniform.

In order to satisfy these conflicting requirements, as well as to make a low mass extruder that is inexpensive to manufacture and is reliable in use, the pressurization process performed by the extruder is split into two stages. Each stage is obliged to optimize for only some of the criteria. A stage of pressurization is a mechanical sub-unit through which the extrusion material flows which generates a pressure at its output that can be higher than the pressure at its input.

In a screw extrusion device, the first stage of pressurization increases the absolute pressure experienced by the solid wire like material from an initial pressure to an intermediate pressure. The initial pressure is usually ambient atmospheric pressure. The intermediate pressure is a level between the initial pressure and the peak pressure required to impel the fused material out of the nozzle. The intermediate pressure is that required to impel the solid wire like material into the inlet of a second stage of pressurization at a maximum required volumetric deposition rate. More specifically, the intermediate pressure assures an adequate flow of the solid wire like material to the second stage of pressurization under all expected

volumetric deposition rates and assures a non-interrupted flow of the fused material out of the nozzle. The second stage of pressurization includes a screw rod and barrel pair which increases the absolute pressure experienced by the fused material from the intermediate pressure to whatever instantaneous pressure is required to impel the fused material out of the nozzle at a predetermined rate, up to the maximum required volumetric deposition rate.

4 DESIGN OF THE SCREW ROD

In fused extrusion systems, the screw extrusion device is mainly composed of three parts: screw rod and barrel; nozzle and control system. The control system controls the motor to drive the screw rod according to rotating speed needed, so fused material under control of a heating and temperature-controlling equipment can be continuously and stably extruded from the nozzle. Capability of each part in the screw extrusion device may directly influence quality of prototypes, but screw rod design is a pivotal link. If parameters and structure of a screw rod are optimised, productivity of extruder may be increased and quality of prototypes may be improved. Since extrudate has higher viscosity, slot screw rod should be selected and used.

4.1 Screw Rod's Parameters

In a screw extrusion device, many parameters may affect screw performance, including screw rod own parameters (D、L_1、L_2、L_3、H_1、H_3、Φ), processing parameters, gap δ as shown in Fig. 2, and these parameters influence each other. Therefore, while a screw rod is being designed with the help of the extrusion theory, design experiences and empirical formula and data must be also adopted.
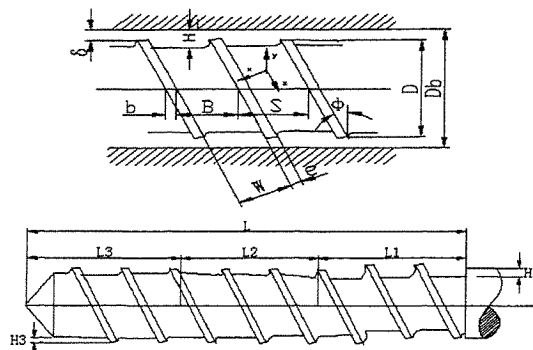
Fig. 2   Geometry parameters of the screw rod

In virtue of flow quantity formula [1] of a screw extrusion device, adverse and leak flow quantity are ignored. If screw lead S equals to D, lead angle Φ equals to 17°42. Therefore, the screw rod diameter D can be decided.

Span of a screw rod can be divided into three sections: loading section L_1, compression section L_2 and metering section L_3. Compression section is defined as the section from slot depth H_1 of loading section to slot depth H_3 of metering section, and its function is to increase press of extrudate at the nozzle exit, and its length is factitiously established (see Fig. 2). It is important to decide length of three sections, and it has close relations with extrusion technology conditions and polymer performances.

Leak quantity is directly proportional to the cube of the gap between screw rod and barrel in terms of analysis of hydrokinetics [2]. Thereby if the gap is too large, productivity of the screw extrusion device obviously descends and to such an extent that fused polymer can not be extruded through the nozzle orifice, furthermore, dwell time of polymer in oversized gap can not be easily controlled and partial fission of fused polymer often occurs. Value of the gap consists of main five tolerances: radial run out tolerance Δ_1 of screw rod, encircle tolerance Δ_2 of screw rod, non-perpendicularity tolerance Δ_3 barrel end relative to axes, hole tolerance Δ_4 of barrel, and gap Δ_5's demand for heat expansion offset, so the maximum gap δ_{max} between screw rod and barrel should not be less than Δ_1 plus Δ_2 plus Δ_3 plus Δ_4 plus Δ_5, and the minimum gap δ_{min} between screw rod and barrel should not be less than $\Delta_1 + \Delta_3 + \Delta_5$.

Besides, viscosity of polymer should also be considered, for example, if polymer material is nylon and the gap is larger, productivity may descent much, more than that the gap will vary along with wear and tear between screw rod and barrel. This shows that wear ability of screw rod and barrel material cannot be ignored yet. The value of the gap directly influences output and energy consumption of the screw extrusion device.

4.2 General Assembly of the Screw Extrusion Device

The screw extrusion device as a pressurization and feeding machine can not only increase flow quantity (not less than 0.5 cm^3/s) through the nozzle orifice, but can also get more smooth surface and more accurate dimensions' prototypes. The assembly drawing of the screw extrusion device is shown in Fig.3.
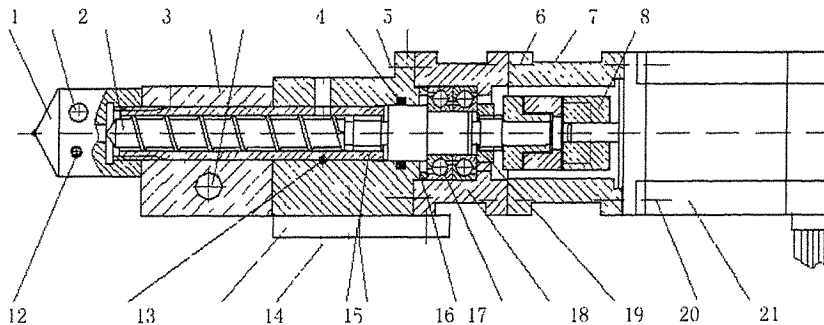
Fig.3 Assembly drawing of the screw extrusion device

1-nozzle 2-heating rod 3-screw rod 4-temperature controller 5-heater 6-gasket ring 7-bolt 8-front end cover
9-bolt 10-motor bracket 11-shaft coupling 12-thermocouple 13-straight pin 14-fixing bracket 15-bolt
16-barrel 17-bolt 18-shaft sleeve 19-bearing 20-back end cover 21-nut 22-bolt 23-step motor

Pre-heater 5 can convert solid-state polymer into fused polymer and temperature here is kept constant by temperature controller 4, and step motor 23 drives screw rod 3 and fused polymer is shoved to nozzle 1 along screw slot, heater 12 can convert fused polymer into melt polymer and temperature here is kept constant by temperature controller 4, and then melt polymer in the nozzle is equably extruded under propulsion of screw rod through the nozzle orifice, and then the extrusion polymer out of the nozzle is sequentially deposited layer by layer on the soleplate, finally, a prototype is rapidly built. The step motor is controlled by ACS Motion Controller (product of A.C.S. Electronics Co., Ltd., Israel). The rotate speed n of the screw rod is given by the following formula:

$$n = 0.08353 + 0.0049W_R H_C S_E \qquad (2)$$

Where W_R, H_C and S_E are the width of "road", distance between two layers (see Fig.1) and the linear velocity for the nozzle.

The screw extrusion device has been manufactured and a RP system (ESW-I-S RP system) with this machine has been also established, they are shown in Fig.4.
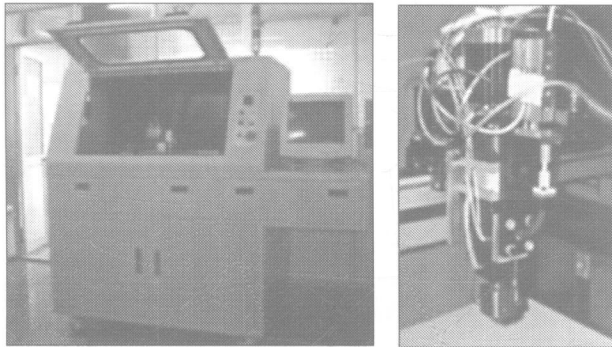
Fig. 4   ESW-I-S RP system and the new screw extrusion device

Many prototypes have been built on ESW-I-S. Two of these prototypes are shown in Fig.5. The left one in Fig.5 is a part of juicer. Its sizes are maximum diameter 80mm and height 118mm. This part built on ESW-I-S took four hours and twenty minutes (H_C equals to 0.2 mm), and its surfaces are greatly smooth.

5   CONCLUSIONS

This screw extrusion device can resolve conflicting requirements in improved performance

extrusion based layered manufacturing, increase the rate that material is extruded from a nozzle so that parts can be built more quickly, and achieve better feature resolution. A RP system with the screw extrusion device has been built, and its productivity increases 5 times higher than RP systems without the screw extrusion device. Moreover, parts built on it have smooth surface and accurate dimensions. Because of higher productivity of the RP system with the screw extrusion device, middle- or large-size prototypes may be built on it, so much as small-batch parts may be directly established on it. Hence its application area is further being broadened.
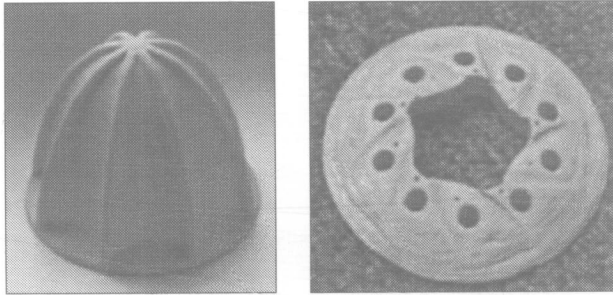
Fig. 5   The parts built on ESW-I-S

REFERENCES

1   **F.H., Zhu** (1984), Screw Rod Design and Theory, Light Industry Publishing Company, Beijing, China

2   **W.D., Xu** (1979), Hydrokinetics, National Defence Industry Publishing Company, Beijing, China

3   **United States Patent**, Patent Num.:5764521, June 9, 1998

4   **Yan Y. N., Zhu Jun** (1998), "Development of Rapid Prototyping in China", Proceedings of the First International Conference on Rapid Prototyping & Manufacturing '98, Beijing. Shaanxi: Shaanxi Science and Technology Press, 1998, pp.30-37

5   **Anne L. Marson, Vinod Kumar, Debasish Dutta, Michael J. Pratt**, "An Assessment of Data Requirements and Data Transfer Formats for Layered Manufacturing", NISTIR 6216, pp.1-47

6   **M. Atif Yardimci, Selçuk Güçeri** (1996), "Conceptual Framework for The Thermal Process Modeling of Fused Deposition", Rapid Prototyping Journal, Vol. 2 Num.2, pp.26-31

7 **Thompson, D.C. and Crawford, R.H.**, Optimizing Part Quality with Orientation, in Marcus, H.L., Beaman, J.J., Barlow, J.W., Bourell, D.L., and Crawford, R.H.(Eds) (1995), "Proceedings of the Solid Freeform Fabrication Symposium", Vol.6, the University of Texas at Austin, Austin, TX, pp.362-8

8 **Comb, J.W. and Priedeman, W.R. and Turley, P.W.**, FDM Technology Process Improvements, in Marcus, H.L., Beaman, J.J., Barlow, J.W., Bourell, D.L., and Crawford, R.H.(Eds) (1994), "Proceedings of the Solid Freeform Fabrication Symposium", Vol.5, pp.42-9

9 **Yardimci, M.A., Guceri, S., Agarwala, M.K. and Danforth, S.C.**, Part Quality Prediction Tools for Fused Deposition Processing, in Marcus, H.L., Beaman, J.J., Barlow, J.W., Bourell, D.L., and Crawford, R.H.(Eds) (1996), "Proceedings of the Solid Freeform Fabrication Symposium", Vol.7 (forthcoming)

10 **K. Elkins, C. Janak, H. Nordby, R.W. Gray IV, J.H. Bøhn, and D.G. Baird** (1997), "Soft Elastomers for Fused Deposition Modeling", Proc., 8th. Solid Freeform Fabrication Symposium, The University of Texas at Austin, August 11-13, pp. 441-448

11 **Jose F. Rodriguez, James P. Thomas, John E. Renaud** (2000), "Characterization of the mesostructure of fused-deposition acrylonitrile-butadiene-styrene materials", Rapid Prototyping Journal, Volume 6 Number 3, pp. 175-186

12 **Robert W. Gray IV, Donald G. Baird, Jan Helge Bøhn** (1998), "Effects of processing conditions on short TLCP fiber reinforced FDM parts", Rapid Prototyping Journal, Volume 4 Number 1, pp. 14-25

13 **Jeng-Ywan Jeng, Jia-Chang Wang , Tsung Te Lin** (2000), "A new flexible layer fabrication method for the jet deposition system to accelerate fabrication speed", Rapid Prototyping Journal, Volume 6 Number 4, pp. 226-234

14 **A.A. Tseng, M. Tanaka** (2001), "Advanced deposition techniques for freeform fabrication of metal and ceramic parts", Rapid Prototyping Journal, Volume 7 Number 1, pp. 6-17

Develop a process planning model for layer-based machining

Z Y YANG, Y H CHEN, and W S SZE
Department of Mechanical Engineering, The University of Hong Kong, Hong Kong

ABSTRACT

Layer-based machining (LBM) is a hybrid of layered manufacturing (LM) and material removal (MR) processes. In the LBM process, a part is built layer by layer. On each layer, part profile is shaped by CNC machining. In this paper, a decomposing-shaping-accumulating (DSA) model is proposed to describe the principles of the LBM process. Based on the DSA model, a hierarchical process planning model is developed for a structured representation of the process planning activities in LBM process. The model consists of three main modules: decomposing module, shaping module, and accumulating module, which are in line with the proposed DSA model. The proposed process planning model can be used as a general guideline for the development of LBM systems and associated computer aided process planning (CAPP)/computer aided manufacturing (CAM). An experimental LBM system is constructed to illustrate the application of the proposed process planning model.

KEYWORDS: layered manufacturing, layer-based machining, process planning

1 INTRODUCTION

There are two popular automated prototyping techniques: layered manufacturing (LM) and CNC machining. LM is developed and employed from the end of 1980s'. Several dominant LM processes are stereolithography (SL), selective laser sintering (SLS), fused deposition modeling (FDM), laminated object manufacturing (LOM), and three-dimensional printing

(3DP). By building models layer by layer, parts with complicated geometry can be build without the need of jigs and fixtures.

However, by comparing with CNC machining, the geometric accuracy and surface quality of the part built in LM are much poorer. Current LM processes also suffer from small build envelope and very few choices of raw materials.

In the last few years, there are increasing efforts on the integration of the LM and CNC processes to get the benefits of both [1]. It is expected that the integration may lead to a hybrid LM system that can eliminate the staircase phenomena and acquire better surface quality without losing the manufacturability for complex features. The common ground of these hybrid LM systems is to build models in additive ways, while shaping each layer in subtractive ways.

The layer-based machining (LBM) process presented in this paper is a typical example of this category. Other examples are: Shape Deposition Manufacturing (SDM) [2], Thick Layered Object Manufacturing (TLOM) [3], Computer-Aided Manufacturing of Laminated Engineering Materials (CAM-LEM) [4], Solvent Welding Freeform Fabrication Technique (SWIFT) [5]. A detailed review on these hybrid LM processes is given in [6].

Most LM techniques have some or all of the operations such as depositing, scanning, or sintering of material and post-assembly as a major step in their working cycles. Obviously, depositing and scanning are time-consuming and limit the types of materials that can be used, while the error incurred by sintering operation is unpredictable and hard to be compensated. Assembly operation with the aid of registration system is also error incurring and can destroy the continuity of material property within the finished model.

Basic researches on the LBM process have been reported in several papers [6-9] by the authors. In the LBM process, the raw material is in the forms of stock layers in different thickness. The stock layer thickness is usually much larger than that in conventional LM systems. The raw stock layers used to build the castle model and toilet seat shown in Fig. 1 have the available thickness of 2*mm*, 3*mm*, 5*mm*, 10*mm*, 20*mm*. The castle model is 380*mm* high with a base cylinder of 200*mm* in diameter. The height of the toilet seat is 240*mm*.
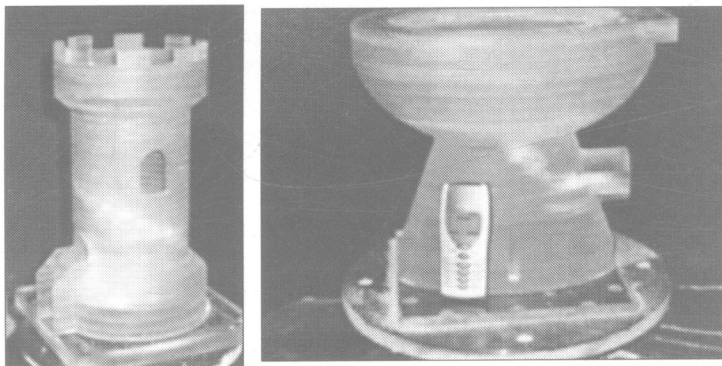
(a) Castle   (b) Toilet Seat
Figure 1. Physical models built by the LBM system

For optimum efficiency, a CAD model is decomposed into slabs according to manufacturability analysis. A slab is a section of the CAD model that can be machined without any interference in a given orientation. One slab is composed of one or more stock layers.
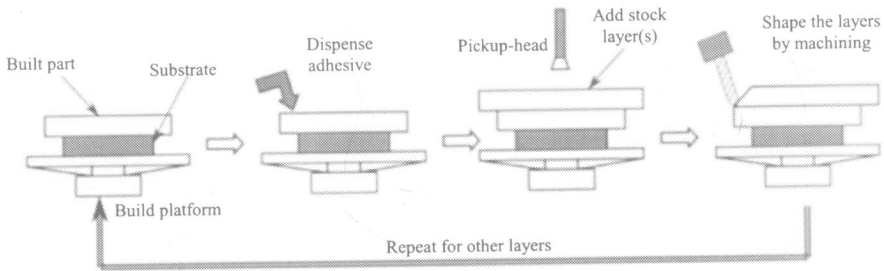
Figure 2. Working cycle of the LBM system

The working cycle of the LBM systems is described in Fig. 2: Before machining, system calibration and process planning are carried out. According to the planning, the first stock layer of the first slab is stuck onto the substrate, which is fixed on the rotary platform. If an adhesive mask is needed, the built part of the model is covered with the corresponding mask. Adhesive dispensing nozzle begins to spread adhesive in a planned path and pattern. Then next stock layer of this slab is stacked and repeated until all the stock layers of this slab are stacked. Five-axis machining is then carried out to shape this slab. Repeat the procedure until all slabs are shaped. An experimental system called Robot Based Layered Manufacturing (RoLM) based on the concept of LBM process has been established as shown in Fig. 3, the experimental LBM system consists of an articulated robot mounted on a linear track and a rotary table. A milling tool is installed on the robot end-effector.
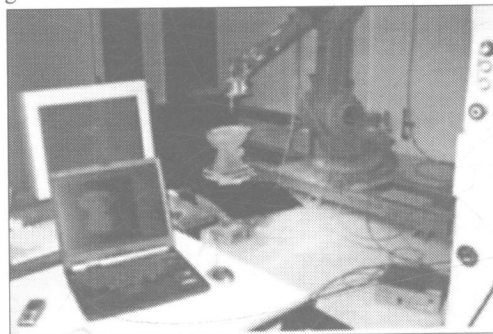
Figure 3. Construction of the experimental LBM system

2 DSA MODEL OF HYBRID LAYERED MANUFACTURING

Modeling is widely used in various engineering fields to describe complex processes or systems in a concise format. Abstraction and generalization are main methodologies of modeling. When a group of similar systems come into existence but not fully developed, modeling could be carried out to grasp the common characteristics of such systems. The

resulting model should be a general construction that could represent the existing systems, as well as a guiding tool for describing and solving the problem in the near future, or even creating new systems.

There are several commercialized hybrid LM systems and on-going research projects on hybrid LM. It is time now to conduct the hybrid LM process modeling. To the best knowledge of the authors, there is no reported model for the hybrid LM process. As a hybrid LM process is usually the integration of LM process and MR process, in this paper, the existing models of LM processes and material removal (MR) processes are analyzed first, and then the decomposition-shaping-accumulation (DSA) model for hybrid LM process is developed.

2.1 Modelling LM processes

The work of modeling LM processes is divided into two parts: one is modeling the principles of LM [10]; the other is modeling a specific LM process [11]. Lin et al. presented a mathematical model of LM processes to characterize and categorize the principles of LM in terms of model decomposition and material accumulation [10]. In their model, a 3D design model is represented by a set of points with sequence functions to correlate the layered processing information. Iso-sequence planes are defined as the processing layers to collect points with the same processing sequence and to define the material accumulation along its gradient direction. However, this decomposition-accumulation (DA) model is not suitable to any existing hybrid LM processes.
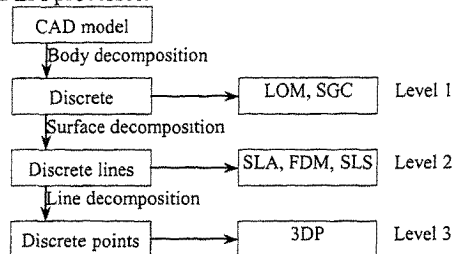
Figure 4. Hierarchy of decomposition-accumulation process

In the DA model, as shown in Fig. 4, a CAD model decomposition is divided into three sub-decompositions: body composition, surface composition, and line decomposition. Accordingly, the accumulation can be classified surface accumulation, line accumulation, and point accumulation. But in hybrid LM processes, the accumulation is *volume accumulation* because the decomposed layers have very large thickness, which should be regarded as volume rather than surface. Therefore, the hierarchical decomposition process cannot be directly applied to the hybrid LM processes.

In the part decomposition of the DA model, the iso-sequence planes intersect and slice the 3D model; in the material accumulation process, the iso-sequence planes form the processing layers on which material will be added. Although the introduction of iso-sequence planes can solve the problem of describing non-planar processing surface, such as Curved Layer LOM (CLOM) [12], it cannot represent the accumulation process of hybrid LM, in which the profile of each layer is usually not parallel to the gradient direction of iso-sequence planes. The profile of each layer here is defined as the design surface between the two slicing planes

of that layer. In hybrid LM, it maybe sculptured surface; while in conventional LM, it is swept surface.

In a word, the decomposition-accumulation model is not suitable to describe hybrid LM systems. The reason is that the part surface in between two slicing planes is ignored in the DA model.

2.2 Modeling traditional MR

A comprehensive survey of various analytical and empirical models of MR processes is given by Jain *at el.* [13]. MR processes include advanced machining processes, ultrasonic machining (USM), abrasive jet machining (AJM), water jet machining (WJM), etc.

One of the most widely used MR processes in industry is NC machining, in which the material is removed by a cutter. As tool path generation is the main problem when modeling the NC machining process, many works have been done on this issue [14]. NC machining is now a well-developed technology. Recent researches are mainly concentrated on sculptured surface machining. A good reference about sculptured surface machining models is given in [15].

One of the simplest models of MR processes is depicted in Fig. 5. In traditional MR processes, one stock is used for one part and the stock must contain the whole part. Different MR processes diverge form each other at the point how the excess material be removed from the stock.
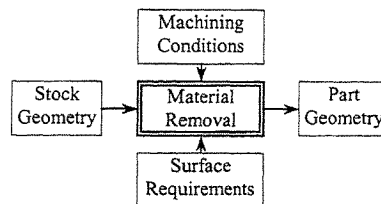
Figure 5. Simplest model of MR process

2.3 Modelling hybrid LM

Based on the above analysis, DA model is not applicable to represent hybrid LM processes. Obviously, MR models cannot describe the material additional nature of hybrid LM processes. We combine the two kinds of models together to form a decomposition-shaping-accumulating (DSA) model to generalize hybrid LM processes, as shown in Fig. 6.
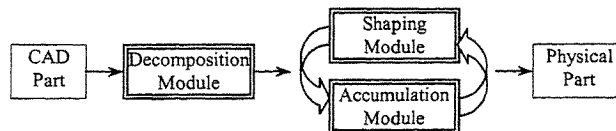
Figure 6. DSA model of hybrid LM process

In all current hybrid LM processes, the main operation on the CAD model domain is decomposing for manufacturability. Before dealing with materials stock, the CAD representation of the part is decomposed into sub-parts. Each sub-part is defined by decomposing surfaces and design surfaces, both could be free-form surfaces. We call such a decomposition is *volume decomposition*. When the decomposing surfaces are parallel planes

that are perpendicular to the build orientation, volume decomposition degenerates to *terrain decomposition* [16]. If the distance between decomposing planes in terrain decomposition is small enough, terrain decomposition degenerates to *surface decomposition*. Analogically, the degeneration goes over *line decomposition* and *point decomposition*. It can be seen that the aforementioned decomposition hierarchy in DA model is a subset of that of DSA model.

Accumulating is performed on raw material stock domain. There are two patterns of the sequence of accumulating and shaping: bond-cut and cut-bond. The task of accumulating is to prepare the material stock for next sub-part (in bond-cut pattern) or assemble the sub-part back to the desired part (in cut-bond pattern). For example, in SDM, raw materials are deposited first to near-net shape of each sub-part (called *compact* in SDM); in LBM, raw stock layers are stacked one by one to form a stock slab, from which the part slab could be shaped; in SWIFT, one raw material sheet is fed first and then designed surface of this sub-part is obtained by CNC machining. These processes are in bond-cut pattern. As for cut-bond pattern, in TLOM, machined part slab are assembled manually; in CAM-LEM, laser carved sheets are stacked together using registration system.

In cut-bond processes, the geometry accuracy and surface quality are partially determined by the post-assembly operations. In order to automate all operations, special grippers for moving shaped stocks have to be designed. This increases the system complexity. The registration systems also destroy the continuity of part properties. Therefore, cut-bond processes is theoretical not so competitive as bond-cut processes. The major merit of cut-bond processes is the instinct of distributed manufacturing.

Both cut-bond and bond-cut processes have the same features: a volume is accumulated in each accumulating operation; desired part surface is (almost) not acquired by accumulating. We call the accumulation in hybrid LM processes *volume accumulation*.

Shaping is performed on stock/part domain. Part surfaces are obtained by MR processes. Any MR process could be employed in hybrid LM system as the shaping method. For example, SDM, SWIFT, TLOM, and LBM use multi-axis NC machining as the shaping method; while CAM-LEM uses laser cutting. Theoretically the shaping method determines the part surface quality. In hybrid LM processes, at least 1-order approximation can be achieved. Comparing to 0-order approximation in conventional LM, hybrid LM brings a significant improvement on surface quality.

From the DSA model, it can be observed that hybrid LM processes tend to obtain better part accuracy with higher-degree approximation of MR processes and increase the manufacturability with decomposition-accumulation operation. However, the proposed DSA model is too generalized and hard to be applied in practice. The following section introduces the proposed process planning model of LBM, which is an elaboration of DSA model.

3 PROCESS PLANNING MODEL OF LBM

Based on the proposed DSA model, we developed a process planning model to represents the concept of LBM process. The objects of developing such a model include:
  1) To implement an LBM system;
  2) To develop corresponding CAPP software;
  3) To provide guidance if practical CAPP are carried out manually.

The process planning model consists three modules in line with the DSA model of hybrid LM processes:

1) Decomposing Module: define sub-parts (which are called *slabs* in LBM process) of the part CAD model;
2) Accumulating Module: prepare raw material stock for each slab;
3) Shaping Module: generate machining stages and operations for each slab and sequence them.

3.1 Decomposition Module

As depicted in Fig. 7, the input of decomposition module includes CAD representation, tool database, machine database, and material database. The main operations are adaptive slicing, accessibility analyzing, raw-stock layer combination. The output is the CAD representation of slabs and corresponding raw-stock layer series.
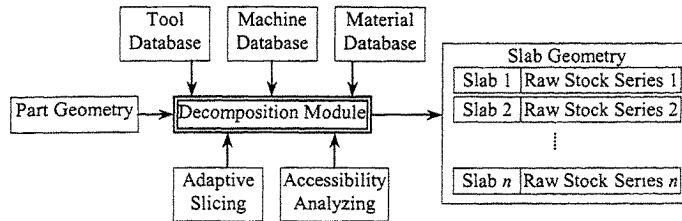
Figure 7. Decomposing Module in LBM

3.2 Accumulating Module

The output of accumulating module, as shown in Fig. 8, includes the bonding sequence of slabs, the bonding sequence of raw-stock layers within each slab. If needed, adhesive mask is also generated and operations of applying adhesive masks are inserted to the queue of bonding. The input "adjacent relationships of slabs/layers" is part of the output of decomposition module "slab geometry".
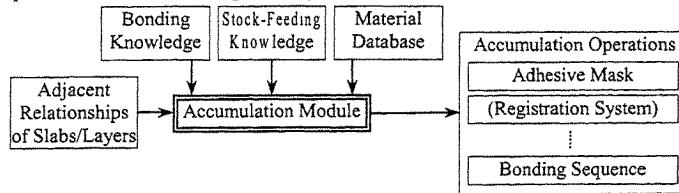
Figure 8. Accumulating Module in LBM

3.3 Shaping Module
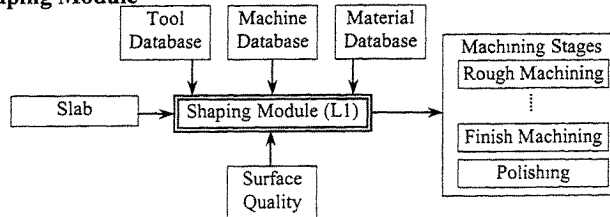
Figure 9. Shaping Module in LBM Level 1: define machining stages

The shaping module of the LBM process consists of two levels, which is similar to the model proposed by Choi et al. [15]. Level 1 (shown in Fig. 9) is to define the machining stages and level 2 (shown in Fig. 10) is to define machining operations for each machining stages. In the shaping module, variant milling tool path topology could be chosen, such as z-constant, helical, or boundary curve parallel.

In the LBM process, bond-cut pattern is adopted. The shaping module and accumulating module breaks in each other and merges together. The shaping operations and accumulating operations are sequenced carefully.
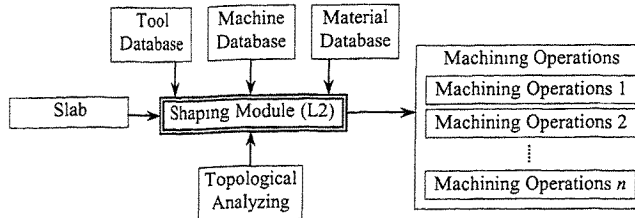
Figure 10. Shaping Module in LBM Level 2: define machining operations

4 APPLICATION OF PROCESS PLANNING MODEL OF LBM

The proposed process planning model is a structured level-of-detail representation of the planning activities in LBM process. An LBM system can be realized by implement both the system hardware and software incorporating the three modules.

4.1 System construction
In the experimental LBM system shown in Fig. 2, the hardware of decomposing module is a PC running CAPP/CAM software developed by the authors. The shaping module is realized by an articulated robot with a milling tool mounted on its end-effector and a rotary fixture. This construction can perform 5-axis milling to acquire the desired surface accuracy and finish. The realization of accumulating module is not automated yet. When implemented, it should contain a magazine feeding system and a selective adhesive spraying system.

4.2 Software integration
Under the guidance of the LBM process planning model, an OpenGL enhanced software package for LBM is developed with Visual C++. The package is independent from any commercial CAD software and integrated with CAPP and CAM. The main modules include build orientation determination and support design, decomposition for manufacturability (adaptive slicing), stock layer combination, interference-free tool path generation, and operations sequencing. The input is the STL file representation of a part. The output is recorded in a data structure called machining-instruction-sheet, including the STL files of each slab, tool paths in robot control format, adhesive masks, and stacking sequence.

5 DISCUSSIONS AND CONCLUSIONS

In this paper, a decomposing-shaping-accumulating (DSA) model is developed to describe the existing hybrid layered manufacturing processes. Based on the DSA model, a hierarchical

process planning model is developed to give a structured representation of the process planning activities in LBM process. The process planning model can be used to realize LBM systems and as a guideline to develop CAPP/CAM software for LBM systems. An experimental LBM system is constructed and sample parts are built to demonstrate the application of this process planning model.

By comparing the LBM model with LM model and MR model, it can be concluded that LBM is different from MR in that the LBM process has a material accumulating operation. LBM uses volume decomposing and volume accumulating, which is more efficient than that in LM.

Different implementation of the modules of the LBM process model, or different sequence of the operations lead to different LBM systems. This is instructive in the creation of new processes. For example, new LM techniques may come into being because of volume decomposing in multiple directions [17].

Improvements could be done in several aspects. A mathematical model is necessary to describe the detail of the DSA model. Process models for heterogeneous part building needs to be considered in the future.

ACKNOWLEDGEMENT

REFERENCES
1 **Kulkarni, P.** and **Dutta, D.** (2000) On the integration of layered manufacturing and material removal process. *Journal of Manufacturing Science and Engineering*. Vol. 122, pp. 100-108.
2 **Ramaswami, K.** (1997) Process planning for Shape Deposition Manufacturing. Ph.D. Thesis, Stanford University.
3 **Horváth, I., Vergeest, J. S., Broek, J. J., Rusák, Z.** and **Smit, B.** (1998) Tool profile and tool path calculation for free-form thick-layered fabrication. *Computer-Aided Design*. Vol. 30, Num. 14, pp. 1097-1110.
4 **Newman, W. S., Mathewson, B. B., Zheng, Y.** and **Choi, S.** (1996) A novel selective-area gripper for layered assembly of laminated objects. *Robotics and Computer-Integrated Manufacturing*. Vol. 12, Num. 4, pp. 293-302.
5 **Taylor, J. B.** and **Cormier, D. R.** (2001) A process for solvent welded rapid prototyping tooling. *Robotics and Computer-Integrated Manufacturing*. Vol. 17, pp. 151-157.
6 **Yang, Z. Y., Chen, Y. H.,** and **Sze, W. S.** (2002) Layer-based machining: recent development and support structure design. *Journal of Engineering Manufacture, Proceedings of the Institution of Mechanical Engineers: Part B*. Vol.216, pp. 979-991.
7 **Yang, Z. Y., Chen, Y. H.,** and **Sze, W. S.** (2001) Determining build orientation for layer-based machining. *International Journal of Advanced Manufacturing Technology*. Vol. 18, Num. 5, pp. 313-322.
8 **Yang, Z. Y.** and **Chen, Y. H.** (2001) Process planning in layer-based machining. *Proceedings of ASME 6th Design for Manufacturing Conference*, Pittsburgh, USA.
9 **Chen, Y. H.** and **Song, Y.** (2001) The development of a layer-based machining system. *Computer-Aided Design*. Vol. 33, Num. 4, pp. 331-342.

10 **Lin, F., Yan, Y. N.** and **Sun, W.** (2001) A decomposition-accumulation model for layered manufacturing fabrication. *Rapid Prototyping Journal.* Vol. 7, Num. 1, pp. 24-31.

11 **Choi, S. H.** and **Samavedam, S.** (2002) Modelling and optimisation of Rapid Prototyping. *Computers in Industry.* Vol. 47, pp. 39-53.

12 **Klosterman, A. D., Chartoff, R. P., Osborne, N. R., Graves, G. A., Lightman, A. Han, G., Bezeredi, A., Rodrigues, S., Pak, S., Kalmanovich, G., Dodin, L.** and **Tu, S.** (1998) Curved layer LOM of ceramics and composites. *Solid Freeform Fabrication Symposium Proceedings.* University of Texas at Austin, Austin, TX, August, pp. 671-680.

13 **Jain, N. K.** and **Jain, V. K.** (2001) Modeling of material removal in mechanical type advanced machining processes: a state-of-art review. *International Journal of Machine Tools and Manufacture.* Vol. 41, Num. 11, pp. 1573-1635.

14 **Dragomatz, D.** and **Mann, S.** (1997) A classified bibliography of literature on NC milling path generation. *Computer-Aided Design.* Vol. 29, Num. 3, pp. 239-247.

15 **Choi, B. K.** and **Jerard, R. B.** (1998) *Sculptured surface machining: theory and applications.* Dordrecht, London: Kluwer Academic Pulishers.

16 **Fekete, S. P.** and **Mitchell, J. S. B.** (2001) Terrain Decomposition and Layered Manufacturing. *International Journal of Computational Geometry & Applications.* Vol. 11, Num. 6, pp. 647-668.

17 **Singh, P.** and **Dutta, D.** (2001) Multi-direction slicing for layered manufacturing. *Journal of Computing and Information Science in Engineering.* Vol. 1, Num. 2, pp. 129-142.

Rapid tooling – producing functional metal parts from fused deposition modelling process using plaster moulding

R NARAIN
Department of Mechanical Engineering, Motilal Nehru National Institute of Technology, Allahabad, India
A SRIVASTAVA
Department of Engineering, Northern Railway, Allahabad, India

ABSTRACT

One of the growing applications of Rapid Prototyping is Rapid Tooling. Rapid prototyping has proved to be a cost-effective and time-efficient approach for producing patterns, moulds, and dies. This paper presents the procedures that were adopted for the production of aluminum casting using a plaster mould prepared from a pattern produced on FDM-1650 type Rapid Prototyping machine. Dimensional accuracies and the problems encountered therein have also been discussed.

1 INTRODUCTION

Sweeping changes in manufacturing practices have taken place during the last two decades and are accelerated with added momentum. These changes have been brought about by the development of several new technologies. Rapid Prototyping (RP) is one such technology that generally refers to techniques that produce shaped parts by gradual creation or addition of solid materials. RP embraces a range of new technologies for producing accurate parts directly from CAD models in a few hours, with little need for human intervention. As a result of it errors are minimized and product development costs and lead times substantially reduced. Other benefits of RP include; production of parts having increased complexity, more organic sculptured shapes for functional or aesthetic reasons, optimize part design, reduce parts count, minimize material and reduce costs.

RP technology finds application in Design; Engineering, Analysis, and planning; and Tooling and Manufacturing (1). Its uses are expanding. Among the new applications are medical modeling, aerospace parts, and building construction etc. (2). While steadily dropping in price it holds a promising future as it can join together liquid, powder, and sheet metal to form parts and fabricate plastic, wood, ceramics, and metal objects. Some of the commercially available systems are Stereo-Lithography (SL), Fused Deposition Modeling (FDM), Ink Jet Printing, 3D-Printing, Selective Laser Sintering (SLS), Laminated Object Manufacturing (LOM), Laser

Cladding, Selective Laser Chemical Vapour Deposition etc. Kruth (3), and Pham et al. (4) have given a detailed review of these processes and other related issues.

2 RAPID TOOLING PROCESSES

Rapid Tooling (RT) is the process of employing RP technology to produce casting, dies, moulds, EDM electrodes etc. The RT techniques are divided into two classes, based on a criterion related to the number of operations required to produce a tool or die. These are:

1. **Direct Tooling**
2. **Indirect Tooling**

The direct tooling processes are the techniques that require no intermediate steps in the manufacture of the tools (e.g. moulds) and can directly supply the production tool from CAD data files. It translates the desired geometry into a negative representation so that the fabricated cavity has a shape that will just hold the desired object. However, the dimensions of the mould may also need to be adjusted to compensate for shrinkage of the moulding material due to cooling.

The indirect tooling processes utilize a pattern made by RP techniques as an aid for mould making because it is sometimes simpler to fabricate a positive pattern on which the mould is cast than to fabricate the negative mould directly. One advantage of this process is that it is easier to check for errors in the design on positive patterns than on the negative. Also it is much easier to sand or polish the positive patterns.

The RT processes can be studied under three different headings (5, 6). These are:

1. **Soft Tooling**
2. **Bridge Tooling**
3. **Hard Tooling**

Soft tooling processes are fairly quick and inexpensive. These include processes such as silicone rubber moulds, epoxy moulds, and spray metal moulds. However these processes have limited utility because of the low mechanical properties of the parts produced and the difficulties in reproducing the finer details in the case of intricate shapes.

Bridge tooling is a mid ground where hard tooling is not economically justified, and the soft tooling cannot satisfy the requirements of injection-moulded parts. It thus fills the gap between "soft" and "hard" tooling. These processes produce tools capable of short prototype runs up to a few hundred parts using the same material and manufacturing process as for final production parts. One of the biggest advantages of bridge tooling is that it permits considerable time savings. An example of bridge tooling is the "Direct AIM (ACES – accurate, clear, epoxy, solid- injection moulding)".

Hard tooling processes produce metal tools capable of generating hundreds of thousands or millions of parts rapidly and at low cost. KelTool process is one such example falling in this category, which claims to have produced complex geometries, and fine details with good reproducibility.

3 PLASTER MOULDING

In plaster mould casting the mould is made of plaster of paris (gypsum or calcium sulfate), with the addition of talc and silica flour to improve strength and to control the time required for the plaster to set. These components are mixed with water, and the resulting slurry is poured over the pattern (7). After the plaster is set the pattern is removed and the mould is dried to remove the moisture. The mould halves are assembled to form the mould cavity and are preheated before pouring the molten metal into the mould. Patterns for plaster moulding are generally made of aluminum alloys, thermosetting plastics, brass or zinc alloys. Plaster mould casting is used only for aluminum, magnesium, zinc and some copper-based alloys. The castings have fine details with good surface finish. Because plaster moulds have lower thermal conductivity than others, the castings cool slowly, and the more uniform grain structure is obtained with less warpage. Fig. 1 shows the schematic of the typical process sequence for plaster moulding.

4 THE FDM-BASED TOOLING PROCESS

The FDM process uses thermoplastic wire-like filaments, which are melted in the delivery head. The material is then extruded from the head and deposited on a layer-by-layer basis. The layering lamination technique is based upon the rapid solidification of the molten laminate material from the modeling filament. The semi-liquid thermoplastic material is deposited onto thin layers, building the model upwards off a fixture base. The plastic or wax material solidifies in place positioned by the XY controlled extrusion head. The extrusion process shears the material and it quickly solidifies while bonding to the previous layer by heating it and then fusing. The model is fabricated upon a piston, which is lowered between layers to make room for the next layer. The process is repeated until the part is fully built.
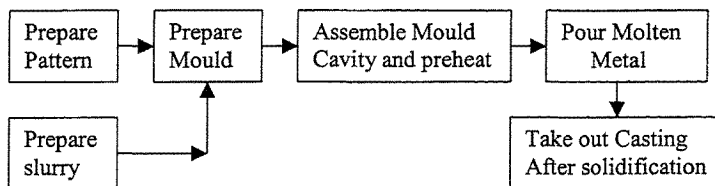
Fig. 1 Schematic of Process Sequence of Plaster Moulding

In this work ABS pattern made on FDM 1650 type RP Machine has been used for aluminum casting because patterns built from ABS offer a number of quality advantages over patterns made by other RP processes (8). These are clear burn out, robustness, the ability to be handled without damage, dimensional stability and ease of pattern preparation. Surface finish preparation of the pattern is important to achieve the best results. Besides aluminum, materials that have been cast include 17-4 stainless steel, cobalt chrome, brass and beryllium copper. The FDM process has been used in investment casting, sand casting, sheet metal prototypes etc. The specific aim of this paper is to report on the procedures that were followed for the fabrication of production tooling using the FDM machine and plaster moulding. The

subsequent paragraphs describe the pattern making, shell formation around the pattern, separation of the pattern from the shell for mould formation, and finally using the mould for aluminum casting. For the purpose of this work a rotor blade of a stepper motor was chosen for casting. Fig. 2 gives the schematic of the process sequence for fabricating the tooling.
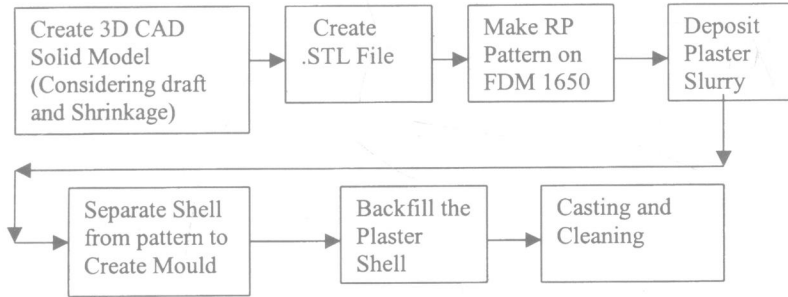
Create 3D CAD Solid Model (Considering draft and Shrinkage) → Create .STL File → Make RP Pattern on FDM 1650 → Deposit Plaster Slurry

Separate Shell from pattern to Create Mould → Backfill the Plaster Shell → Casting and Cleaning

Fig. 2 The Rapid Tooling Process For Plaster Moulding

5 PATTERN MAKING

First of all a 3D CAD solid model of the required pattern is to be made using any CAD package such as Pro/Engineer, I-DEAS, etc. In this case Auto CAD- R14 has been used and the rotor blade was modeled in two halves as shown in Fig.3. The model was then converted into an .STL file using a specific translator. The .STL file was then sent to the FDM slicing and preprocessing software i.e. QuickSlice, where selection of proper orientation, creation of
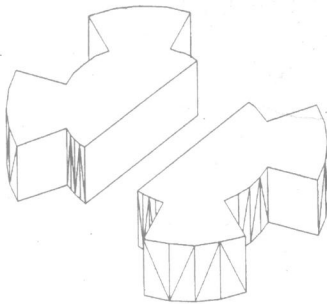
Fig. 3 The 3D Model of The Rotor Blade
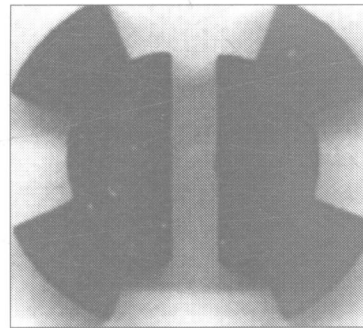
Fig. 4 The Build FDM Model

supports, slicing, and other related parameters are determined. The sliced model and supports were then converted into a SML (Stratasys Modeling Language) file. This contains actual instruction codes for the FDM machine to follow. These include the specific tool paths i.e. the roads, the extruded material to be deposited to create each road etc. The SML file is then sent to the RP 1650 machine, where the head creates each horizontal layer by depositing molten extruded material on a foam foundation until the part is completed.

The material used for making the model was ABS (P400) plastic and for making the supports was ABS (P301). The temperature at which the material and the supports were extruded from the nozzle tip was 270 ^{0}C and 265 ^{0}C. After the part was completed, it was taken out, supports carefully detached, and was ready for use as the FDM parts do not require post curing. Fig. 4 shows the build FDM model.

6 MOULDING AND CASTING

To create plaster mould, Plaster of Paris is deposited around the RP pattern. The first layer was fine coated and then dried in sunlight for an hour. After this it is dried in oven at about 105 ^{0}C for another one hour. After depositing the first layer the subsequent layers were coated with the slurry and dried in sunlight and then oven. The process was repeated several times so that a layer of sufficient thickness was uniformly deposited around the pattern. To separate the shell from the pattern it was placed in a hot air oven at 300^{0}C for about 24 hours. The complete plastic pattern had melted leaving behind a cavity for pouring the molten metal. The small amount of residual plastic sticking to the mould walls was cleaned mechanically. Figs. 5 and 6 show the RP model coated with plaster and the plaster mould after the removal of the pattern. The moulds were then backfilled to provide necessary supports. Top gating system was used for pouring the molten metal into the mould. After pouring the molten metal, the mould was broken and the casting taken out. Fig. 7 shows the final casting made in aluminum.
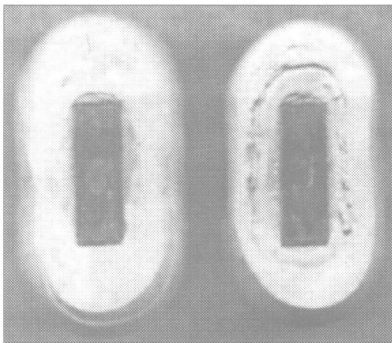
Fig. 5 RP Model Coated With Plaster   **Fig. 6 Plaster Mould For Casting**
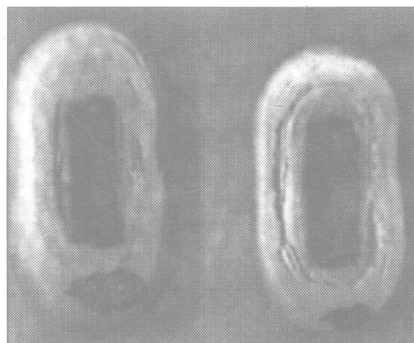
Fig. 7 Aluminum Casting

7 DISCUSSION

Some of the difficulties encountered during the fabrication process are now discussed in this section. The biggest difficulty that was encountered was during the withdrawal of the RP plastic pattern from the mould in the oven. Initially five layers of slurry were deposited. But then when the shell was placed in the oven it developed cracks on some portions of the mould because of the thermal stresses. So in the subsequent trial five more layers were deposited on the shell so as to make it thick enough to withstand the thermal stresses. Now no cracks were formed when the pattern was withdrawn. Another difficulty was with the proper alignment of the two halves of the moulds. A slight shift was observed at the parting line upon casting. However this can be removed by providing proper supports. Yet another difficulty was with the surface finish, which could not be obtained of a very high order because of the low permeability of the plaster mould.

For the purpose of accuracy, measurements were taken of different dimensions of the RP model, and the casting and it was then compared with the dimensions of the CAD model. The dimensions included the pitch, the blade thickness, and the diameter of the rotor blade. The error on these dimensions between the CAD model and the RP model varied in the range of .1% to .27%. However the deviations in the dimensions of the CAD model and the aluminum casting varied from 2.95% to 6.4%. One of the reasons for this larger deviation in the dimensions of the final casting and the low quality of surface finish seems to be the incomplete removal of the plastic pattern from the mould when heated in the furnace. The residual plastic sticking to the mould walls had to be removed mechanically causing thereby some damage to the mould walls. This led to increased dimensions of the mould, and thus the casting, and poor surface finish as well. Another reason for over sized casting could be improper allowance on the pattern. Adequate allowances should be provided on the pattern particularly for the shrinkage of the metal. Also for a good quality casting a very smooth pattern with excellent surface finish is required. For this sanding and filing of the pattern

could have been done. Another alternative worth exploring could be the use of some low melting point material such as wax for making the pattern itself so that it can easily be melted and removed from the shell leaving behind a clear and clean cavity.

8 CONCLUSIONS

Rapid tooling offers an economical alternative to traditional machining. This paper discusses the use of FDM type Rapid Prototyping process to produce rapid tooling for plaster moulding. Making patterns, moulds, and dies for producing parts having complex geometry is very time consuming, expensive, and tedious. Use of RP process is very helpful in eliminating these drawbacks. However as is clear from the discussion of the previous paragraph a detailed investigation is further needed into the reasons, and the possible remedies for the problems encountered, so as to enable rapid and reliable manufacturing of the mould and thus enhance the applicability of the plaster moulding process.

9 REFERENCES

1   Chua Chee Kai & Leong Kah Fai. (1997) Rapid Prototyping: Principles and Applications in Manufacturing. John Wiley & Sons, (Asia) Pte Ltd, Singapore.

2   Aronson, Robert. B. (2000) Its Not Just RP Anymore. Journal of Manufacturing Engineering, Vol. 5, pp. 98-110.

3   Kruth, J. P., Leu, L. C., and Nakagawa, T. (1988) Progress in Additive Manufacturing and Rapid Prototyping. Annals of the CIRP, Vol. 47/2, pp. 525-540.

4   Pham, D. T., and Gault, R. S. (1998) A comparison of rapid prototyping technologies. International Journal of Machine Tools & Manufacture, Vol. 38, pp. 1257-1287.

5   Jacobs, P. (1997) Recent advances in rapid tooling from stereolithography. In proceedings of the Second National Conference on Developments in Rapid Prototyping and Tooling, held at Buckinghamshire College, UK, Edited by Graham Bennett, pp. 135-151.

6   Pham, D. T., Dimov, S., and Lacan, F. (1998) Techniques for firm tooling using rapid prototyping. Proc. Instn. of Mech. Engrs., Vol. 212, part B, pp. 269-277.

7   Kalpakjian, S., and Schmid, S. R. (2000) Manufacturing Engineering and Technology. Addison Wesley Longman (Singapore) Pte. Ltd. 4th Edition, pp. 276.

8   Gouldsen, C., and Blake, P. (1998) Investment Casting Using FDM/ABS Rapid Prototype patterns. A literature on Investment Casting from Stratasys INC. Bangalore, India.

Stereo-thermal-lithography – a new principle for rapid prototyping

P BARTOLO
Mechanical Engineering Department, Polytechnic Institute, Leiria, Portugal
G MITCHELL
Polymer Science Centre, University of Reading, UK

SYNOPSIS

A new fabrication process for rapid prototyping is proposed in this paper. Optical and thermal effects are simultaneously used in this process to locally induce a phase change in a liquid resin. This phase change phenomena is used to "write" complex three-dimensional shapes or patterns. Such objects or patterns can involve macroscopic engineering prototypes through to nanostructures for exploitation in waveguiding and photonic crystals. Several advantages can be achieved through this new process, in terms of accuracy, cost and time.

1 INTRODUCTION

The most important rapid prototyping technologies currently available or under research are indicated in fig 1. Among them, stereolithography (1) is the most popular process. Stereolithography involves the curing or solidification of a liquid photosensitive polymer by a laser beam scanned across its surface. The laser supplies energy that induces a chemical reaction, bonding large numbers of small molecules and forming a highly cross-linked polymer. As the reaction proceeds, the viscosity of the resin in the direct vicinity of the light spot increases until vitrification occurs. The first layer is formed on an elevator platform, which is then lowered to allow new liquid resin to flow onto the working surface. This process is repeated, with each new layer adhering to the previous one. Finally, due to the insufficient extent of cure obtained during the building process a post-cure operation is needed to complete the solidification process. Important consequences of insufficient polymerisation are warping due to relaxation, diffusion and evaporation of low-molecular weight components, and post-cure shrinkage due to density changes associated with the post-cure of the liquid resin trapped in the lattice structure (2-4). Shrinkage, which can cause internal stresses in the model, is another important problem associated with stereolithography (SL) and is a logical consequence of the formation of large molecules from small ones during the polymerisation process, which results in an increase in density (3,4).

An alternative process to stereolithography is proposed in this research paper. This process, designated by stereo-thermal-lithography (STLG), comprises both heat and ultraviolet (UV) radiation effects, solving some of the major limitations of the conventional SL process, such as efficiency, speed, accuracy and tunability (5). The fabrication conditions used in this new process were established from the understanding of the physical and chemical transformations of thermosetting resins under thermal and photo-initiated curing reactions.
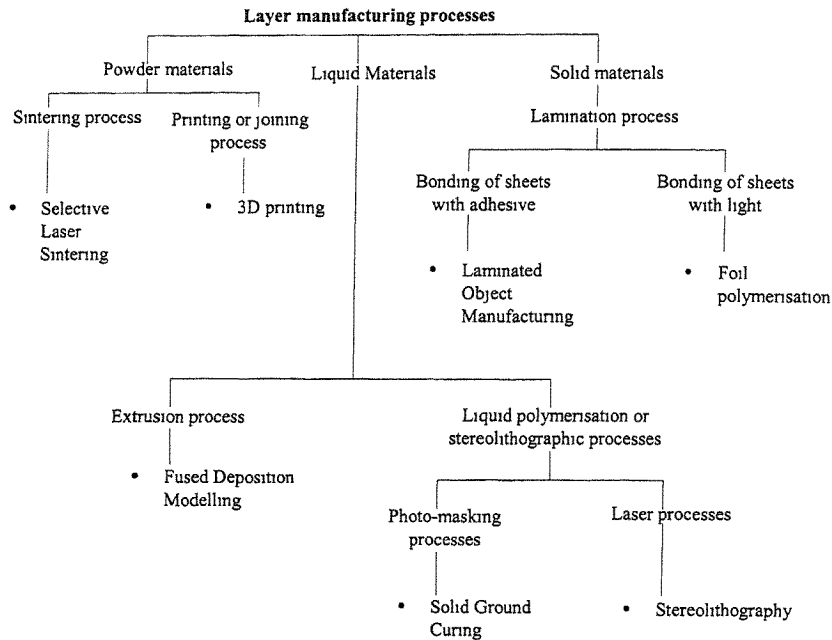
Figure 1. Classification of rapid prototyping technologies (6,7)

2 LIGHT-MOLECULES' INTERACTION

Since most monomers or pre-polymers do not produce initiating species upon light exposure, it is necessary to introduce low-molecular weight organic molecules called initiators that will start polymerisation, through photophysical and photochemical processes (8).

When a molecule absorbs light, electrons are set into motion by the oscillating electric field of the light (8-11). As the incident photon exchanges its energy with the molecule, a valence electron is promoted from the highest occupied molecular orbital to an unoccupied molecular orbital, with the formation of an excited singlet state molecule (8-11). However, this excited molecule is a short living specie that disappears by various competitive processes, dissipating the excited energy (8-11). The absorption of light by a molecule and the subsequent evolution of its excited states are illustrated in fig 2 through the Jablonski energy diagram (9). Two processes can be identified: photo-physical processes (radiative and non-radiative) and photo-chemical processes (8-11).

Radiative mechanisms involve the absorption of a photon by a molecule in its ground state (S_0) and the emission of a photon from an electronically excited state by either fluorescence (de-excitation of an excited state with the same spin multiplicity of the ground sate) or phosphorescence (de-excitation of an excited state with different spin multiplicity of the ground sate) (9,11). Non-radiative processes include internal conversion (IC) and the intersystem crossing (ISC), which occurs among states of different spin multiplicity creating excited triplet states (T_1) (9,11).
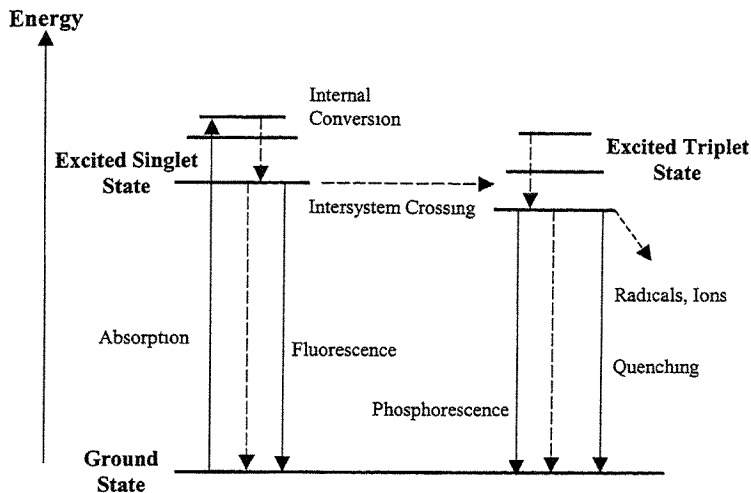
Figure 2. The Jablonski energy diagram.

The photochemical process is a transformation of the starting molecule, through cleavage processes, electron transfer reactions, hydrogen abstraction, etc. Due to the short life of the excited singlet state molecules many photochemical reactions occur only via excited triplet states, which are longer living species (10-12). Singlet or triplet states are electronic states where the molecule possesses all paired electrons or two unpaired electrons, respectively (8,10).

3 THE SL CHEMICAL PRINCIPLE

In stereolithographic processes, resin samples containing a certain amount of photo-initiator are irradiated using UV radiation. The interaction of a photon of UV radiation with the photo-initiator can give one of the two results:

- the photon is not absorbed

- the photon is absorbed. In this case the initiator molecule forms an excited singlet state, which then undergoes intersystem crossing to form a triplet state. This triplet state will generate radicals by undergoing α-cleavage fragmentation (8). The excitation and fragmentation process can be described by:

$$S_0 \xrightarrow{\ h\nu\ } S_1 \xrightarrow{\ ISC\ } T_1 \xrightarrow{\ \alpha-cleavage\ } Radicals \qquad (1)$$
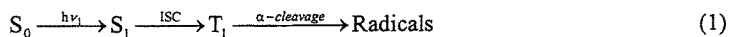
Finally, these radicals can undergo three different possible reactions:

- the radicals recombine to reform the photo-initiator molecule (reaction 1)

- the radicals escape from each other but are quenched by the inhibitor or oxygen molecules (reaction 2)

- the radicals are not quenched, and react with monomers to start the curing reaction (reaction 3).

Among these three possible mechanisms, reaction 2 is the dominant one during the first stage of the curing reaction (13). This period is called the induction time, and can be reduced by increasing the light intensity, photo-initiator concentration or temperature.

Optimisation of this subsystem can be achieved by balancing the effects of photo-initiator concentration and light intensity. However it is important to point out that:

- the increase in photo-initiator increases the initiation and reaction rates and consequently the amount of solid material formed, but reduces the UV penetration depth. In the case of a stereolithographic process, this means smaller (*i.e.* thinner) layers

- the increase in light intensity increases the initiation rate and the rate of gel formation. However, systems using high power UV laser beams are more expensive (14)

- if the balance between photo-initiator concentration and light intensity is not the ideal one, this could also means slow reaction rates. In this case, diffusion may be an additional problem (the reaction becomes less localised), with implications in terms of the accuracy of the obtained models.

4 THE STLG CHEMICAL PRINCIPLE

The STLG process (fig 3) uses ultraviolet radiation and thermal energy (produced, for example, by infrared radiation) to initiate the polymerisation reaction in a medium containing both photo- and thermal-initiators. The concentrations of both initiators are carefully selected and the reaction starts only when there is a combination of UV radiation and thermal energy. This way, the amount of each initiator must be low to start the polymerisation by only one of these two effects. However, at the point where the two effects intersect each other, the amount of radicals generated is sufficiently high to initiate the polymerisation process. On the other hand, the temperature is used to produce radicals through the fragmentation of thermal-initiators and simultaneously to increase the rate of gel formation of the photo-initiated curing reaction. Consequently, the extent of cure is increased, so no post-cure is needed. This is an important advantage of this process, because if products of high quality are required it is usually necessary to ensure a fairly high degree of gel formation. Moreover, if a small amount of solid material is formed, a large number of unreacted molecules are still present, so it is possible that an excessive heat generation will occur during the post-cure phase, leading to the distortion and warping of the built model.

The initiation process of in this new approach can be described as follows:

- **path 1 (thermal energy effects):**

$$\text{Initiator} \xrightarrow{\text{heat}} \text{Radicals} \rightarrow \text{Radical recombination and/or inhibition effects} \qquad (2)$$
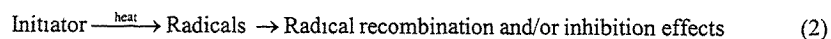
- **path 2 (UV radiation effects):**

$$\text{Initiator} \xrightarrow{\text{UV radiation}} \text{Radicals} \rightarrow \text{Radical recombination and/or inhibition effects} \qquad (3)$$

- **intersection of path 1 and path 2 (thermal energy + UV radiation effects):**

$$\left.\begin{array}{l}\text{Initiator} \xrightarrow{\text{heat}} \text{Radicals} \\[2em] \text{Initiator} \xrightarrow{\text{UV radiation}} \text{Radicals}\end{array}\right\} \rightarrow \text{Sufficient numbers of radicals to start the polymerisation} \qquad (4)$$
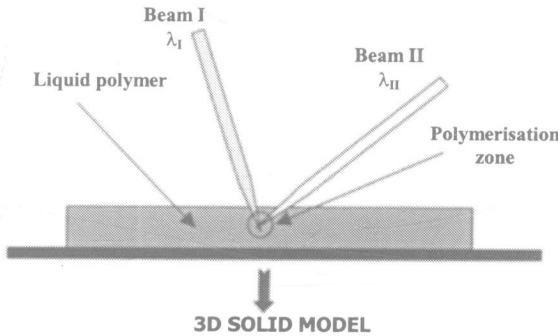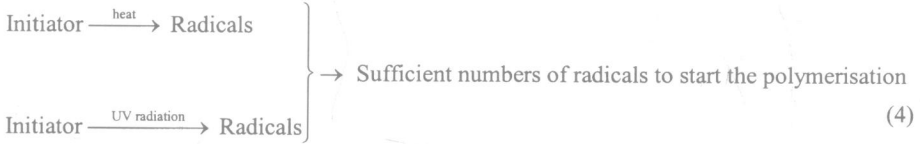
Figure 3. The dual laser curing principle of the STLG

The chemical principle of this system corresponds to a different approach from the Swainson's previous one (15-16), who tried to develop a stereolithographic system using radiation of two different wavelengths to write inside a vat containing a liquid resin. The work of Swainson, presupposes that the curing reaction is initiated through the sequential bi-photonic excitation of a single type of initiator molecule. Different bi-photonic excitation processes were subsequently tested and the most important among them, in terms of the initiator molecules used, was the following one(15-16):

$$S_0 \xrightarrow{h\nu_1} S_1 \xrightarrow{\text{ISC}} T_1 \xrightarrow{h\nu_2} T_n \qquad (5)$$

In this case, the initiator molecule absorbs a photon of an appropriate wavelength and forms an excited singlet state, which then undergoes intersystem crossing to form the triplet state. This triplet state decays into the ground state if not further stimulated. However, if it is exposed to radiation of appropriate energy, it is capable of forming a higher excited triplet state, which is highly reactive.

The excitation mechanisms tested by Swainson presented two important problems. If a high concentration of initiator is used, the penetration depth of the radiation is reduced. However, if low concentrations of initiator are employed, the radiation penetration depth increases, but the reaction becomes slower and less localised. Swainson was never able to solve these problems.

This new system presents several advantages with respect to conventional stereolithography and the Swainson bi-photonic process:

- the generation of radicals is more efficient

- small concentrations of the two types of initiator are used, this way enabling the radiation to penetrate deeper into the polymer

- the combination of UV radiation and temperature increases the reaction rate

- the curing reaction is more localised, which will improve the accuracy of the produced models

- the system has more tunability.

This new fabrication approach represents a better system in terms of both the speed of the process and the quality of the models produced. It represents a truly one-step process for the generation of physical models because it avoids the time-consuming step of post-curing (performed in a separate apparatus). In addition, it represents a less expensive system compared with conventional stereolithography, as the entire fabrication process is carried out in a single apparatus.

In the future, the principles used in the STLG process can be applied to a new commercial system that will enable to generate models inside a liquid medium. This will turn the building process into a real three-dimensional process.

5 IMPLEMENTATION AND RESULTS

To test the STLG principles an in-house device was made by machining an aluminium block. In the initial stage of this research, the mask-based method of irradiation was the selected method because it is easier to implement experimentally. The device was placed on a metallic structure, which supports a mercury lamp (Blak-Ray B100 AP, UVP Inc.) transmitting light in the range 300-400 nm with λ_{max} at 365.5 nm, and is equipped with thermocouples connected to a temperature controller. Temperature gradients are produced by using heating elements and a refrigeration system, which uses cold water flow. Moreover, the light intensity can be controlled by changing the distance between the plate containing the apparatus and the upper level of the metallic structure where the UV lamp is placed.

Different shapes of polymer structures were produced using two different types of masks created on black card (see fig 4 and table 1), and placed 2 mm above the sample position. The accuracy of the STLG process was evaluated by comparing the dimensions of the produced models with the dimensions of the corresponding mask as indicated in table 2. Because no mechanism was used to define a regular thickness value, the accuracy is only investigated in terms of the lateral dimensions of the models, which should be as close as possible to the dimensions of the masks.

In this study, a thermosetting resin system containing 0.5 wt% of photo-initiator were irradiated at room temperature using the conventional UV irradiation mechanism, while a thermosetting resin system containing 0.01 wt% of photo-initiator and 0.01 wt% were thermal and photo-cured at 110 ^{0}C using the STLG principle.

Mask type 1:   Mask type 2:
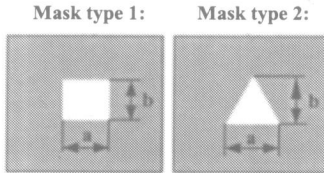
Figure 4. Mask types

Table 1. Dimensions of each mask type

	Mask type 1	Mask type 2
Dimensions (mm)	a = 11	a = 11.5
	b = 11	b = 11

Table 2. Dimensions of models produced by conventional UV radiation and by stereo-thermal-lithography. The error is the difference between object and mask dimensions

Mask type 1	Dimensions (mm)	Error (mm)
Mask dimensions:	11	
	Conventional UV irradiation	
Object dimensions:	11.557	0.557
	Stereo-thermal-lithography	
Object dimensions:	11.235	0.235
Mask type 2		
Mask dimensions:	a = 11.5 ; b = 11	
	Conventional UV irradiation	
Object dimensions:	a = 12.159	0.659
	b = 11.52	0.512
	Stereo-thermal-lithography	
Object dimensions:	a = 11.712	0.212
	b = 11.131	0.131

6 CONCLUSIONS

The findings of this experimental research show that the accuracy of the models obtained through this new STLG principle can be greatly enhanced.

This new process involves two distinct chemical reactions occurring simultaneously, the thermal-initiated curing reaction initiated by heat and the photo-initiated curing reaction initiated by UV irradiation. Moreover, the resin samples have small concentration of thermal and photo-initiators, so they show a very low reaction rate when they occur separately. However, whenever the two effects intersect each other the population of radicals produced will be sufficiently high to initiate the curing reaction with a higher reaction rate. On the other hand, the coupling of temperature and UV radiation induces an increase in the temperature in the regions where both effects are present, with advantages in terms of reaction rate and fractional conversions, improving the process in terms of a better localised curing reaction. The stereo-thermal-lithography chemical principle developed for this new process was recently submitted for patent registration under the name "*Dual beam photo-fabrication*" and is being further developed through two ongoing projects: the LRAMP programme

(www.rdg.ac.uk/psc/lramp.pdf) and the project called "Two photon writing processes in organic polymers", both supported by UK funding.

REFERENCES

1 **Hull, C.W.** (1984) US Pat. 4575330.

2 **Weidemann, B., Dusel, K.H.** and **Eschl,** (1995) Rapid Prototyping Journal, 1(3), 17.

3 **Klaus, I.S.** and **Knowles, W.S.** (1966) J. Appl. Polym. Sci., 10, 887.

4 **Flach, L.** and **Chartoff, R.P.** (1994) Proceedings of the fifth international conference on rapid prototyping, Edited by A.J. Lightman and R.P. Chartoff, University of Dayton, 181.

5 **Bartolo, P.J.** (2001) Optical approaches to macroscopic and microscopic engineeering, Unpublished PhD Thesis, University of Reading, UK,.

6 **Kruth, J.P.** (1991) Annals of the CIRP, 40, 603.

7 **Campbell, R.I.** and **Martorelli, M.** (2000) Proc. of the 9th European Conference on Rapid Prototyping and Manufacturing, Edited by R.I. Campbell, University of Nottingham, 37.

8 **Fouassier, J.P.** (1995) Photoinitiation, photopolymerization, and photocuring: fundamentals and applications, Hanser, New York.

9 **Olayan, H.B. Hamid, H.S.** and **Owen, E.D.** (1996) J.M.S. Rev. Macromol. Chem. Phys., C36(4), 671.

10 **Gilbert, A.** and **Baggot, J.** (1991) Essentials of molecular photochemistry, Blackel Science Publishers, London.

11 **Ushiki, H.** and **Horie, K.** (1989) in Handbook of polymer science and technology, Vol. 4, Edited by N.P. Cheremisinoff, Dekker, New York.

12 **Guillet, J.** (1985) Polymer photophysics and photochemistry, Cambridge University Press, Cambridge.

13 **Selli, E.** and **Bellobono, I.R.** (1993) in Radiation curing in polymer science and technology, Vol. III: Polymerisation mechanisms, Edited by J. P. Fouassier and J. F. Rabek, Elsevier Science Publishers, London.

14 **Chen, H. Corbel, S., Allanic, A.L.** and **Andre, J.C.** (1992) Proceedings of the third international conference on rapid prototyping, Edited by A.J. Lightman and R.P. Chartoff, University of Dayton, 1.

15 **Swainson, W.K.** and **Kramer, S.D.** (1982) US Pat. 4471470.

16 **Swainson, W.K.** and **Kramer, S.D.** (1975) US Pat. 4078229.

Benchmarking for decision support in RP systems

M MAHESH, Y S WONG, J Y H FUH, and H T LOH
Department of Mechanical Engineering, National University of Singapore, Singapore

SYNOPSIS

Benchmarking is invaluable for evaluating the performance efficiency of RP processes and systems. Appropriate benchmark parts can be designed for performance evaluation of RP systems and processes, and provide helpful data for use in decision support systems. Besides the process and the material, other factors such as the building style and specific process parameters may also affect the accuracy and finish of the part. In RP benchmarking, it is necessary not only to standardize the design of the benchmark part, but also the fabrication and measurement processes. This paper presents issues on RP benchmarking and attempts to identify factors affecting the definition, fabrication, measurements and analysis of benchmark parts.

1 INTRODUCTION

A geometrical benchmark part is proposed and designed to evaluate the efficiency and performance of the rapid prototyping machines/processes. The benchmark part aims to incorporate key shapes and features that are currently employed in better-known benchmark parts (1-8). A review on these benchmarks is given in (10). Additionally, it also includes new geometric features, such as freeform surfaces, certain mechanical features and pass-fail features that are increasingly required or expected of RP processes/systems. The benchmark part has been fabricated via four widely used RP processes: liquid-based Stereolithography Apparatus (SLA), solid-based Laminated Object Manufacturing (LOM) and Fused Deposition Modelling (FDM) and powder-based Selective Laser Sintering (SLS). A discussion on the problems encountered and its influence on the geometric accuracy are discussed in this paper. The applicability of the proposed design as a generalised benchmark is discussed with respect to its suitability for evaluation of the four widely used RP processes. Apart from the

geometrical benchmark part, the mechanical and process benchmarks are discussed as well.

2 PROPOSED BENCHMARK PART

Figure 1 shows the proposed geometric benchmark part. The detailed description of the benchmark and its comparison with other reported benchmarks are discussed in Reference 10.
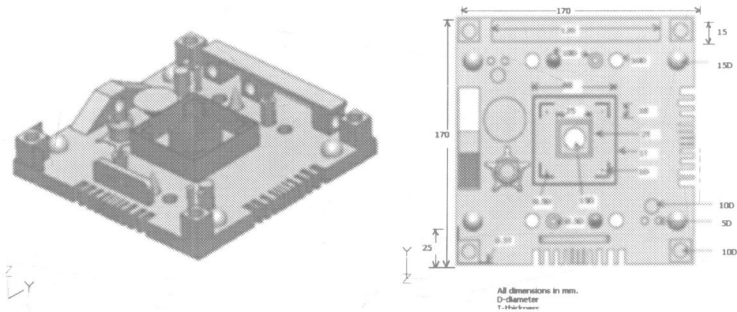
Figure 1. Benchmark stl file and dimensions

An experimental study was conducted to demonstrate the importance of such a benchmark in estimating the performance of the RP processes and machines. The four well-known RP processes of SLA, SLS, FDM and LOM were used in the experimental study. The benchmark parts were fabricated with the default set of machine parameters.

2.1 Fabrication the benchmark part on the SLA machine
The benchmark part was fabricated on the SLA-190/250 by 3D System's. The material used was an epoxy resin. It was generally found that the features on the benchmark part were built quite well. All the features on the benchmark part, including the pass/fail features could be visually observed (Fig.2). The benchmark part was measured using a CMM machine to determine the geometric accuracy of the fabricated part.
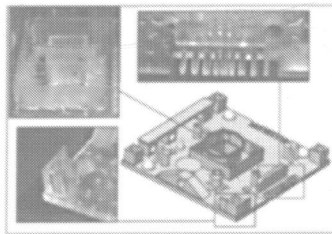
Figure 2. Ability of SLA to build all features including Pass/fail features

The thin cylinders and thin walls were warped (Fig.2). Part shrinkage is a common problem in SLA processes and any of the following could affect the part accuracy:

- Prolonged exposure of the green part to air before curing
- Long solvent bath
- Less/over curing in the UV chamber.

It was observed that the base of the benchmark part was not really bottom-flat and had some undulations but later it was attributed to the less initial preheating of the resin before the actual building, and possibly due to a drop in the laser power during the building of the first few layers. So it is important to ensure that the resin reaches the operating temperature before building the part. The part accuracy is also dependant on the post-processing techniques and the experience of the user.

2.2 Fabrication of the benchmark part on the SLS machine

The parts were fabricated on the DTM Sinterstation 2500 and the material used was ProtoForm Composite (LNC- 7000). All the features except some of the pass-fail features could be built from the SLS process. Two-benchmark parts were built. The first part was built with the default parameter setting of the machine and the laser power was about 8W. It was found that the part was warped. In addition, the thin cylinders, 0.5mm holes, 0.5mm slots, and thin wall could not be built. This was attributed to the laser power and parameter setting. The benchmark part was built again with a new set of parameters and the laser power was raised slightly to about 9W. The new set of parameters was obtained by doing a number of trial-and-error experiments. This time the part turned out to be better and most of the features could be built. But there were problems to separate the part from the part cake. It was very delicate to brush off the unsintered powder around the fine features like the 0.5mm cylinder. The reason for this was that the increase in the part bed temperature had caused the unsintered powder to adhere together.

2.2.1 Effects of Warpage on the Surface accuracy

Warpage has an adverse effect on the geometric accuracy of the fabricated parts. Warpage could be either the form that affects the edges causing the part to curl upwards (Fig.3), or the type that affects one side like the one shown in Fig.4.

Figure 3. Part warpage drifting towards the edges

Figure 4. Part warpage drifting towards a side

This was the first part fabricated using default parameters setting. The warpage was high towards a side of the part as shown in Fig.4.The part was distorted from the centre towards the corner. From Fig.5, the shape and size of the square boss towards the corner can be seen to be effected due to the warpage. Another part was fabricated on the same machine but with a different orientation and was placed more towards the inner side. The laser power was also increased by a watt (9W). The part bed temperature and the powder bed temperature was 192 and 101 °C. The warpage of the fabricated part was less compared to the earlier one as can be seen from Fig.6. It should be noted here that the benchmark part could also serve the purpose of optimising the machine. To optimise the machine parameters, a few experiments were
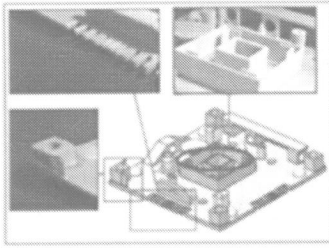
Figure 5. Effect of warpage on the first SLS part
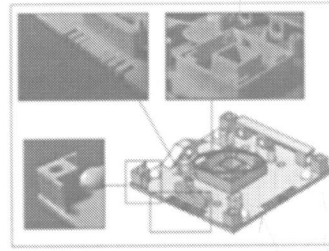
Figure 6. Second SLS part built including pass/fail features.

initially conducted by fabricating the benchmark part and adjusting the parameters by visual inspection during the process. However the machine parameter setting and optimising essentially rely on a trial-and-error method and differ with the type of material and to a certain extent, the experience of the operator.

2.3 Fabrication of the benchmark part on the LOM machine

The benchmark part was fabricated on the Helisys LOM 1015 system and the material used for the fabrication was laminated paper of 0.09652 mm thickness. The benchmark part by the LOM machine showed some interesting results (Fig.7). All the features could be built but it was practically impossible to separate some of the features from the support structures. For example some of the pass/ fail features like the thin cylinders could not be separated from the support squares. Delamination is a severe problem in the case of the LOM systems. The benchmark part could help in the identification and evaluation of delamination, such as those associated with certain geometric features.
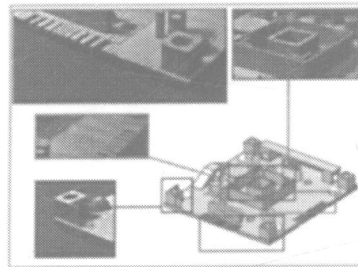
Figure 7. Highlighted LOM part showing some results of fabrication like the delamination, air holes, thin walls and brackets

Delamination greatly restrains the geometric accuracy of a part. Delamination generally occurs if the bonding between the layers is not strong enough to hold them together. Any separation between the bonding layers could cause the part to distort, thereby affecting the geometric and dimensional accuracy of the part. Any of the following could influence delamination: the material selection, and heating, weak bonding between the base of the part and platform. The humidity and induction of air holes in between the layers also cause delamination to occur.

2.3.1 Effect of delamination on the benchmark part

The LOM-fabricated benchmark part was delaminated more towards the base as can be seen from Fig.7 An analysis on the possible cause of delamination indicates that it was caused by the induction of air holes during the fabrication process. Humidity of the environment was the root cause that induced air holes in the model that ultimately caused the model to delaminate. The features above the base of the benchmark part were affected as well by delamination resulting in easy break away of features during the post processing. In general the dimensional accuracy of the LOM part is quite satisfactory although not as accurate as the SLA part. Most of the features, including some of the pass-fail features, could be built. Post processing is most delicate and time-consuming in LOM.

2.4 Fabrication of the benchmark part on the FDM machine

The benchmark part was fabricated on the Stratasys FDM 3000 using ABS-400 as part material and ABS-400R as the support material. All part features, except the pass/ fail features, could be built (Fig.8). The failure of the pass/fail features could be attributed to FDM as basically a deposition process. This shows that the FDM process was least suitable to build fine features as compared with the other processes. In addition, the surface finish and dimensional accuracy were not as good as those from the other processes.
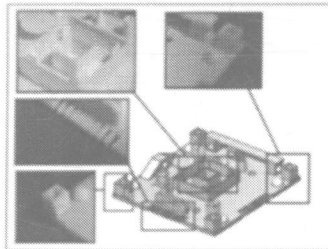
Figure 8. Highlighted areas on the FDM benchmark part showing the warpage, failure to build- very thin walls, cylinders.

2.4.1 Warpage

The base was warped but less compared to that of the SLS part. The warpage was seen to be more progressive towards the corners and thus curling the base upwards. Based on the experience of the users, it could have been due to the temperatures setting in the build area or late removal of the part from the FDM machine after fabrication was over. Using proper temperature setting for particular materials is important to fabricate a good part.

3. MEASUREMENT

A CMM (Co-ordinate Measuring Machine) is well suited for the measurement of the benchmark parts because of its versatility and speed. Most CMMs have high accuracy compared and can be programmed to carry out a variety of automatic measurements, ranging from simple to complex. A CMM determines the measured dimensions and shape errors, namely flatness, parallelism, angularity, straightness and roundness. Some basic measuring instruments such as the vernier calliper and screw gauge with high accuracy have also been used in the measurement of the fabricated parts.

Figures 9 show the plots of the % deviation in terms of accuracy of certain features in the benchmark part. The percentage deviation is measured as the change in value from the nominal dimensions of the STL file. The surface roughness R_a (arithmetic mean) is measured using the Rank Taylor Hobson's surface texture measuring equipment based on ISO 468 and other international standards as appropriate and followed by Rank Taylor Hobson equipment. The result of the surface roughness measurement is as shown in Figure 10.
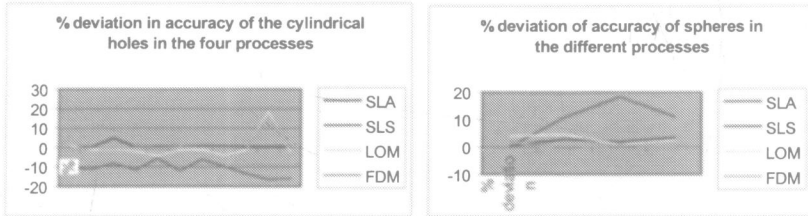
Figure 9. % Deviation in accuracy
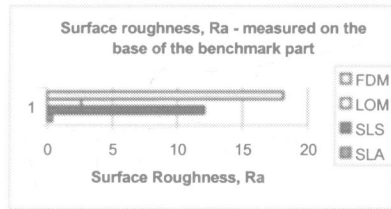
Figure 10. Surface Roughness, Ra

Table 1 shows a comparison of the feature built on the various RP processes. The results are tabulated and rated from A- B in the order of the best to the worst performance respectively. The results indicate that SLA is still the best in terms of accuracy and surface finish followed by SLS, LOM and FDM. It is also clear that FDM and LOM are least suitable in building very fine features. For medium-sized features, the order of performance is SLA, LOM, SLS and FDM. The order of best performance in terms of surface roughness is SLA, LOM, SLS and FDM.

Table 1.Comparison of the RP processes based on the geometric benchmark

Features	SLA	SLS	FDM	LOM
Square base	A	B	A-B	B-C
Cube	A	A-B	A-B	A
Flat beam	A-B	A	A-B	A-B
Cylindrical holes (z-direction)	A	A-B	B-C	A
Cylindrical holes (x, y)	A	A-B	B-C	A
Spheres	A	A-B	A-B	A-B
Solid cylinders	A	A-B	A-B	A-B
Hollow cylinders	A	B	B	A
Cones	A	A-B	B	A-B

Slots	A	B-C	B-C	A-B
Hollow squares	A-B	B	A-B	A-B
Brackets	A	A	A-B	A-B
Circular holes	A	A-B	A-B	A
Mechanical features	A	B	C-D	A-B
Pass/fail features	A	A-C	D	B-C
Surface Roughness	0.4	12.1	18.4	2.6
Part Warpage	A	B-C	B	B
Rating A to D represent best to least desirable property respectively				

4. MECHANICAL BENCHMARKS

There are certain mechanical properties that need to be evaluated, which includes: shrinkage, tensile and compressive strengths, curling, and creep characteristics (Fig.11). The obvious influence of shrinkage and curling are geometric distortions affecting dimensional inaccuracy.
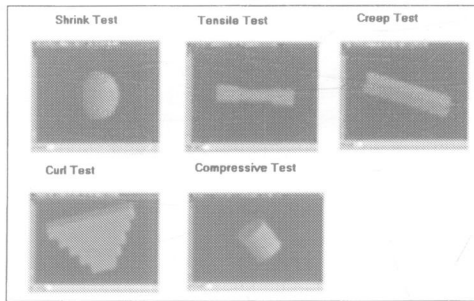
Figure 11. Mechanical tests.

We can generally use the standards that are used by the ASTM for testing a particular property. It would be more convenient to customise the standards pertaining to the RP industry.

5. PROCESS BENCHMARKS

Besides identifying an appropriate benchmark part design, care has to be taken with regard to the fabrication and the process related to it. Hence, in this experiment, the process benchmark has been another important consideration. The goal of the process is to improve the accuracy and surface finish, and obviously reduce the build time. By fabricating the benchmark part, the objective is to optimise appropriate parameters and evolve with a successful process benchmark.

6. DECISION SUPPORT

The data obtained by measurement from the fabricated benchmark parts on the various RP process will be processed for the evaluation of a particular process and the system based on that process. A database will be the central repository of such data. All useful information will be captured from the different RP processes and systems from the fabrication of the benchmark and will later be used by decision support tools or systems to select suitable systems or processes to meet specific objectives (9-10).

CONCLUSION

RP benchmarking would greatly facilitate decision-making to identify and determine a suitable RP process or system for the fabrication of RP parts that could meet specific requirements. The data obtained will be stored in a benchmark database to be used by appropriate RP decision support systems to identify suitable RP machines, materials and processes to meet specific requirements, and even suggest vendors and bureau services.

REFERENCES

1   **Juster, N.P. and Childs, T.H.C.** (1994) Linear and Geometric Accuracies from Layer Manufacturing, Annals of the CIRP, 43, No.1, 163-166

2   **Kruth, J P.** (1991) Material Incress Manufacturing by Rapid Prototyping Techniques, CIRP Annals. 41st General Assembly of CIRP. Stanford, CA, USA, 40, No. 2, 603-614

3   **Gargiulo, E.P.** (1992) Stereolithography Process Accuracy: User Experience, Proceedings of the 1st European Conference on Rapid Prototyping. Univ. of Nottingham, July 6th –7th,187-207

4   **Lart, G.** (1992) Comparison of Rapid Prototyping Systems, Proceedings of the 1st European Conference on Rapid Prototyping. Univ. of Nottingham, U.K, 6th-7th July, 243-254.

5   **Juster, N.P. and Childs, T.H.C.** (1994) A Comparison of Rapid Prototyping Processes, Proceedings of the 3rd European Conference on the Rapid Prototyping and Manufacturing, University of Nottingham, 6th –7th July, 35-52

6   **Ippolito R, Iuliano L, De Filippi A** (1994) A New User Part for Performances Evaluation of Rapid Prototyping Systems, Proceeding of 3rd European Conference on Rapid Prototyping and Manufacturing, Nottingham 6-7 July, U.K, 327-339.

7   **Mike Shellabear**, EOS Gmbh (1999), Benchmark Study of Accuracy and Surface Quality in RP&M Models, *RAPTEC*, Task 4.2, Report 2.

8   **Reeves, P.E., Cobb, R.C.** (1996) Surface Deviation Modelling of LMT Processes – A Comparative Analysis, *Proceedings of the 5th European conference on Rapid Prototyping and Manufacturing, Helsinki, 4–6th June.*

9   **Fuh, J.Y.H., Loh, H.T, Wong, Y.S., Shi, D.P., Mahesh M. and Chong T.S.,** A Web-based database system for RP machines, processes and materials selection, *Chapter2, Software Solutions for RP*, PEP Ltd, UK, 2002.

10   **Wong Y.S, Jerry Fuh.Y.H, Loh.H.T, Mani Mahesh.,** Rapid prototyping and manufacturing (RP&M) benchmarking, *Chapter 3, Software Solutions for RP*, PEP Ltd, UK, 2002

Process parameter optimization using a feed-forward neural network for direct metal laser sintering process

Y NING, J Y H FUH, Y S WONG, and H T LOH
Department of Mechanical Engineering, National University of Singapore, Singapore

ABSTRACT

As one of the Rapid Prototyping (RP) processes, Direct Metal Laser Sintering (DMLS) method is used to build prototype parts by depositing and melting metal powders layer by layer. Using DMLS, metal powder can be melted directly to build functional prototypes. The resulting properties of interest to the users include the processing time, mechanical properties, and geometric accuracy. Some main process parameters could affect the results significantly. These process parameters, which involve the laser scan speed, laser power, hatch density, and layer thickness, can be determined by the operator before building the prototype parts. But the relationships between these parameters and resulting properties are complicated. In many cases, the effects of different parameters on the resulting properties contradict one another. A method based on the Feed-forward Neural Network(NN) is described in this paper for predicting the resulting properties of the laser-sintered metal parts. After continuous training by using the data pairs, this NN constructs a mapping relationship between the process parameters and resulting properties. The objective of this research is to obtain the statistical relationships of the selected process parameters and the achieved process results.

1 INTRODUCTION

The Selective Laser Sintering (SLS) method can directly sinter different materials such as metal, sand, and plastic. With the capability of directly fabricating the metal parts, the SLS process has become more competitive. An experimental laser sintering system has been developed by the National Univ. of Singapore (NUS) RP Group to study the Direct Metal Laser Sintering process. Quality improvement of the metal prototype is widely expected in the industry. Unfortunately, using a single rule to judge the process quality from the resulting properties is not adequate. Many important goals, including dimension accuracy, mechanical strength, processing time, cost, surface finish and so on, need to be considered. Very often, these goals do not necessary result in a similar trend as the change in the process parameters.

A good scenario is to weight some of them and ignore others according to the user's requirement. A fixed set of the parameter values that can achieve the best outcome for all desired properties inevitably does not exist. For example, to take less time to finish the process, a larger layer thickness and faster scan speed are needed; but normally higher mechanical properties cannot be achieved. In practice, even for an experienced operator, it is difficult to set the correct parameter values when the customers have different objectives for the final parts.

By changing the process parameters, the operator could control the metal part's resulting properties effectively. The main process parameters include laser scan speed, laser power, hatch density, and layer thickness. The purpose of this paper is to construct a mapping relationship between the process parameters and part properties achieved. At the first stage, the research focused more on the processing time, mechanical properties, and geometric accuracy. For processing time, there is a clear quantitative relationship with the parameters, so a good mapping function could be built. But for the other two properties, the relationship is somewhat more complicated and hard to deduce directly. Therefore, a feed-forward NN is introduced to build a mapping between the two goals in relation to the process parameters by using the experimental data pairs for NN training. An orthogonal arrays design method [1] was adopted and 27 experimental data pairs were used to train the network. Five more data pairs were used to verify the training results. This has been shown to achieve good mapping.

2 PROCESS PARAMETERS AND BUILD-TIME ESTIMATION

The laser scan speed, laser power, hatch density, layer thickness and scan path are the important process parameters that should be carefully selected by the operators. A zig-zag scan path (shown in Fig. 1) was chosen during all the experiments.
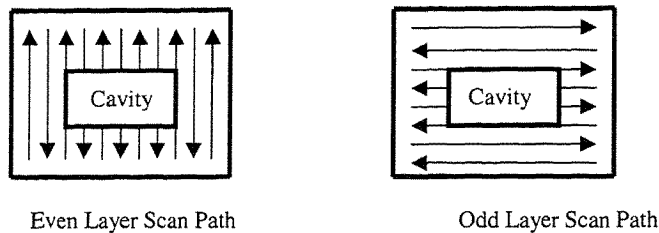
Even Layer Scan Path   Odd Layer Scan Path
Fig. 1 Zig-zag scan path

Experimental results have shown that this scan path will bring more homogeneous properties to the final parts than most of other scan paths.

2.1 Build-time estimation
To understand the effect of the process parameters on the total build-time by the zig-zag scan method, the following equation has been derived based on the volume of the part.

$$t_{Total} = \frac{1}{D_{Thickness}} \left(\frac{V_{Parts}/S_{Scan} + V_{Cavitys}/S_{Jump}}{D_{Hatch}} + H_{Height} t_{Layer} \right) \tag{1}$$

where t_{Total}, denotes the total build time, $D_{Thickness}$ and D_{Hatch} the layer thickness and hatch distance, V_{Parts} and $V_{Cavitys}$ the total part volume and the cavities volume in the part, H_{Height} the part height, S_{Scan} the scan speed set by the operators, and S_{Jump} the jump-scan speed, (the default setting is 2000mm/sec for the experimental machine). t_{Layer} denotes the time used between each layer building. Some unimportant factors were ignored to simplify the function. These factors are mainly related to the time used in the changing of scan state. It includes the scan speed acceleration and deceleration time and laser-on delay (laser-off delay) when the laser switches on or off. Each time of these state changes is less than 50μs by the default laser system setting [2].

2.2 Laser scan speed and laser power
In the DMLS process, through absorbing enough laser energy, part of the powder melts into a liquid phase quickly. The presence of the liquid phase results in rapid sintering since mass transport can occur by liquid flow and particle rearrangement [3]. It is similar to the welding process. The energy needed to melt the metal powders is far more than that needed to melt polymer binder, as often used in SLS process. Therefore, a high laser power and slow scan velocity are used in the metal sintering. Normally, a higher laser power and slower scan speed bring a higher strength because more energy is absorbed by the loose metal powders. It results in a higher density in the parts. But over-sintering will occur when the energy is too high. The resulting properties will decrease sharply at that time. The higher laser energy will bring a bigger laser beam spot but will decrease the part accuracy. Therefore, it is important to make a trade-off between the scan speed and laser power setting. For the Cu-base metal powder sintering, the power used ranges from 80 to 200W and laser scan speed is from 80mm/s to 250mm/s.

2.3 Hatch density
Hatch density is another important process parameter. It affects all the three resulting properties. If the distance between two adjacent scan lines is larger than the diameter of the laser beam, the metal powders do not bond well together. When other parameter values are kept unchanged, a large hatch density often brings a high mechanical strength because more energy is absorbed by the metal powder. But it will take more time to process because a larger hatch distance can decrease the processing time effectively (Eq. 1). The hatch density was changed from 0.08mm to 0.3mm during the experiments.

2.4 Layer thickness
Layer thickness has a close direct inverse relationship with the total processing time (Eq. 1). It is the most important factor when the processing time is of greater concern than other resulting properties. But a thinner layer thickness could make the final parts more dense. The strength of the part, which is primarily a function of fractional density (or porosity) [4], has a reverse trend with layer thickness.

3 DESIGN OF EXPERIMENT

3.1 Apparatus and software

The experimental High-Temperature Laser Manufacturing System (HTLMS) has been developed by the NUS RP Group. A 200W CO_2 laser is employed in the HTLMS system. The laser scan speed can be set up to 4000mm/s. The developed software can perform functions such as transferring the CAD models in STL files to the layer process data required by the machine. Besides, it could combine several parts to be built on one base and based on different process requirements of each part to select the drive data. The different process parameters to be selected include the laser scan speed, laser power, hatch density and scan path. Then the operator can build several parts together on one base with different process parameter values but the value of layer thickness must be the same with these different parts. This feature could be used to save the processing time.

Table 1 Parameter setting (for training samples)

Levels Parameters	1	2	3
A: Scan Speed (mm/s)	100	150	200
B: Laser Power (W)	100	150	200
C: Hatch Density (mm)	0.25	0.20	0.15

Layer thickness=0.05mm		Layer thickness=0.10mm		Layer thickness=0.15mm	
Test No.	SS,LP,HD level*	Test No.	SS,LP,HD level	Test No.	SS,LP,HD level
#1	A1B1C1	#10	A1B1C1	#19	A1B1C1
#2	A2B1C2	#11	A1B2C2	#20	A2B2C1
#3	A3B1C3	#12	A1B3C3	#21	A3B3C1
#4	A2B2C1	#13	A2B1C2	#22	A1B2C2
#5	A3B2C2	#14	A2B2C3	#23	A2B3C2
#6	A1B2C3	#15	A2B3C1	#24	A3B1C2
#7	A3B3C1	#16	A3B1C3	#25	A1B3C3
#8	A1B3C2	#17	A3B2C1	#26	A2B1C3
#9	A2B3C3	#18	A3B3C2	#27	A3B2C3

*SS, Scan Speed; LP, Laser Power; HD, Hatch Density.

3.2 Process parameters

All these considered parameters were first set at three levels. Using the full factorial designs [5], a total of 81 test parts will need to be built. To reduce the experimental trials, one of the fractional factorial design methods, the orthogonal arrays experimental design method, was adopted. Because the layer thickness cannot be changed in one processing time, we set the other three parameters level according to orthogonal arrays with a fixed layer thickness level. The layer thickness also has 3 levels, so 27 test parts were built in this work. The setting of the parameter levels is shown in Table 1.

In addition to the 27 samples used to train the NN, five more samples were built to test and validate the training results.

3.3 Test part

According to the ASTM standard E8 [6] for the tensile testing of metallic materials, specimens (Fig. 2) were built by the machine to evaluate the resulting properties. The thickness and reduced section width at the specimen are both 6.35mm. The over-all length L is 92mm. All the size values are set following the ASTM standard. Five test specimens were built on one base each time (Fig. 3).
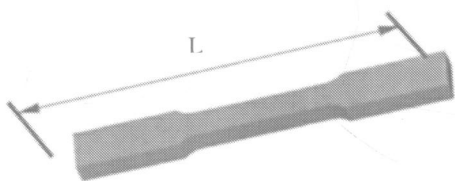
Fig. 2 CAD part model of specimens   **Fig. 3 Five parts built in one base**
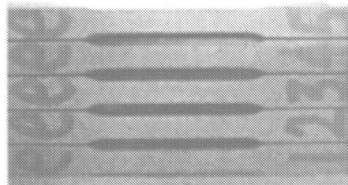
The value of the length L (Fig. 2) was measured accurately to determine the dimensional accuracy. In fact, many factors have an effect on the dimensional accuracy, such as the material, powder size, machine position error and process parameter values. Since part distortion often happens, one part size is not enough to fully evaluate the dimensional accuracy. But if one dimension is very close to the design size, most of the other sizes of these parts will show a consistent result to the design sizes. Most of the time, when the part is built, all of the part show similar distortion trend.

4 DATA ANALYSIS USING NEURAL NETWORK

A NN is a massively parallel distributed processor made up of simple processing units (neuron) [7]. The development of NN is inspired by the biological models of the human brain. A mass of interconnected neurons work together following the changing environment. Each neuron receives information from some other neurons through connection. By summing all the effects, the result will be transported to some others. All these processes work in parallel. The final outcome help to achieve the information needed. NN is a simplified mathematical model to simulate neural behavior. Following the fast development of computer technology, intensive calculations are no longer the bottleneck to NN. Since the mid-80's, many novel NN models, for examples, back-propagation (BP) network, radial-basis function(RBF) network, adaptive resonance theory (ART) network and Hopfield network, were widely applied to optimization problem, pattern classification, image processing, regression problems, simulation and so on. NN application has much success in many fields now.

4.1 BP Network

The backpropagation network is a multi-layer feed-forward network with a different transfer function based on the artificial neuron and powerful learning rules [8]. It was discovered by Parker [9] and Werbos [10] independently. Many books give a detailed description about the multi-feedforward network and BP algorithms (i.e. [7, 8]). To simulate the network, a NN software called SNNS (Stuttgart Neural Network Simulator) [11] is used for this purpose. A multi-feedforward NN with one hidden layer is built and the number of hidden layer's units is

2. Fig.4 is the multi-feedforward networks with one hidden layer used for mapping the parameters with the accuracy and strength.
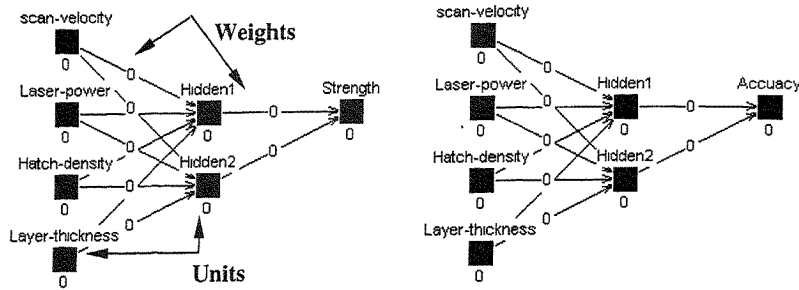
Fig. 4 Multi-feedforward networks used in the study

Using the standard BP algorithm, the nth training inputs first propagate forward by a fixed weight to give the network's output $\hat{y}_j(n)$. These paths are called forward pass. The error could be achieved by

$$e_j(n) = y_j(n) - \hat{y}_j(n),\tag{2}$$

Neuron j is an output unit and $y_j(n)$ is the corresponding target output (given by experimental data). Then the error is used to backforward layer-by-layer until it reaches the input layer. During this process, the weights will be corrected according on the delta rule [7]. Each sample data pair is used to train the network in the two-pass circulation. The NN is trained by all training samples iteratively until the error can be accepted. During the training process, the validated sample pairs do not join in the training. They are used as a measuring criterion for the training effect. All the detailed definition and calculation methods used in the forward and backward passes can be found in the SNNS user manual [11].

4.2 Training result

Unfortunately until now, no effective methods can help to decide the network structure. In practice, the training results are fed back to direct the hidden layer and the correlated unit number chosen. In this study, both the tensile strength and dimensional accuracy are trained by a multi-layer feed-forward NN with one hidden layer of two units, as shown in Fig. 4 The model performance can be evaluated by the Root-Mean-Square Error (RMSE) ν,

$$\nu = \frac{\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 / N}}{\bar{y}} \times 100\tag{3}$$

where N, denotes the total number of sample pairs, $\bar{y}$ the average value of the corresponding target outputs. The strength and accuracy of the prediction errors ν are 8.8% and 19.3%,

respectively. The comparison between the training results and desired results is shown in Fig. 5 and Fig. 6, respectively. The training outputs are well matched with the target data.
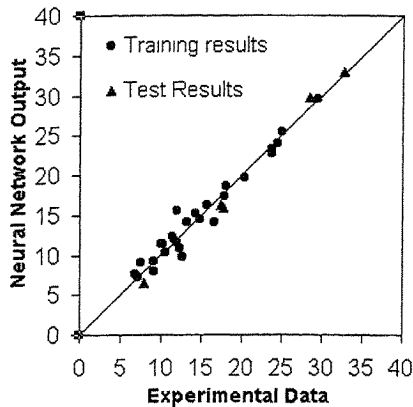
Fig. 5 Result of tensile strength (MPa)
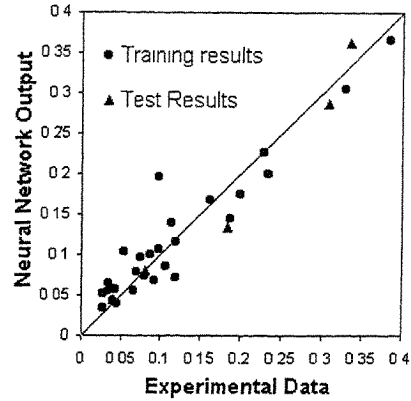
Fig. 6 Result of dimension accuracy (mm)

5 CONCLUSION

NN can provide a good mapping between inputs and outputs without many assumptions and simplifications on their relationships. These advantages make it a powerful tool to predict complicated process relationships. The proposed NN approach is able to predict the properties of the part built by DMLS based on the experimental data training. In future work, more resulting properties and parameters could be identified and considered together. By classifying the results according to different user requirements, a database can be built to help customers choose the suitable parameter setting used in DMLS process.

REFERENCE:

[1] **T.P., Bagchi,** Taguchi Methods Explained, Practical Steps To robust Design, Prentice-Hall of India Press, 1993, pp. 90.

[2] SCANLAB Gmbh "Installation and Operation, PC Real Time Control Board RTC2, for Scan Head and Laser Control in Time,", 1998.

[3] **M., Agarwala,** "Direct selective laser sintering of metals," Rapid Prototyping Journal, Volume 1, Number 1, 1995, pp. 27.

[4] **R.M. German,** Powder Metallurgy Science, Metal Powder Industries Federation Press, Princeton, NJ, 2nd Edition, 1994.

[5] **N. L. Frigon** and **D. Mathews,** Practical Guide to Experimental Design, John Willey &Sons, Inc Press, 1997, pp137.

[6] ASTM, "Standard Test Methods for Tension Testing of metallic materials", Annual Book of ASTM Standards, ASTM Designation:E8-99, 1999.

[7]   **S., Haykin,** Neural Networks, A Comprehensive Foundation,   Prentice Hall International, Inc. Press, Second Edition,1999, pp. 2 & pp. 167.

[8]   **L.M. Fu,** Neural Networks In Computer Intelligence, McGraw-Hill, Inc. Press, 1994.

[9]   **D.B., Paker,** Beyond regression, new tools for prediction and analysis in the behavioral science, Ph.D. thesis, Harvard Univ., 1974.

[10] **P.J., Werbor,** Learning logic, Technical Report TR-47, MT, Cambridge, MA, 1985.

[11] Stuttgart Neural Network Simulator (SNNS) user manual, version 4.1, University of Stuttgart.Institute for Parallel and Distributed High Performance Systems (IPVR), 1995.

Rapid prototyping of a differential housing using three-dimensional printing technology

D DIMITROV and K SCHREVE
Global Competitiveness Centre in Engineering, University of Stellenbosch, South Africa

ABSTRACT

The objective of this paper is to demonstrate the use of 3D printing (3DP) Technology as an effective tool for rapidly making patterns for prototype investment castings. This is done through the discussion of a case study. In this case patterns are made for casting housings for an automobile differential. This is a complex part that presented several challenges before a successful process chain was established, e.g. ensuring dimensional stability and coping with a part that is substantially larger than the build volume of the 3D printer from Z Corporation.

INTRODUCTION

Rapid Prototyping (RP) enables users to create physical prototypes early in the design cycle so that flaws can be detected and corrected before they mushroom into costly expenditure. Initially conceived for design approval and part verification, RP now meets the need for a wide range of applications from building of test prototypes with material properties close to those of production parts to fabricating models for art and medical applications. Furthermore, these systems allow the optimisation of functional performance through an iterative process of prototyping, testing and analysis.

This paper reflects one of the experiences gained at the GCC in the use of 3D printing technology in conjunction with investment casting. A new process chain was developed for applications in the automotive, consumer goods and packaging industries.

OVERALL CHARACTERISTIC OF THE 3D PRINTING PROCESS

The Z402 3D printing system from Z Corporation is one of the most basic RP systems available today. It utilises the principle of ink jet printing, perfected in the 1990's, along with relatively cheap and readily available standard components and materials. This combination ensures a stable and robust RP system that delivers parts, within 1% overall accuracy, at remarkable speed and low cost. For this reason, leading automotive companies, such as Daimler Chrysler, Germany, make wide use of the Z402 system concurrently throughout their design life cycle to help reduce development time and improve quality [1].

The Z402 system uses a specially modified printer cartridge to deposit a two dimensional profile of resin onto a layer of powder that forms the build surface. Similar to the Selective Laser Sintering (SLS) process the base material is rolled onto the build platform to form each new layer. The prototype is formed as hundreds of layers of powder are built on top of one another. At present there are mainly two types of powder, namely the ZP100 plaster based powder, which allows a layer thickness of 0.075 mm to be built, and a cheaper more robust ZP14 starch based

powder, which enables a layer thickness of 0.175 mm. The different materials and post processing possibilities that form part of the system also mean that there is a vast range of applications. Apart from the characteristics of the base powders, several resins can be used to harden the prototypes and provide different properties for design verification and testing. Using the Automated Waxer, parts can easily be infiltrated with surgical wax to provide strength and to form patterns for investment casting.

Since the commercialisation of the Z402 RP system, much research was done to improve various aspects of its system hardware, software and materials used. It has however been recognized that urgent opportunities for further research of this technology are needed, especially in the areas concerning the improvement of model quality. The system manufacturer gives the accuracy of the device simply as 0.5 % in X- and Y- axis and 1% in Z-axis [2]. More information regarding the accuracy, for different powder and infiltrant types, and the factors that influence it, is needed.

A recent study was conducted with the main objective to investigate how build parameters – build orientation and material - influence process improvements. Experience has proven that different build orientations, for example, provide improved model characteristics for different model designs. However, experiences need to be quantified and explained so that a platform for further increased performance can be realised. In the particular context of providing services to industry it is of extreme importance to know the real capabilities of the process in order to be able to meet customer requirements. Typical build parameters the operator can control are layer thickness, build orientation, material type, type of binder, depth of curing and the scan/hatch pattern. The choice of build orientation very often involves a trade-off between maximising part accuracy and surface finish and minimising build height and cost [3].

The following figures summarises an extensive investigation done on the issue of dimensional accuracy of the 3DP technology. More details can be found in reference [4].

Parts built with the longest dimension in a horizontal direction, using ZP100 powder (not infiltrated and infiltrated), show errors normally distributed about the zero point, indicating a sufficient random character (Figure 1, left).
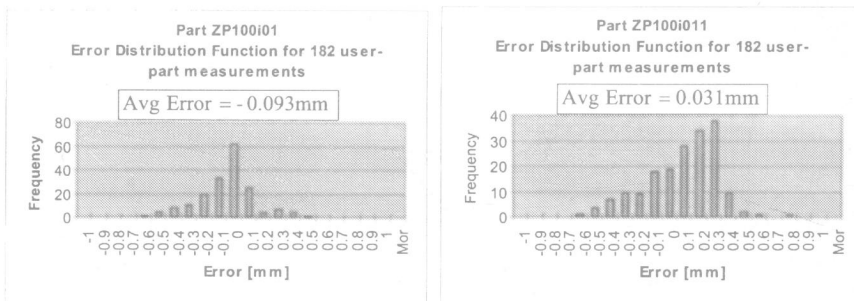
Figure 1 Parts orientated horizontally (left) and vertically (right) using ZP 100 powder.

Parts built with the longest dimension in a vertical orientation, using ZP100 powder, were found not to have a random error. Most parts show a positive bias. This clearly indicates a systematic error, for which the reason must still be identified (Figure 1, right).

For the starch based parts normality in distribution was observed for all the parts in both horizontal and vertical build orientation. This was also true for parts before and after wax

infiltration. Though the distributions were observed to be normal, the Error Distribution Function (EDF) of each part was centred largely at about 0.5mm and not zero.

CASE STUDY: DIFFERENTIAL HOUSING

Background

The objective of this paper is to demonstrate the use of 3D printing technology as a tool for making patterns for investment castings. In the course of this study, several issues were addressed, e.g. the dimensional stability and accuracy, patterns that are larger than the build volume of the 3D printer as well as establishing a process chain for this new technology. Patterns were made for investment castings for differential housings for the Ford Motor Company SA.

The Z402 3D printer used by the Global Competitiveness Centre at the University of Stellenbosch has an effective build volume of 230mm×230mm×190mm. The overall dimensions of the differential housing are 264mm×236mm×281mm, see Figure 2.
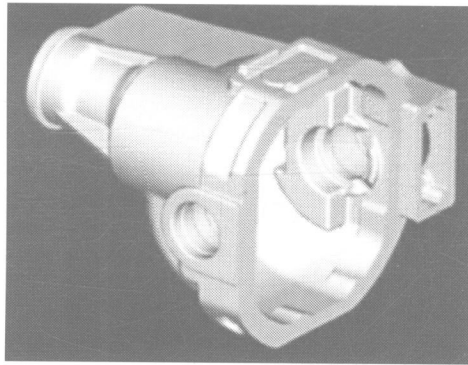
Figure 2 CAD model of the differential housing.

In order to create patterns for this part on the 3D printer, it had to be cut into smaller parts, the parts grown separately and then assembled into a complete pattern. The parts were grown in starch based powder (ZP14) and infiltrated with wax. This means that a ceramic slurry coating can be applied around the pattern. The infiltrated wax melts, and the remaining starch burns out in the investment casting process. Thus, in this process, the pattern is destroyed. Other powders used for the 3D printer give better strength and accuracy, but with the important disadvantage that they do not burn out of the ceramic cavity as easily as the starch based powder.

Three iterations were completed to arrive at a successful process chain. After the first iteration one successful casting was created, three further patterns were made during the second iteration, but the castings were unsuccessful due to a problem at the foundry. Another eight successful castings were produced during the third and final iteration.

First Iteration

With no previous experience to rely on, the differential housing was divided into three parts as shown below.
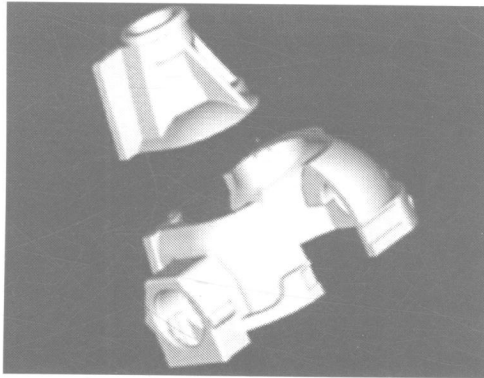
Figure 3 Three parts of differential housing.

The parts were grown, but the result was not satisfactory. Figure 4 shows the significant deformation that occurred. The wood pattern shows the desired shape where the housing was cut. Despite this severe deformation, the pattern maker was able to assemble the parts and force them to be within the tolerances by repeatedly heating the parts in the oven and then deforming them until they reached the desired shape. This is a long and tedious process and it takes the "rapid" out of Rapid Prototyping!
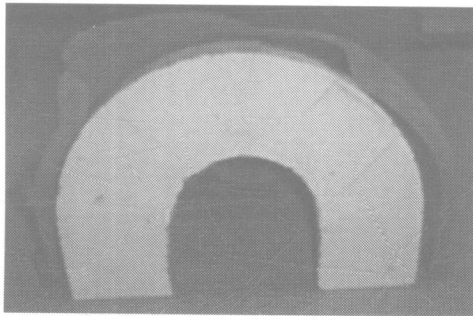
Figure 4 Significant deformation of parts. The white wood pattern is the desired shape of the cavity. Note the large gap on the right hand side.

However, a casting was made from this pattern to the satisfaction of the client.

Second Iteration

For this iteration a wooden jig was used to assemble the parts of the housing. The jig simply focused on alignment of the holes for the shafts. From a prototyping point of view, these are the only critical dimensions. However, with generous machining allowances, great accuracy was not required on this part. These parts still required all the hand work of the first iteration, but was speeded up a bit by the use of the jig.

In order to provide a flat base surface from which to assemble the parts, a rectangular piece was removed from the bottom of the housing and grown separately. This division of the housing into four parts was used for the rest of the case study.
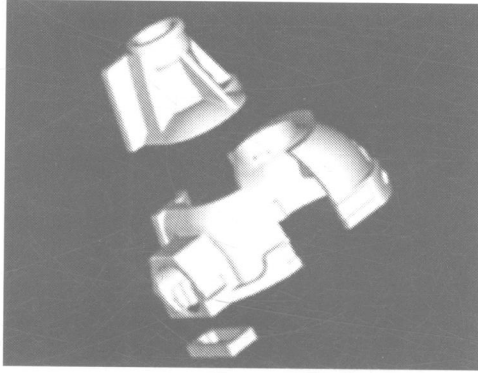
Figure 5 Removal of rectangular section at bottom of housing to facilitate assembly.

Third Iteration

It was decided to limit the amount of deformation of the parts by inserting webs at the open sections created by cutting the housing, see Figure 6. The experience with this part showed that the weight of the parts, combined with the thin wall thickness, caused the parts to sag once they are taken out of the 3D printer. Wax infiltration further increases the deformation, because a lot of wax is sucked into the heavy sections. The overhanging and uneven weight distribution then press the part open, as can be seen in Figure 4.[*] This is a significant observation, since intuition assumes that the powder surrounding the parts in the 3D printer can support them, thus support structures that are often needed with other Rapid Prototyping processes are not needed for the 3D printer. The mass distribution was further improved by removing some of the bulk material above the shaft entrance, as can be seen in Figure 6.
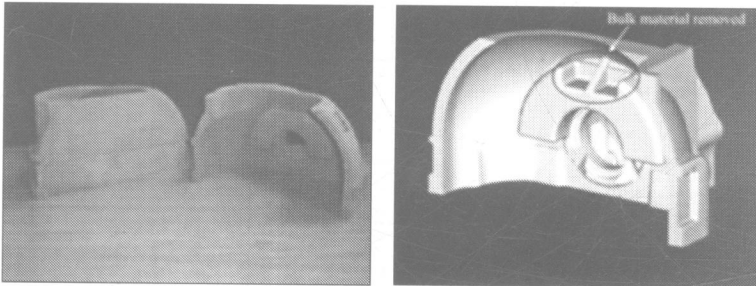
Figure 6 Webs inserted in the two bottom sections (left) and bulk material removed (right).

The webs are grown as part of the component. This has the disadvantage that the web must be removed by hand and the place where the web was connected to the rest of the part must be

---

[*] Recent experience with other projects seems to indicate that heavy, overhanging sections of open parts are not the only cause of deformation. It was observed that the open section at the end of a flat, forked part closed by a significant amount. This seems to indicate that anisotropic contraction of the part, while drying, also causes deformation. In this case inserting webs may also help.

cleaned and repaired. Note that the webs are only removed after the part was infiltrated with wax and thoroughly dried out. In this case the parts were dried out in an oven, at 70°C, for about 15 minutes before infiltrating them with wax. This minimises the effect that any moisture inside the parts might have on the accuracy and dimensional stability. (The starch based powder is very absorbent, and thus very sensitive to changes in the atmospheric humidity.) The significant improvement in dimensional stability can be seen by comparing Figure 7 with Figure 4.
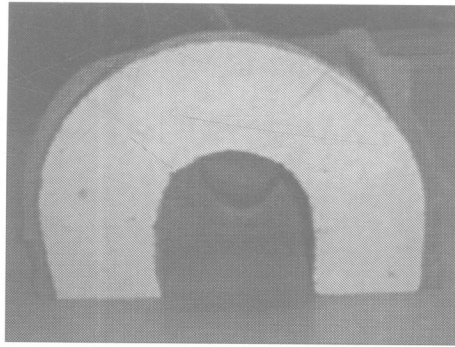
Figure 7 Comparison of the parts made with the webs. The wood template is the desired shape at the cut. Note the improvement by comparison with Figure 4.

Figure 8 shows the division of time for the production of one pattern. The total manufacturing time for one pattern was 31.4 hours. Clearly, the growing time is the most significant. Drying time and machine set up time are unproductive times about which little can be done. Ideally, for Rapid Prototyping processes, the assembly and finishing time must be as short as possible. In this case, it is a large proportion of the total manufacturing time, because the housings could not be grown as a single component and inserting the webs caused a lot of post processing work.
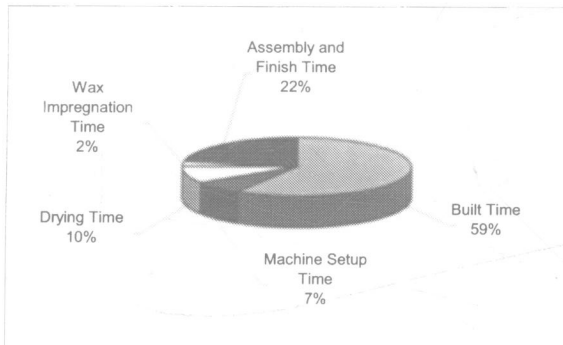
Figure 8 Division of time for the work elements in the process.

Except for a few dimensions, most of the critical dimensions were less than 1% larger than required, as shown in Figure 9. This gave a safe contraction allowance for the casting and is exactly what is required. The castings made from these patterns came out at the required cast sizes. The observed errors also closely correlate with the accuracy reported by Z Corporation (mentioned earlier in this paper).
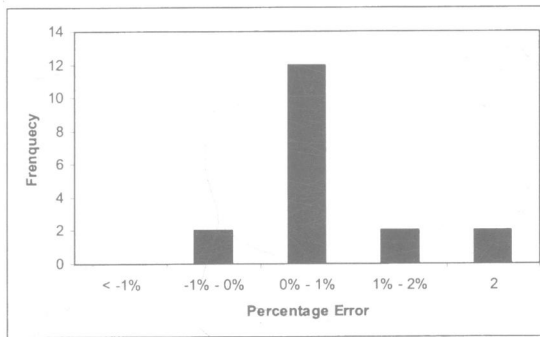
Figure 9 Frequency distribution of deviation from principle dimensions.

The complete housing (after casting and machining) and the assembled pattern are shown in Figure 10.
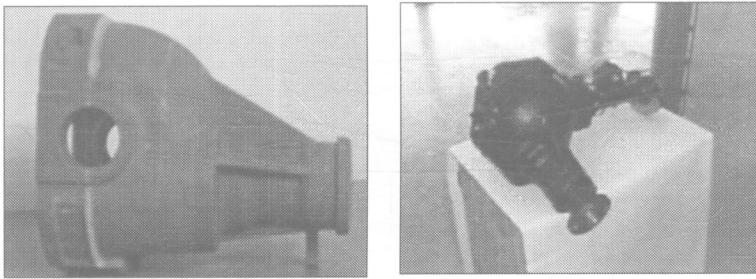
Figure 10 Assembled pattern (left) and completed casting (right).

COMPARISON WITH TRADITIONAL PATTERN MAKING

In order to compare the Rapid Prototyping process chain with the traditional route, the authors consulted a senior pattern maker. The alternative, assuming that no layered manufacturing process is available, is to make a wood pattern from 2D drawings of the differential housing. The pattern would then be used to make a core box, from which sand castings can be produced. This particular part requires an assembled core, thus increasing the complexity of producing a pattern. Based on the information available, the pattern maker quoted 200 hours for producing the pattern and one core box, allowing roughly 100 hours for making the pattern and 100 more for the core box.

The advantage of the 3D printer and the process chain described above is clearly the speed at which the first pattern, ready for investment casting, can be produced. This gives the developer a time advantage of about 4 weeks to test the first prototype. Also, the cost of producing the first printed wax pattern is about 3.7 times less that the cost of producing the pattern and core box as quoted by the pattern maker. Another advantage is that it is much quicker and easier to do another design iteration using this process if any problems turn up during manufacturing, e.g. if some features are not manufacturable, or if the functional testing of the part indicate that changes are required.

CONCLUSIONS

This case study established a process chain that is now regularly in use at the Global Competitiveness Centre at the University of Stellenbosch. Conventional wisdom dictated that the 3D printer is primarily a tool for producing models for conceptualisation purposes quickly, but this case study shows that it can also be used effectively as a tool for rapidly making patterns for investment casting. Also, the process is not limited to parts that fit in the build volume of the 3D printer. Intelligent cutting of the object and effective support of its parts make it possible to produce patterns much larger than the build volume. These patterns can be produced in less than five working days.

The very positive feedback obtained from Ford Motor Company confirms the conclusions drawn here.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the valuable insights of Mr. Peter Humphreys during the course of this project. Inserting the webs was his idea!

REFERENCES

[1]   Schell, T., Rapid Prototyping /Rapid Tooling – Universelles Werkzeug fuer das gesamten Produktentwicklung Prozess, Proceedings, 3D Erfahrungsforum Werkzeug- und Formenbau, 25-26 May, 2000, Dresden, Germany.

[2]   Z-Corporation: http://www.zcorp.com.

[3]   Pham, D.T. and Dimov, S.S., Rapid Manufacturing, Springer-Verlag, London Ltd, London, 2001.

[4]   Dimitrov, D. and De Beer, N., 3D Printing – Process Capabilities and Applications, Proceedings, RAPDASA 2nd Annual Conference with International Participation on Rapid Technologies, 14-15 November 2001, Stellenbosch, South Africa.

The rapid tooling testbed – a distributed design-for-manufacturing system

D W ROSEN, J K ALLEN, and F MISTREE
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA
Y CHEN
3D Systems Inc., Valencia, California, USA
S SAMBU
Align Technology, Santa Clara, California, USA

ABSTRACT

A new design-for-manufacturing method, called geometric tailoring, and associated digital interface concept have been developed that enables design activities to be separated from manufacturing activities. Conditions for the successful application of this method are investigated. The geometric tailoring method is demonstrated for rapid prototyping and rapid tooling technologies, where prototype parts are required to match production properties as closely as possible. This method is embodied in a system called the Rapid Tooling Testbed. Research work is presented on geometric tailoring and the distributed computing environment underlying the Rapid Tooling Testbed. Examples are summarized from the usage of this method and Testbed.

1 INTRODUCTION

One tenet of concurrent engineering and design for manufacturing is the need for early involvement of manufacturing and other groups in product development projects. In apparent violation of this tenet, there is a push within the rapid prototyping (RP) community to separate design and manufacture activities. It is common practice to create CAD models and STL files, ftp them to service bureaus, and get physical parts delivered within days. RP technologies (such as stereolithography and selective laser sintering) enable this separation between design and manufacturing activities by virtue of their capability to fabricate complex shapes directly from a CAD or STL model.

In the mid-1990's, the US National Science Foundation created the Distributed Design and Fabrication Initiative to investigate the separation of design and manufacturing activities in the context of RP technologies. The hypothesis was that a standard interchange format for RP

processes can be developed that enables design activity to be separated from manufacturing activity and that little additional communication between design and manufacturing organizations is necessary. Under this initiative, a Rapid Tooling TestBed (RTTB) was developed to investigate this hypothesis. We created the technological infrastructure for rapid prototyping, rapid tooling (using RP to fabricate injection molds), and distributed product realization. The key question that we addressed was: *How early in the product realization process, and under what conditions, can design be separated safely from manufacture?*

The problem defined by the research question can be restated more informally as "who is responsible for design-for-manufacturing (DFM)?" DFM is often difficult for mechanical parts since significant manufacturing knowledge is required to adjust part designs to aid manufacturability by a specific process. Small design changes can cause large changes in the manufacturing process or may render that process infeasible. However, if the manufacturer understands the purpose of a design and its functional requirements, then the manufacturer can adjust the design to facilitate manufacturing without compromising functionality.

The overall context for the RTTB is shown in Figure 1. The three main stages in the product realization process of relevance to this project include Functional Design, Design-For-Manufacture, and Manufacturing (tooling and fabrication). Design involvement is separated from manufacturing involvement, but the software tools and information formats for the design and manufacturing organizations overlap. The key question of the project relates to the timing of the transfer from design to manufacturing and the scope of the design for manufacture stage.
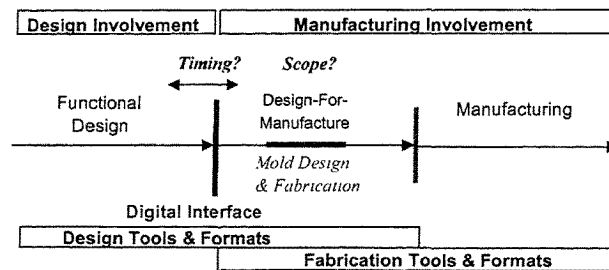
Figure 1 RTTB Process and Digital Interface.

Overall, our approach to answering the key question was to experiment with different timings of design-to-manufacture transfer and by scoping the DFM tasks differently. We separated design and manufacturing activities by requiring that the designer transfer as much information about the design, its requirements, and design freedom as possible, then allow the manufacturer to perform DFM, process planning, and manufacturing. The digital interface between designer and manufacturer denotes the information package that was transferred to the manufacturer. We experimented with STL files, CAD models, and design decision formulations with integrated CAD and FEA models as the digital interface. Design freedom refers to the design's attributes that the manufacturer can adjust, and the ranges of the adjustments, in order to perform DFM.

The research work consisted of four main thrusts: (1) Development of a Fabrication Process Compiler to translate product design descriptions into fabrication process plans (RP-produced molds and molded components), (2) RP Characterization to encode process characteristics for use in the Compiler, (3) Injection Molding to characterize polymer and powder molding for use with the Fabrication Process Compiler, and (4) Distributed Computing to provide the

computational environment for distributed design and fabrication. Thrusts 1 and 4 will be presented in this paper, as will a summary of the overall RTTB experiments.

2 FABRICATION PROCESS COMPILER

The objective of this thrust was to translate a product design description into fabrication process plans, including process plans for any required tooling, that result in production-representative prototype parts. The primary tasks of the Fabrication Process Compiler (FPC) were to select materials and fabrication processes and to design the processes and tooling. We followed the Decision-Based Design paradigm in which the progress of a design implementation is determined by the types and nature of decisions that are being made.

Process and Material Selection: This was accomplished by extending the Selection Decision Support Problem (DSP) (1). It is especially suitable for solving problems involving coupled selection. The essence of the Selection DSP is to rate a set of alternatives against a set of attributes, then rank-order the alternatives. In this work, alternatives are candidate materials and fabrication processes, while the attributes are derived from product requirements. Two major lines of research were conducted on material and process selection: (1) extensions of the Selection Decision Support Problem (DSP) to include target-matching and uncertainty, and (2) a utility-theory based selection method. In many prototyping problems, it is necessary to match target values of the attributes. For example, it may be important for a prototype part to match the mechanical properties of a production part. This can be achieved only if the production part's mechanical properties serve as the target values for the selection attributes. The target-matching selection DSP formulation is a hybrid of the selection and compromise DSP (2). The second major effort focused on adopting a more rigorous decision making method anchored in utility theory. This also enabled much better models of designer preference through utility functions. A Utility-based Selection Decision Support Problem (U-sDSP) problem formulation and solution method were developed and tested as part of this project (3,4).

After appropriate materials and processes are selected, DFM activities are typically performed in order to "tweak" part designs to improve manufacturability. We developed DFM methods, called **geometric tailoring**, for both part designs and mold designs. To support geometric tailoring, we developed stereolithography (SL) process planning capabilities. Also, limited injection molding process design was also accomplished. Each aspect is discussed below.

Fabrication Process Design: After a part or mold is designed, it is necessary to design the fabrication process to be utilized to make the part or mold. Our approach is to develop experimental and analytical models that relate process variables to measures of part quality.

Based on experimental work, fabrication process design methods were developed that enable the selection of appropriate process variable values to achieve build goals of accuracy, surface finish, and build time. The compromise DSP was used as the problem formulation. We developed a process planning strategy that consists of three stages: orientation, slicing, and parameter. Trade-offs among build time, accuracy, and finish are made in each stage so that only the most appropriate process plans are passed to the next stage. Three generations of SL process design software tools were developed (5,6). The latter two versions incorporated a new adaptive slicing algorithm.

Mold Design: Given that the input to the RTTB will be a part or product design, it must be possible to design a mold for the part(s) that can be fabricated using RP techniques. We

developed a library of mold insert CAD models (ProEngineer, SolidWorks) with suitable mounting hole, ejection hole, gate, runner, and sprue features. After fabrication, the inserts fit into standard mold bases for injection molding parts. The types of rapid tools (molds) that we have studied include (1) solid SL inserts, (2) SL shelled inserts that are backfilled with epoxy or low-melting-point metals, and (3) epoxy inserts that are cast into rapid prototyped dies.

We developed a new method of automated mold design that was suitable for simple two-piece molds (consisting of core and cavity), as well as molds with many additional moving sections (7). In our region-based approach, part faces are partitioned into regions, each of which can be formed by a single mold piece. By seeking to minimize the number of mold pieces, different partitions of faces into regions are explored until the smallest number of regions is found. During this process, a linear programming problem was adopted for finding a satisfactory parting direction of a region. CAD models of molds were constructed from the region information. Determining how to split multi-piece molding inserts manually is a tedious process that can take a long time. A new approach based on "reverse glue" operations was developed. "Glue" operations are known more formally as Euler operations in solid modeling. The key to the reverse glue operation is to generate "glue faces" that span the space between parting lines and the inner part cavity. Algorithms were developed to generate glue faces and to construct CAD models of molds from these glue faces (8).

Geometric Tailoring Design-For-Manufacture Method: The purpose of DFM is to ensure that parts can be manufactured to meet their requirements in the most cost and time-effective manner. To accomplish this, it is often necessary to modify part designs to better facilitate the selected manufacturing process. This is especially true when building prototypes that are to be representative of production parts. For example, a prototype gear train is to be fabricated in a different material and with a different process than the production gear train. Further, assume that the prototype gear train will undergo functional testing and that gear tooth stress is of concern. Then, gear dimensions must be modified so that the prototype gear teeth have the same stress characteristics as the production gear teeth under similar operating conditions. The process of modifying prototype parts' dimensions is called "geometric tailoring." For the most part, it is necessary to compensate for differences in material properties (between production and prototype materials) when fabricating prototype parts in RP technologies. The situation is more complicated when injection molding prototype parts, since there are two manufacturing processes to be considered: the process to fabricate the mold, and the molding process itself. Both processes must be considered when performing geometric tailoring.

Material-Process Geometric Tailoring for RP: Consider that a functional prototype is needed to test a property of the production part. Using concepts from similitude (e.g., Buckingham Π theorem), the 'property of interest,' X, for a prototype part and a production part can be formulated as a function of a set of part dimensions (9). Part dimensions can be divided into two categories based on their effect on X. Assume dimensions $d_1, d_2, ..., d_k$ have significant effect on X, while dimensions $d_{k+1}, d_{k+2}, ..., d_n$ have negligible effect. The principle of the approach developed in this research is to change model (prototype) dimensions $d_{m,j}$ (for $k+1 \leq j \leq n$) such that $d_{m,i}$ (for $1 \leq i \leq k$) match as closely as possible production dimensions $d_{p,i}$ ($1 \leq i \leq k$) based on design and process goals and constraints. These dimension changes are referred to as *geometric tailoring*. Continuing the gear example, the face width and diametral pitch dimensions would have a significant effect on gear tooth stress, while the diameter of the hole for the shaft would have a negligible effect.

For RP-produced prototype parts, the problem formulations are called Material-Process Geometric Tailoring / Rapid Prototyping (MPGT/RP). These problems are the result of

combining a problem that captures functional requirements and a problem that captures manufacturing and material capability. The fabrication process design work was used for the latter problem. All problems here are based on the Compromise DSP formulation (10,11).

Material-Process Geometric Tailoring for Rapid Tooling: For the case where prototype parts are injection molded, additional considerations must be included, specifically, the variation of molded material properties. The material and mechanical properties of SL molded parts can be different from those of production parts (12). For example, tensile modulus and strength of parts molded in SL are typically lower than for parts molded in steel, while flexural properties are enhanced. These mechanical property differences cause prototype parts to have different behaviors than production parts in their intended in-use situations. The MPGT/RT problem compensates for these mechanical property differences. It combines three problems: one for functional design, one for SL fabrication process design of the molds, and one for the design of the molding process. The situation is illustrated in Figure 2.

MPGT/RP and MPGT/RT Solution Procedure: The solution procedure is based on the Robust Concept Exploration Method (13). The MPGT problem must be decomposed into smaller subproblems that are simpler to solve. Typically, the evaluation of goals and constraints that model functional requirements requires finite-element models, which can be time-consuming to solve. Rather than embed FEA in an optimization loop, surrogate (approximate) models of the part's functional behaviors are computed using Design of Experiments and Response Surface Methodology techniques (14). Additionally, SL process planning can require substantial computing times due to the need for adaptive slicing of CAD models of parts or molds. As a result, surrogate models of manufacturability for different part dimension values are built to further decouple the problem. Coupling also results from the dependence of both part dimensions and process settings on project cost and time. As a result, the MPGT/RT (or MPGT/RP) problem is decomposed into two subproblems: Modified RT-PP – performs SL process planning, and Modified MPGT/RT – performs geometric tailoring of parts and molds using surrogate models of functional requirements (from FEA) and manufacturing capability.
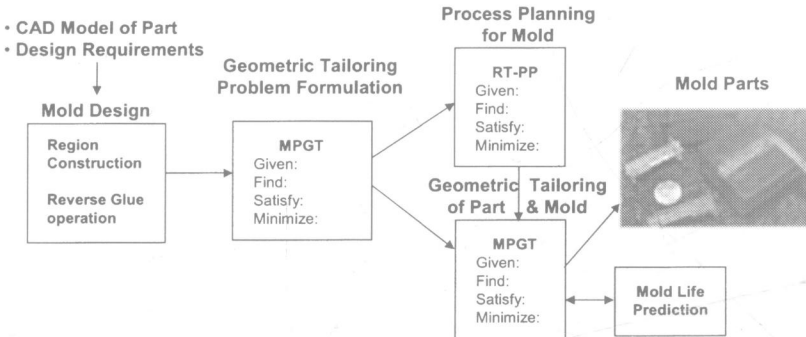
Figure 2 Work flow of SL rapid tooling process.

The results obtained from Modified RT-PP problem are used to solve the Modified MPGT/RT problem. Modified RT-PP results are used to build surrogate models of process capability for various settings of part design parameters. Additionally, surrogate models of functional goals (stress, deflection, etc.) are constructed by performing design-of-experiments using FEA experiments. The Modified MPGT/RT problem is solved for each solution obtained from Modified RT-PP problem and a selection (based on the objective function value) is performed

to determine the best of the obtained solutions. OptdesX with a simulated annealing (SAN) algorithm is used to solve the problem. The calculation of surrogate models, deviation variables and objective function are performed in C++ code integrated with OptdesX.

3 WEB-BASED DISTRIBUTED PRODUCT REALIZATION ENVIRONMENT

A distributed computing environment is essential for the implementation, testing, and deployment of the Rapid Tooling TestBed. Research efforts were focused on two major aspects: development of a suitable computing framework, and modeling the information that flows through the framework to enable design and fabrication. Three versions of distributed computing environments were developed, along with three methods of information modeling.

The third computing framework was called Web-DPR, a web-based Distributed Product Realization environment (15). Web-DPR enabled users to interact with product models, perform geometric tailoring, and explore the effects of changes in project requirements through web browsers. Communications between agents in Web-DPR occurred using Events that were broadcast through Event Channels. However, instead of encapsulating message, control, and information within an Event, Events contained only message/control information, while application content was routed through a separate data flow. Message information was encoded using XML. Interoperability of distributed objects was accomplished using Java-RMI. The Web-DPR framework is shown in Figure 3. Note the separation of message flows (through Event Channels) and data flows (through the Data Vault).
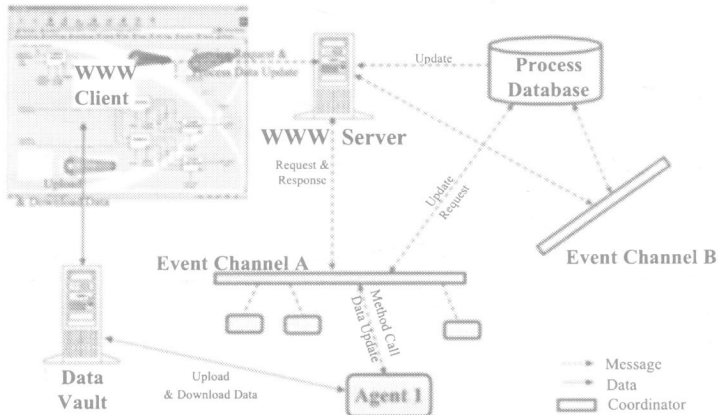
Figure 3 Web-DPR Framework.

Another addition to Web-DPR was the capability to identify suitable distributed computing resources (agents) throughout the web. An information model of capability descriptors was developed to represent the capabilities of agents, and implemented in XML. Agents were broadly classified as Analysis, Selection, or Synthesis resources. Capability descriptors contain information regarding input and output parameters required and the design freedom associated with them, a brief description of their solution strategy, and estimates of cost and time involved in the usage of the resource. A utility theory based selection procedure was incorporated into Web-DPR to perform the selection of agents for a particular usage scenario.

4 RESULTS

We conducted geometric tailoring experiments with many parts, five of which are summarized in Table 1. For four of those parts, geometric tailoring worked very well, as it was possible to improve the performance characteristics of prototype parts, relative to the production parts' designs. In the fifth case, not enough design freedom was available to enable a significant change in part performance. Process selection early vs. late indicates the extent of design changes that the manufacturer could perform. For geometric tailoring purposes, we assumed that parts would be fabricated on a SLA-3500 machine in CibaTool SL7510 resin. Molded parts would be produced using SL tools in polystyrene.

Material-Process Geometric Tailoring for RP: Most designs tested showed significant improvement in matching production-like performance characteristics. We could integrate effectively design and manufacturing models using the compromise DSP formulation. The integrated problem formulation was effectively decoupled into a RP process planning problem and a geometric tailoring problem. Design-of-experiment and response surface methods worked well in modeling the coupling between the problems. We did find, however, that the response surfaces did not always fit the design space well, but the fit could be improved if the design space size was reduced. For the simple robot arm (Expt. 3), a two-stage solution procedure was used. In the second stage, a smaller design space was formulated, based on the results of the first stage. A significant improvement in prototype performance was achieved. For the truss robot arm, shape optimization was performed using truss element diameters as design variables. Objectives included maximizing stiffness and minimizing weight.

In one GT problem (Expt. 2), significant improvements in prototype performance could not be achieved. This is because the design variables could not be modified without violating other design requirements. The conclusion here is that geometric tailoring success depends on providing the manufacturer with sufficient design freedom to enable some DFM.

Material-Process Geometric Tailoring for RT: The MPGT method for rapid tooling was applied to three parts, one gear in the gear train, simple robot arm, and camera roller (Expts 1, 3, 5 in Table 1). In these cases, the experiments were very successful. The injection molded robot arm parts were tested for their strength, stiffness, weight, surface finish, accuracy, and mold life. Strength and stiffness were improved while meeting most of the other requirements. Draft angle of the mold was modified in order to achieve the mold life objective.

For the camera roller, both geometric tailoring and configuration design were performed by the manufacturer. That is, the number and arrangement of rib and slot features were modified, as were their dimensions. This indicates an early transfer from design to manufacturing, before the designer specified a lot of design details. Finite element analysis results showed an improvement in stress and deflection performance of molded parts.

WEB-DPR: The framework efficiently delivers message and data to the appropriate agents. Details of the complex information communication activities are hidden from users. The separation of message and information flows is beneficial for two reasons. First, it enables the usage of a standard Event class. Second, it greatly reduces network traffic since data files are not routed through the main web server. It was easy to construct and add new agents, using the standardized Agent Templates, in order to extend the distributed environment.

The examples, experiments, and Web-DPR are more completely presented in Refs 10, 11, 15.

Table 1 Major Examples and Experiments.

	1. Gear Train	2. Light Switch Cover Plate	3. Simple Robot Arm	4. Truss Robot Arm	5. Camera Roller
Geometric Tailoring Status	Yes, Part and Molded Part	Unsuccessful	Yes, Part and Molded Part	Yes, shape optimization	Yes, Molded Part
Timing of Process Selection	Late	Late	Late	Late	Early
Molded Part	Yes	No	Yes	No	Yes

5 CONCLUSIONS

From our activities and experiments, we found the following:

- Communicating a part's nominal geometry was insufficient to enable the manufacturer to design the RP process plan to meet designer requirements. If only CAD or STL models are transferred, the designer must complete the part design, including DFM.

- By communicating tolerance and surface finish requirements, the manufacturer can design a RP process plan to attempt to meet as many of these requirements as possible. By communicating designer preferences among time, cost, accuracy, and surface finish, the manufacturer can better meet designer requirements by exploring trade-offs among various process plans. This effectively answers the key question from the Introduction.

- Geometric tailoring can be performed effectively, provided that the designer provides sufficient design freedom to the manufacturer. This applies to both geometric tailoring for RP as well as for Rapid Tooling.

- The concept of a digital interface appears to be a promising construct with which to define communication paths between organizations. This work demonstrated that selection and compromise decision templates can serve as the interchange format and can capture functional requirements, design freedom, tolerance and surface finish requirements, and project constraints (time and cost).

- Distributed design and fabrication requires a computing environment that enables participants to share information and collaborate. Results demonstrate that a web-based environment can effectively integrate distributed and heterogeneous computing resources (hardware and software) for engineering design and manufacture using client/server architectures. Both synchronous and asynchronous collaboration can be accomplished in such an environment. XML is useful as a standard for baseline data exchange among computing agents in this environment. This demonstrates a working computing environment that supports the separation of design and manufacturing activities.

ACKNOWLEDGEMENT

REFERENCES

1 Mistree, F., O. F. Hughes and B. A. Bras (1993). "The Compromise Decision Support Problem and the Adaptive Linear Programming Algorithm." *Structural Optimization: Status and Promise.* Washington, D.C., M. P. K. (Ed.), 247-289.

2 Herrmann, A., and Allen, J.K., "Selection of Rapid Tooling Materials and Processes in a Distributed Design Environment," ASME Design for Manufacturing Conference, September 12-15, 1999, Las Vegas, DETC99/DFM-8930.

3 Fernández, M.G., Seepersad, C.C., Rosen, D.W., Allen, J.K., and Mistree, F., "Utility-Based Decision Support for Selection in Engineering Design," ASME Design Automation Conference, September 9-12, 2001, Pittsburgh, PA, DETC2001/DAC-21106.

4 Rosen, D.W. and Gibson, I., "Decision Support and System Selection for RP," Book chapter in *Software Solutions for RP*, PEP, Ltd., UK, 2002.

5 Lynn-Charney, C.M. and Rosen, D.W., "Accuracy Models and Their Use in Stereolithography Process Planning," *Rapid Prototyping Journal*, 6(2):77-86, 2000.

6 West, A.P., Sambu, S., and Rosen, D.W., "A Process Planning Method for Improving Build Performance in Stereolithography," *Computer-Aided Design*, 33(1):65-80, 2001.

7 Chen, Y. and Rosen, D.W., "Problem Formulation and Basic Elements for Automated Multi-Piece Mold Design," ASME Computers and Information in Engineering Conference, paper #DETC2001/CIE-21293, Pittsburgh, Sept. 9-12, 2001.

8 Chen, Y. and Rosen, D.W., "A Region Based Approach to Automated Design of Multi-Piece Molds with Application to Rapid Tooling," ASME Computers and Information in Engineering Conference, paper #DETC2001/CIE-21294, Pittsburgh, Sept. 9-12, 2001.

9 Cho, U., K. L. Wood and R. H. Crawford, 1999, "Error Measures for Functional Product Testing," *ASME Design for Manufacturing Conference*, Paper #DETC99/DFM-8913, Sept. 12-16, Las Vegas.

10 Chen, Y., "Computer-Aided Design for Rapid Tooling: Methods for Mold Design and Design-for-Manufacture," Ph.D. Dissertation, Georgia Institute of Technology, 2001.

11 Sambu, S.P., 2001, "A Design for Manufacture Method for Rapid Prototyping and Rapid Tooling," Masters Thesis, Georgia Institute of Technology.

12 Dawson, K., 2001, "The Effects of Rapid Tooling on Final Product Properties", Ph.D. Dissertation, Georgia Institute of Technology, Atlanta.

13 Chen, W., Allen, J. K., and Mistree, F., "The Robust Concept Exploration Method for Enhancing Concurrent Systems Design", *Concurrent Engineering: Research and Applications*, vol. 5, no. 3, pp. 203-217, 1997.

14 Myers, R.H. and Montgomery, D.C. (1995) *Response Surface Methodology: Process and Product Optimization using Designed Experiments*, John Wiley & Sons, New York.

15 Xiao, A., Choi, H-J, Kulkarni, R., Allen, J.K., Rosen, D., and Mistree, F., "A Web-Based Distributed Product Realization Environment" ASME Computers in Engineering Conference, September 9-12, 2001, Pittsburgh, PA, DETC2001/CIE-21766.

Internet-based Systems

The virtual design system for individualized product based on Internet

T ZHENG and **Y HE**
Department of Mechanical Engineering, Chongqing University, China

ABSTRACT: Recently customers have taken high demands for the individualization of the product. Meeting the market demand has been the key factor to the enterprise's survival facing to drastic competition. So it is very important for an enterprise to develop product according with customer's demands. This paper puts forward the virtual design system for individualized product development based on internet. This system has the structure of browser/server. And it is developed in VRML. Customers can design products in the virtual environment. Enterprises can develop individualized products according to the virtual product. It can assist the enterprise in its efforts to rapidly and effectively react to the changes in market.

KEY WORDS: virtual design, VRML, individualized product

1 INTRODUCTION

With the improvement of living environment and living standard, customers take high demand for product. A customer wishes that the product not only meets the functional demand but also satisfies the aesthetic demand. The product must represent a kind of culture and a kind of sensibility. It is said that customers are pursuing for the individualization of product. Traditional enterprise's rigid design mode lacks flexibility. The developed products always lag to the market's demand. Last year China joined in WTO, the impost will be lowed down, and national enterprise will be impacted with the rush into of international product. How to reply the international impact and find a path in drastic competition has been an urgent problem for national enterprise. It is very useful for enterprise to develop individualized product and occupy the market rapidly that the customers take part in product design and exert their imagination fully. Fortunately, VRML provide us with the ability to construct interactive virtual environment to realize the individualized product development.

VRML (Virtual Reality Modeling Language) is used to describe virtual scene. It has been the standard format for the transmission of VR image on internet. We can construct interactive

three-dimensional virtual environments using it. It has clear structure, abundant function of construction and alternation. It supplies the methods to construct point, curve, surface and solid. And more it supplies the manipulation of light source, material feature, texture and so on. It can also realize rotation, scroll, cruise and so on. In addition VRML provide the function of selection and collision detection. In a word it provides us with abundant tools to construct virtual environment. As discussed above, VRML not only suits for the transmission of VR image on internet but also can construct interactive 3D virtual environment. So we choose VRML to develop the system.In this paper we will discuss the virtual design system based on internet.

2 THE SUPPORT OF VRML TO THE SYSTEM

2.1 The features of VRML

VRML describe three-dimensional scenes by various objects. These objects are called nodes. Each node has one or more fields. The standard of VRML97 has defined most functions in 3D applications [1,2]. They are briefly listed below.

1. VRML provide abundant nodes of modeling, orientation, group and so on;
2. Third dimension and render ability. It provides abundant render nodes to realize light, color, texture and three-dimensional sound;
3. The means of observe and alternation;
4. Animation: VRML supply convenient control manner of animation;
5. Level of details and collision detect: LOD, Collision;
6. Hyperlink and embed: Anchor, Inline.

2.2 The requirement of system

The essential of virtual design system is to construct a virtual environment suitable for people to create virtual model. This system involves two kinds of information. They are environment model and product model. The information is digital object. The environment model includes the geometrical model of environment, video information, feeling signal and so on. The product model includes the geometrical model of product, kinematics model of product, physical model of product and so on. A virtual design system should have the ability to describe environment information and product information. As described above the abundant modeling methods of VRML not only fit the demand of CAD for three-dimensional modeling but also fit for the communication of 3D information on internet. So VRML is suitable to virtual design based on internet. It can create the environment for customers to take part in product design.

3 THE STRUCTURE OF THE SYSTEM

3.1 The type of virtual design system

Image output is very important for virtual design system. According to its device of image output the virtual design system is classified into two classes: immersed virtual design system and desktop virtual design system. The immersed virtual design system has high demand for

hardware, and its cost is high. Head mounted display should be equipped. The desktop virtual design system has low demand for hardware. Display or raster glasses can meet its demand. Because the investment of desktop virtual system is acceptable to most enterprises and virtual reality based on VRML can satisfy the demand of virtual design, this system is designed as desktop virtual design system.

3.2 The structure of the system

In order to maintain all databases conveniently, the system adopts the structure of browser/server. It is said that the databases are up to the enterprise to maintain, so that all databases can be updated in time. For example, new style product or new material can be stored in database directly. Its structure is shown in figure 1. All databases are connected with the server. The client is equipped with data glove or 3D mouse. The detailed function is discussed below.

Client: The system of virtual design connected with server through internet supply the tools for design product. The devices of input and output are connected with it. The software of virtual design is installed in the client. Customer can design virtual product in the virtual environment.

Server: The server is the World Wide Web server. Its main task is to accept the request from client and make response to client.

Database of environment: This database stored kinds of the model of adornment, such as television set、 flowers、 fog 、 curtain and so on.

Database of light: This database stores the models of light, such as headlight、 reading lamp、 droplight、 daylight lamp、 wall lamp and so on..

Database of material: This database stores kinds of material, color, textures and so on.

Database of product: This is the place store the models of existing products which are exported from the models of product created by CAD. Customers can choose these models or design new product referring to them.
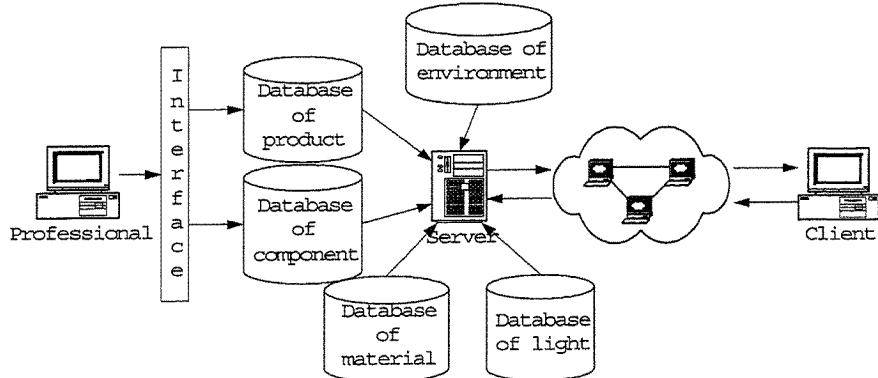
Figure 1. The structure of virtual design system based on VRML

Database of component: This database stores the models of component which are used to construct product. The models of component are also exported from the models of component created by CAD. The components are convenient for user to design new product.

Interface: It is the interface through which the model created by CAD can be translated into VRML.

Professional: It is the CAD system used by professional.

All materials the virtual design needed are stored in these databases in the format of VRML file. When needed, they are downloaded to client. And except the client, all others are located in enterprise.

4 VIRTUAL DESIGN ENVIRONMENT

Virtual design system had two basic modes: cruise mode and construction mode. Under cruise mode, user can observe the geometric models. The system provides the user with 3D image and the feedback of sound and touch. Under construction mode, user can create and modify the geometric models by the movement of hands and eyes and the command of speech. The system also provides 3D image and the feedback of sound and touch [3]. It is said that a whole virtual design system should have cruise mode and construction mode and other modes closely related to them. Correspondingly this system has two modes: cruise mode and construction mode. Under cruise mode, user could observer the product in virtual environment. Each mode included several modules. The functional configuration of the system was shown in figure 2. The detailed function of each module will be introduced.

4.1 Cruise mode

Light: In general, light is very important for user to look at the appearance of product in cruise mode. This module provides all kinds of light in the format of VRML. They are stored in server. Designer or customer can drag and drop it into virtual environment simply and light is created. We can create several light sources according to the needs, such as point light, ambient light and so on. The light can be turned on or off and its location, direction, color and intensity can be adjusted. And it can be deleted when useless or unsatisfied.

Color: It provides a palette for the user to paint the product at random.

Texture: As we know, product is made of wood, glass, steel, plastic and so on. In order that the product looks vivid, we must endow them with material features. The texture module gives us the opportunity to design product with all kinds of material. It looks as if we are artists not designers.
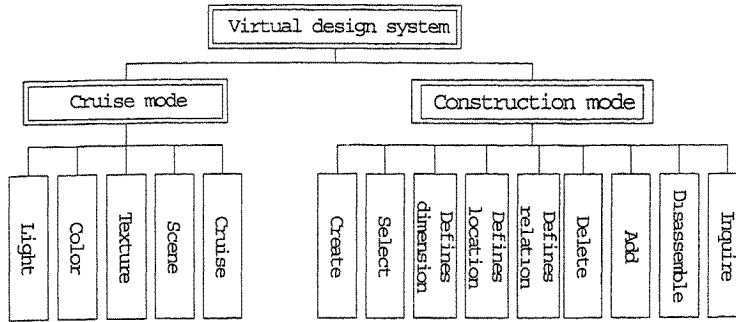
Figure 2. Functional configuration of virtual design system

Scene: We can accessorize the virtual environment with plants, TV, sculpture and so on, as if the virtual environment is just the place the product will be located in.

Cruise: The function of cruise is vitally important for virtual design. Under this state we can observe the virtual model in all direction and inquire the property of models selected.

4.2 Construction mode

Create: It is the basic function of construction mode. We can create a solid from begin, or reuse the model existing in database.

Select: It helps us to select the model we want. Only the model selected is active. And only the active model can be manipulated

Defines dimension: Generally the initial dimension of model is not proper. We can define the dimension of the model to constrain its shape.

Defines location: Perhaps the model is not in a suitable location. We can use this tool to define its exact location in WCS or UCS.

Defines relation: The product is composed of several assemblies. And assembly is made up of parts. They are assembled by certain laws. The final product is determined by the relations between the parts and assemblies. This module is used to define the relations between parts and assemblies.

Delete: This module is very simple. It is just like the "delete" function in all software. We can delete the unwanted part or assembly.

Add: We can add a component to a subassembly or a subassembly to an assembly.

Disassemble: This module used to disassemble assemblies from product or parts from assembly.

Inquire: We can inquire the features or states of model.

4.3 The realization of system

Although it is already possible to create VR models from within VR packages, for the use of VR in construction industry, the transfer of geometrical data between CAD and VR is desirable to avoid repetitive work. There are three kinds of approaches for data exchange from CAD to virtual reality: a) library-based approach, b) straightforward translation approach, c) database approach [4]. But the second kind approach is time required and not convenient for

user to interactive with the virtual environment in real-time. As for the third kind approach, it has complex structure and is difficult to maintain. So we choose the first kind approach to translate the CAD model into the VR model as shown in figure 1. As discussed above image output is very important for virtual design system. In order to promote the performance of the system we adopted the simplified VR model. It is said that the VR model translated from CAD model is optimized. Optimization involves the reduction of the number of polygons to be processed in the VR model and it is done to increase performance, by increasing the frame rate and speed of reaction to user input.

The material, color and other stored in other databases can not be translated from CAD model. They are created directly with VRML.

As for the browser side, we adopt EAI (External Authoring Interface) to control VR model. EAI is a suit of Java classes designed for VRML explorer. External environment can visit the present VRML, so it can operate, control and modify the internal object in VRML directly.

It is difficult for a non professional to design product using professional software. But virtual design system provides them with the tools to design in virtual environment. It is very effective to design individualized product. Designer can free from sterile geometric and structure design and dedicate themselves into innovation design in the virtual environment. They can exert their creativity fully. When customer is not satisfied with the product in shop, he can design the product he wanted in the virtual environment. Owing to lack of professional knowledge, he just designs the rough shape, and does not think about the detailed structure and industry standards. Then the virtual model of product is uploaded to server and then the professional perform detailed geometric and structure design according to the virtual model customer designed. At last the product is transported to customer.

5 CONCLUSION

Virtual design system based on internet had not been studied yet. This system described a rudiment of virtual design system based on internet. But it has the features of virtual design system. And the system provides abundant tools for customer to design virtual product. It gives enterprise the opportunity to know the customer's demands for individuation. Then the enterprise can develop the product meeting the demand of market in time. With the development of internet and the technology of virtual reality, virtual design based on internet will be a direction of virtual design.

ACKNOWLEDGEMENT

REFERENCES

1 **Wagner, M G, Hatekeyama**. Visualization of non-precision instrument approaches using VRML. Proceeding of SPIE - The international Society for Optical Engineering, Apr 5- Apr 6, 1999, 3691: 81~88.

2 **Carson, George S,Puk, Richard F, Carey Rikk**. Developing the VRML 97 International Standard. IEEE Computer and Graphics and Applications. Mar- Apr,1999, 19(2):52~58.

3 刘宏增，黄靖远　虚拟设计　机械工业出版社　1999,11 P30

4 **J. Whyte, N. Bouchlaghem, A. Thorpe, R. McCaffer**　From CAD to virtual reality: modelling approaches, data exchange and interactive 3D building design tools ,Automation in Construction 10 2000 43–55

Development platform for networked sale and customization systems

Y YANG, X ZHANG, F LIU, and S LIU
Institute of Manufacturing Engineering, Chongqing University, China

ABSTRACT

Based on former research work, this paper proposed to set up a development platform for networked sale and customization systems Using this platform, enterprises can develop their networked sale and customization systems effectively in short time and at low cost Firstly, the paper presented an overall structure framework of the platform and a general process on how to use the platform to develop systems Secondly, common technologies of the platform are researched including general working process and organizing model Finally, a networked sale and customization system that was developed using the platform was presented as an example

1. INTRODUCTION

The 21-century's market has been characterized by increasing individualized requirements of customers, and the developing trends of manufacturing are customized, networked and customer-centered As networked sale and customization can improve reactive abilities of enterprises powerfully, many enterprises are now eager to introduce it At present, product sale and customization systems have been widely used for customers to select or configure preferred products in an e-business environment, and the products for customization have involved in personal computers, cars, cosmetics, jewelry, houses, health / fitness services etc [1-3] Aiming at requirements of some enterprises, we have developed several systems for enterprises such as "networked sale and customization system of ceramics products", "networked sale and customization system of electronic products", and so on As a result, those enterprises benefit a lot from these systems [14] [15] And during the development process of these systems, we found that although detailed contents of these systems are different, the function module, working process, and organizing structure of these systems are similar on a large scale From this point of view, we summarized those common regularities and set up a general development platform for networked sale and customization systems, so that enterprises can develop networked sale and customization systems with their own

characteristics. The platform plays an important role to help enterprises to set up their systems effectively in relatively short time and at relatively low cost.

There have been a lot of ASPs (Application Service Provider, ASP) providing networked sale and customization application services. For example, DELL ASP e-commerce solution services aim to help companies to build their business an E-Commerce enabled website presence including customer service, selling products, etc. Iron Speed automatically generates feature-rich application programs, including the web-based user interface, database schema, access code, and workflow logic based on the data model and business rules. E-Commerce systems INC provides complete outsourced commerce management tools for direct sales channels. USi has partnered with BroadVision to provide a full complement of electronic business services including relationship marketing, and personalized electronic commerce [8-13]. However, most existing services are focused on development platform service of direct sale. Development platforms for product customization are seldom concerned. Several ASPs provide customization application systems, and these systems devote great effect to process customer requirements automatically [4-7].

We proposed a development platform for networked sale and customization system in this paper. The platform can help enterprises to develop networked sale system, as well as product customization system. As to product customization system, we provide more customization approaches according to the customer's individual demands instead of retrieving and adapting approaches. An overall structure framework of the platform and the developing process are proposed, and common technologies of the platform including general working mode, functional modules, and organization model are discussed in this paper. At the end of paper, a networked sale and customization system that was developed using the platform was presented as an example.

2. AN OVERALL FRAMEWORK OF DEVELOPMENT PLATFORM

An overall framework of the development platform is illustrated as Figure 1. Based on general models' layer, the platform is consisted of enterprises' users management template, enterprises' information releasing template, and sub-systems including product browsing and sale process definition, product customization process definition, orders management process definition, customers management process definition, delivering management process definition, user interfaces definition, product data structure definition, and external interfaces definition. Functions of main sub-systems are described as follows:
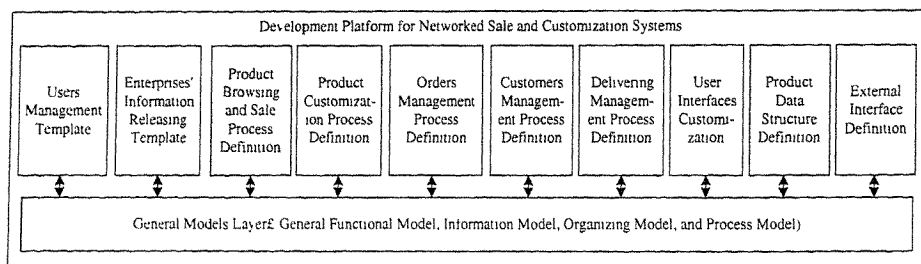
Figure 1. An overall framework of the development platform

(1) Sub-system of product browsing and sale process definition: This sub-system helps enterprises to define working processes such as product browsing, consultation, ordering, and information query of shopping cart and orders.

(2) Sub-system of product customization process definition: Product customization is the focus of a networked sale and customization system, so the sub-system of product customization is the focus of the development platform. This sub-system helps enterprises to define processes such as configuring product based on product structure or function, and collaborative design by customers and enterprises.

(3) Sub-system of order management process definition: This sub-system helps enterprises to customize working processes such as ordered-product query, order's status query, order treatment, order information statistic and analysis.

(4) Sub-system of customer management process definition: This sub-system help enterprises to customize working processes such as customers resources query, customers information statistic and analysis, customers' orders track, and customers' suggestion.

(5) Sub-system of delivering management process definition: This sub-system helps enterprises and distributors to customize working processes such as delivering task management, delivering treatment, and delivering information query, statistic and analysis.

(6) Sub-system of product data structure definition: This sub-system helps enterprises to define product data structure and information management process based on a product database of the system.

(7) Sub-system of user interfaces definition: After those development works discussed above, this sub-system provides enterprises several interface stencils to choose, helps them to customize user interfaces of their networked sale and customization systems.

Based on the overall framework, the developing process of a networked sale and customization system was illustrated as figure 2. The whole developing work is organized by a service center of networked sale and customization, which is generally the developer of the platform. In a web-based collaborative environment, the center worked with enterprises to develop networked sale and customization systems. Developing work includes function design, processes definition, database definition, external interfaces definition and user interfaces customization. After the system is developed, the center would continue to provide services for enterprises including system integration, system management, system maintenance, and technology support, in order to assure the system to work well.
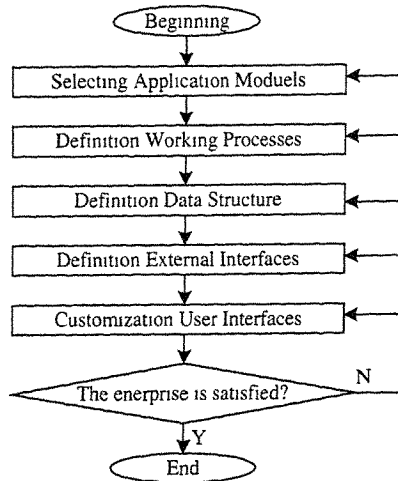
Figure 2. Developing process of networked sale and customization systems

3 COMMON TECHNOLOGIES OF THE PLATFORM

(1) General working model of networked sale and customization systems

Although requirements of each enterprise for networked sale and customization are not same completely, the working process among these networked sale and customization systems has common regularities. So an essential step for the platform development is to propose a general working model of these networked sale and customization systems, as illustrated in figure 3.
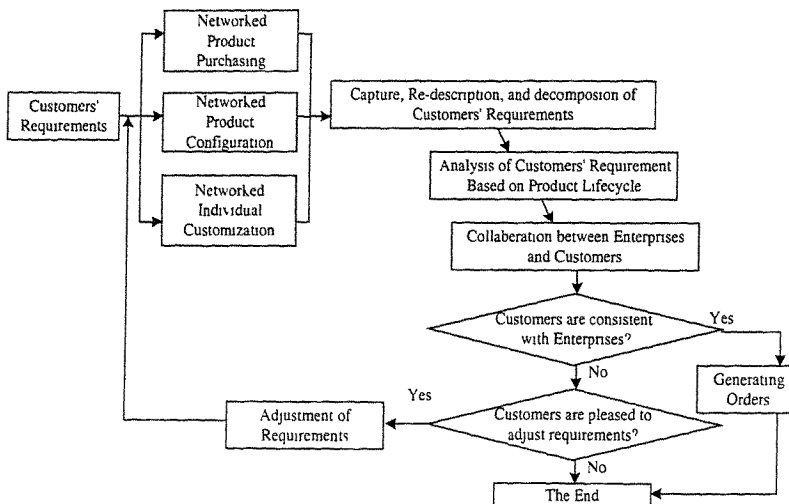
Figure 3. General working model of networked sale and customization systems

The characteristic of this model lies in integration of customer's action and enterprise's action so that the enterprise can run coordinately around customers' requirements and the customer's creativity can be fully utilized. The working process is described as follows:

a. Based on networked system, a customer presents networked order requirements, such as product's shape, function, and performance.

b. Next, the requirements should be re-described from customers' point of view to enterprises' point of view based on product's lifecycle, so that the requirements can be reflected to internal sections of the enterprise, such as product's functional modules, design parameters, requirements of manufacturing, assembling, service, delivery time, logistics, and quality.

c. After re-description, requirements should be analyzed based on an evaluating architecture towards product's whole lifecycle. The evaluating architecture is consist of five aspects which is product design, operation management, manufacturing, service, and quality, so that important considerations of the product such as designability, manufacturebility, assemblebility, manufacturing cost, and time data can be took account before the order has been confirmed. If the order was confirmed, this step can save enterprises' delivery time and product cost to a large extent because more aspects are considered beforehand.

d. Feedback messages would be generated after requirements' analysis. Sequentially, the enterprise negotiates with customer about those messages in a networked collaborative environment. Each uncertain requirement would be clarified and confirmed by both enterprise and customer. The destination of this step is trying to satisfy both customers' need such as price and enterprises' need such as profit.

e. There are two results after collaboration between the enterprise and the customer. One is that the customer is satisfied and the product order is confirmed. Another result is disagreement between the enterprise and the customer. In the latter case, the enterprise usually proposes an adjusting suggestion to the customer. If the customer would like to adjust his requirements and originalities, steps would be repeated from a to e.

Through the general working mode we proposed, the customer's requirements are transferred to internal links of the enterprise. As a result, satisfactions descent caused by requirements' illegibility is avoided. At the same time, the delay of delivery time and increasing of product cost caused by customized production are cut down also.

(2) General organization model of networked sale and customization system

Cooperated with general working mode, a general organizing model of networked sale and customization systems is illustrated as figure 4. Customers, networked sale and customization service center, multi-function team, manufacturer, providers, and distributors compose the organization. Customers are individuals or groups who have purchase intents.
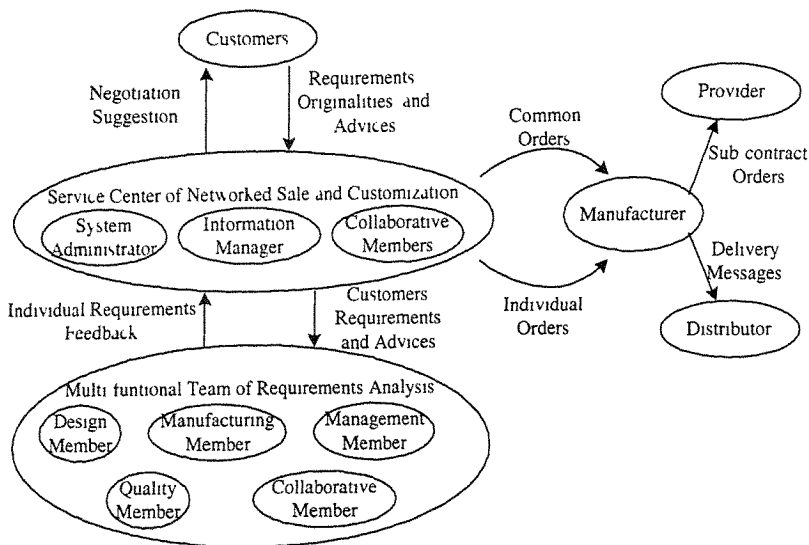
Figure 4. General organization model of networked sale and customization systems

The service center is executive unit that collaborates with customers to complement net-based product browsing, configuring and customization The multi-function team processing requirements consists of various members coming from various departments of the enterprise, working co-ordinately to re-describe and analyze customers' requirements The manufacturer arranges production based on orders generated by the networked sale and customization system Some parts of the product would be sub-contracted to providers in order to save delivery time and product cost Distributor agents locating all over the world would deliver products to customers in an optimal way Based on the general organization model, customers manufacturer, providers, and distributors are connected tightly As a result, not only customers' requirements are satisfied on a large scale, but also a networked global manufacturing system can be complemented

4 A CASE STUDY

Based on principles discussed above a platform prototype is set up As a case study, we choose furniture products of a Chinese company "Chongqing STK Furniture Co Ltd" Facing to the growing individualization of customer demands, the company has to embrace a closer reaction to the customers' needs by providing networked sale and customization service in the evolving e-business world Collaborated with the company, we developed a networked sale and customization system of furniture products base on the platform Customers can browse select, and customize those virtual furniture products in a network environment, and the system also provides functions including orders information management and customer information management The user interface of the product customization sub-system was illustrated in figure 5 Suppose a customer wants to customize the student desk and chair of a classroom He can select a prototype first and then give his individual requirements of the surface material the leg material and the size parameters such as length width and height The virtual product model can be displayed online through

Cult3D Viewer. The customer can zoom, drag and rotate the model to see if the product is desirable. If he is satisfied, he can propose his order to the system, and the customization is finished. However, if he can not be satisfied with current customization, he can propose his requirements to the system also, and the product design engineer will contact with the customer individually. The system was put into work in an ASP service way, in which we provide services of system development, data infrastructure, system maintenance and management, while enterprises use the system to benefit themselves. The system was proved to be effective after a period time of working.
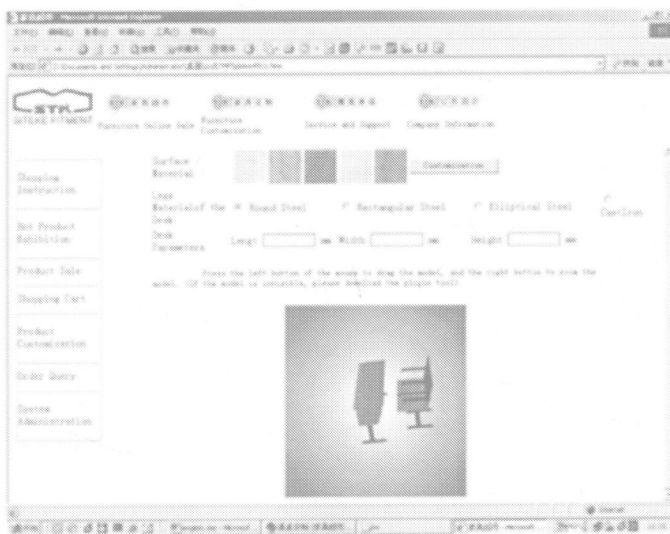
Figure 5. The user interface of the customization sub-system

5 CONCLUSION

Networked sale and customization can help enterprises optimize their operation processes focused on customers' requirements so that enterprises can fully satisfy their customers and implement global manufacturing. Therefore, the work to set up a development platform for networked sale and customization systems is quite important. From application practice, we can drew a conclusion that utilizing a general platform is a quick, effective, and low-cost way for enterprises to develop their networked sale and customization systems.

ACKNOWLEDGEMENTS

REFERENCES

1. **Rautenstrauch et al. (ed.).** Moving towards mass customization, Springer: Heidelberg/Berlin/New York, 2002.
2. **CHOI S Y, et al.** (1999) The future of e-commerce: integrate and customize Computer; Vol. 32(1), pp. 133-134.
3. **Rautenstrauch et al. (ed.).** Moving Towards Mass Customization, Springer: Heidelberg/Berlin/New York, 2002.
4. **Armin Stahl, Ralph Bergmann, Sascha Schmitt.** A Customization Approach for Structured Products in Electronic Shops. Proceedings of the 13th Bled Electronic Commerce Conference, Bled, 2000.
5. **Tseng, M.M. and Du, X.** (1998) Design by Customers of Mass Customization Products, CIRP Annals, Vol. 47(1), pp.103-106.
6. **Jiao Jianxin and Mitchell M. Tseng.** (1998) Fuzzy Ranking for Concept Evaluation in Configuration Design for Mass Customization, Concurrent Engineering: Research and Application; Vol. 6(3): pp. 189-206.
7. **Sun Microsystems Inc.** SUN Microsystems White Paper. Customization of services in internet time, 2000.
8. **WOL's ASP Solutions.** http://www.wol com/asp.html.
9. **Antarix ASP Solutions.** http://antarix.net/.
10. **DELL ASP Solution.** http://www.dellhost.com/html/ordernow.asp.
11. **Iron Speed Development Platform.** https://exchange.buckaroo.com/Home/Home.asp.
12. **E-Commerce systems INC.** http://ecommerce.ecommercesys.com/.
13. **E-Business Powered by BroadVision.**
    http://www.usi.net/solutions/ecomm/broadvision.html.
14. **F, Liu, X. D. Zhang.** Manufacturing System Engineering, Defence Industry Publishing, China, 2000.
15. **F, Liu, Yin, C.,Liu, S.** (2000) Regional Networked Manufacturing System, Chinese Journal of Mechanical Engineering; Vol 13, Iss S, pp. 97-103

Collaborative part manufacturing via an online e-service platform

P JIANG, Y ZHANG, and H SUN
School of Mechanical Engineering, Xi'an Jiaotong University, Shaanxi, China

SYNOPSIS

This paper describes an e-service platform to implement the web-based online part manufacturing in the form of synchronously collaborative work. Java solution is used for constructing the correspondent systematic architecture based on the three-tier browser/server mode. Under the support of this architecture, a working logic of the online part machining is presented. Furthermore, some of the key issues, including selecting a feasible manufacturer, queuing a manufacturing task, visualizing the manufacturing process, and using the synchronously collaborative work environment, are described in details. At the same time, the discussions about the application range and the supplier involvement are given. Finally, conclusions are drawn accordingly.

1 INTRODUCTION

Globalization is a kind of new trend for enterprises. At present, traditional mass production mode is being replaced with current mass customization one. The major factor to go to the success in global market will greatly depend on the agile and rapid response speed to the customers' requirements. In addition, customers' direct participation during the design and manufacturing procedure is becoming popular. In fact, Internet just provides a computing environment for customers, manufacturers, and suppliers to be able to work together on Web.

For a long time, we have thought of something in the stage of after-manufacturing when we talk about the concept of e-service. Actually, e-service can be introduced to the in-manufacturing phase if we consider a design or manufacturing activity as a group of service operations. So it is possible to use an e-service platform for the online part manufacturing on Web under the premise of being sure that customers, manufacturers, and suppliers can work in synchronously collaborative mode. Starting from this point of view, in the paper, authors construct a systematic architecture and study the correspondent enabling techniques.

Furthermore, a prototype system based on above the idea was developed to verify this approach.

2 SYSTEMATIC ARCHITECTURE AND WORKING LOGIC

2.1 Defining Concepts
Before describing the systematic architecture, we have to define following concepts for the best understanding of upcoming sections.

Customer: a person role who submits a manufacturing task to a special site for fabrication.
Manufacturer: a person who locates the front-end in a manufacturing site and has a capability to provide the special manufacturing services for customers.
Supplier: a person who provides outsourcing parts for manufacturers. During the part manufacturing service procedure of this approach, supplier only plays a role in giving an advice for the feasible manufacturing plan or the collaboration to solve some troubles.
Administrator: a person who possesses the e-service platform and uses this platform from the angle of information confirmation and role-based privilege management.
Manufacturing Facility: a kind of tool workers can operate. It makes the part shaped partially or completely with cutting, net-shape fabrication, rapid prototyping manufacturing, etc.
Manufacturing Cell: a kind of logic and physical layout according to special production demands, which consists of several manufacturing facilities, the correspondent software, workers, technicians, and managers.
Manufacturing Site: a kind of combination of the manufacturing cells.

2.2 Implementing the Architecture and Correspondent Functions
The e-service platform for online part manufacturing uses three-tier browser/server mode as its information architecture. As shown in Fig.1, there are a Java-enabled web server and the correspondent web database in the server side. Resource information and functions are integrated into the server-side of this architecture under the supports of fundamental technologies like network technology. The correspondent software implementations are just a combination of Java applet-servlet pairs, HTML files, JDBC connections, and a large number of resource data stored in the Web database. Here, the resource information integration focuses mainly on describing, handling and storing
● machining messages of ongoing manufacturing tasks,
● configuration information of manufacturing site,
● manufacturing capability information,
● customers' manufacturing task information,
● manufacturers' business-related information, and
● customers' business-related information.

The function integration deals with using the resource information to enable the relevant operations for online part manufacturing activities. It depends on enabling
● customers relationship management,
● machining queue scheduling and management,
● manufacturing capability evaluation,
● manufacturing tasks bidding,

- manufacturing process visualization including the output of on-site machining video stream and the report of machining status, and
- synchronously collaborative work.

In the client-side, customers, suppliers, manufacturers, and administrators use the e-service platform only through browsers without the help of the special platform-related local client programs.
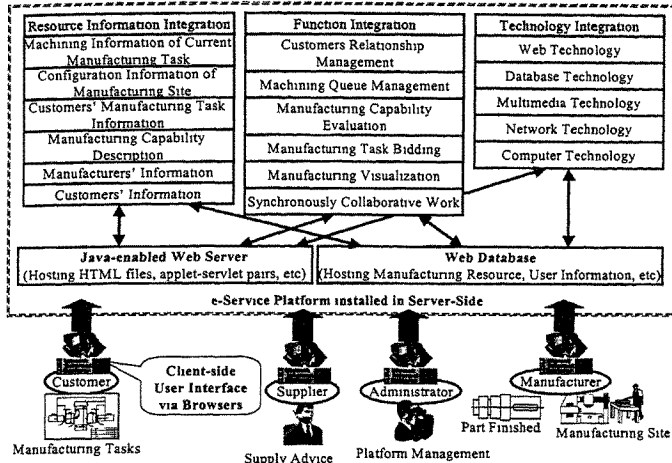

Figure 1 A System Architecture of e-Service Platform for online Part Manufacturing

It should be pointed out that the role of supplier in the part manufacturing is only to give some advices (like dimension demands to match his part/assembly to be sold, etc). Here, no further discussion on this issue is given.

2.3 Main Working Logic of the e-Service Platform

An online part manufacturing procedure using above e-service platform can be seen in Fig.2. Here, a customer firstly must register to the e-service platform when he/she has a part manufacturing task. As soon as the customer has a successful registration, the manufacturing task would be inputted to the web database. Now both customer and manufacturers who have registered to the e-service platform can query the manufacturing task. For the customer, he/she will query the potential manufacturing sites and their capability information, and then calculate the manufacturing cost. By means of using the evaluation criterion provided by the e-service platform, the customer can determine several potential manufacturers and put them into a bidding queue. At the same time, any manufacturer can declare a manufacturing price for the part to be fabricated and put himself into the bidding queue if he/she has an interesting to join the competition to obtain this manufacturing order. After waiting for a specified time, the customer and all manufacturers listed in the bidding queue will participate a bidding process in the mode of collaborative work and select a suitable manufacturer to accept the manufacturing order. After determining the manufacturer and the correspondent manufacturing site, the current manufacturing task will be appended into the manufacturing queue of facilities and wait for machining. When the customer's manufacturing task is being

fabricated, the video stream of on-site machining can be viewed by the customer. In addition, the collaborative work between the customer and the manufacturer can be started if necessary. In this way, parts can be fabricated in the form of online.
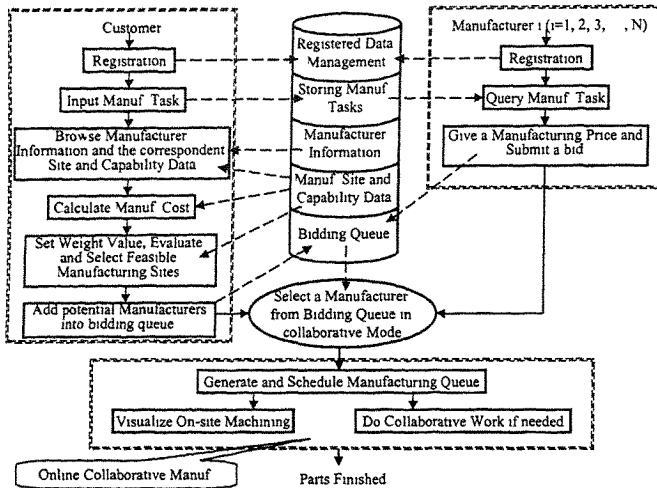
Figure 2 Working Flow of An Online Part Manufacturing via e-Service Platform

3 SOME KEY ENABLING TECHNIQUES

As mentioned above, making the e-service platform run well needs a lot of enabling techniques. Here, we only discuss some of them.

3.1 Selecting a Feasible Manufacturing Site

The key point to choose a feasible manufacturing site or manufacturer is to build a good evaluation method considering the factors like manufacturing price, quality, delivery time, and manufacturing capability, etc. As shown in Fig.3, an Analytical Hierarchical Planning (AHP) model is created. Based on the model, we use the weight-driven synthetic method to calculate the reference value that presents the performance of the manufacturing site evaluated.

Referring to the model in Fig.3, the correspondent steps for getting the evaluation value of one manufacturing site are described as follows.

● Set the weight coefficients for leaf nodes "Machining precision of facility" and "Product Eligibility" and calculate the evaluation value of the node "Manufacturing Precision" by using the following formula:

$$P_i = \sum_{j=1}^{n} P_{ij} \times W_{ij} \quad (P_{ij} \leqslant 1; \sum_{j=1}^{n} W_{ij} = 1; n=2)$$

● Calculate the value P_i of other leaf nodes in the criterion level according to the resource information of manufacturing site to be evaluated and make all the values be equal to or less than 1.

- Referring to the evaluation node "techniques" in the goal level B, set the weight coefficients respectively for correspondent nodes, which are connected with it, in the criterion level, and calculate the evaluation value of the node "techniques" by using the following formula:

$$P_{tech} = \sum_{j=1}^{n} P_{tech\,j} \times W_{tech\,j} \quad (P_{tech\,j} \leqslant 1; \sum_{j=1}^{n} W_{tech\,j} = 1; n=3)$$

- Calculate the evaluation values P_{prod}, P_{serv}, P_{cost}, and P_{other} of the other nodes in the goal level B according to the same method like calculating P_{tech}.
- Calculate the final evaluation value P_{site} by using the following formula:

$$P_{site} = P_{tech} \times W_{tech} + P_{prod} \times W_{prod} + P_{serv} \times W_{serv} + P_{other} \times W_{other}$$
$$(W_{tech} + W_{prod} + W_{serv} + W_{other} =1)$$

We repeat above steps till obtaining the evaluation values of all the manufacturing sites and arrange such values in order. It must be declared here that all weight coefficients can be chosen freely by the customer. If not, default weight coefficients will be used.
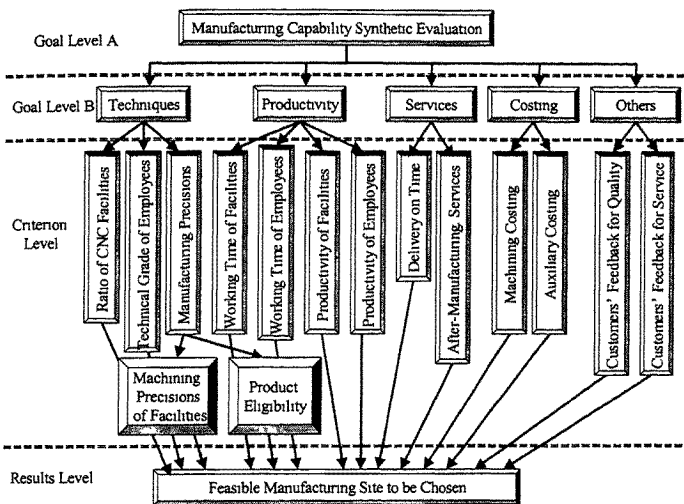
Figure 3 Analytical Hierarchical Planning Model for Synthetic Manufacturing Capability Evaluation

In this way, the customer can put some of potential manufacturing sites into the bidding queue. Furthermore, a feasible manufacturing site can be selected through the collaborative work for the assessment of all the bids submitted in the bidding queue.

3.2 Queuing a Manufacturing Task

Queuing a manufacturing task will refer to following three cases.
- *The first coming and then the first manufacturing,*

For a manufacturing queue related to a machining facility, it may be expressed as $S_1\{T_1, T_2, \cdots, T_i, \cdots, T_n\}$.

When the new manufacturing task T_{n+1} comes to a manufacturing site, the first thing we should do is to check the machining status of all the facilities according to the delivery time and the machining time of this task. When a facility leaves unused or has an ability to accept new tasks, this task is appended to the end of queue, that is, $S_i\{ T_1, T_2, \cdots, T_i, \cdots, T_n, T_{n+1}\}$.

- *Using the delivery time as the judging criterion of queuing,*

When the new manufacturing task T_{n+1} comes to a manufacturing site, it is inserted to a suitable position in the manufacturing queue, depending on the delivery time of this task. Under this case, the manufacturing queue is described as $S_i\{ T_1, T_2, \cdots, T_i, T_{n+1}, T_{i+1}, \cdots, T_n\}$.

It should be mentioned that inserting this task into the manufacturing queue is not able to influence the delivery time of tasks $\{ T_{i+1}, \cdots, T_n\}$. This is also the judging criterion to find the inserting position of the new manufacturing task.

- *Using the priority as the judging criterion of queuing,*

Sometimes the new manufacturing task T_{n+1} must be finished by the specified time. It hints the customer would give the manufacturer much more benefits through setting up the priority to this task. In general, the priority deals with three modes, that is, "urgent mode", "express mode", and "immediate mode". Here, "immediate mode" means the manufacturing task T_{n+1} is done immediately when it comes. It must be pointed out that this queuing case is just similar to the queuing case based on measuring the delivery time mentioned above.

3.3 Visualizing the Manufacturing Process

As soon as customer's manufacturing task is being done in the manufacturing site, he/she can join the fabricating process through the remote visualization and synchronously collaborative work environment. Here, the last issue will be discussed in the next section.

In fact, visualizing the manufacturing process can be considered from both the customer's view and the manufacturer's view. Here, the visualization from the manufacturer's view is mainly used for internal manufacturing scheduling, monitoring, and controlling. We will not discuss it in this paper. As regards the visualization from the customer's view, it plays an important role in making the manufacturing process become "transparency" and "online". This kind of visualization issue will be extended in this paper.

Generally, visualizing the manufacturing process from the point of view of the customer deals with two aspects. The first is to take the on-site video and audio streams and to present them in the front of the customer. The second is to broadcast the manufacturing status data, for example, executable NC code status, executable process plan status, a variety of statistical tables and charts based on SPC (Statistic Process Control), QC (Quality Control), etc, to the customer. For the first aspect, shown in Fig.4, web camera is used to take the on-site video and audio streams. It is configured with an independent IP address through the embedded hardware like AXIS products or via a host personal computer with which the camera is connected by using the standard USB interface. Java programs are included as important elements to take the video and audio streams. The customer can "see" the manufacturing site only by means of clicking a URL link to an embedded applet. Here, the second implementation is a kind of "soft" implementation. We have used Java Media Framework API to develop the correspondent programs.
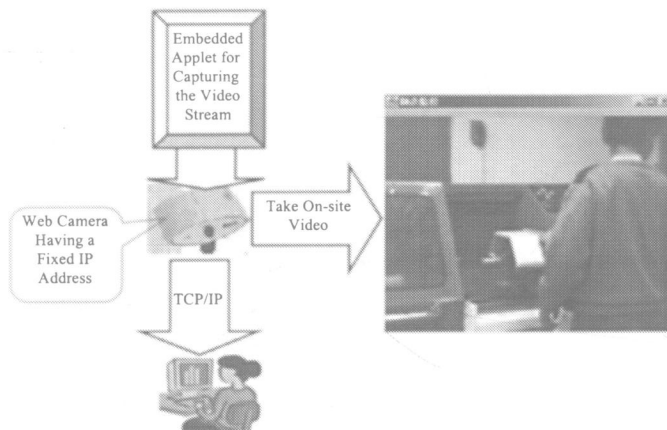
Figure 4 Viewing the Remote On-site Manufacturing Process with Web Camera

As for the manufacturing status data, they are collected with the interactive method through an applet-servlet pair. The sampled data are saved to the server and then broadcast to the correspondent persons.

3.4 Realizing the Synchronously Collaborative Manufacturing Work

As mentioned above, collaborative work plays a very important role in enabling the online part manufacturing. At least, it not only can support the interaction among customers, manufacturers, and suppliers, but can also help the customers finish the bidding activity during determining a feasible manufacturing site. At present, a project-based synchronously collaborative tool has been developed in our research group[1]. Fig.5 just shows a screen snapshot of using this tool during the online part manufacturing.
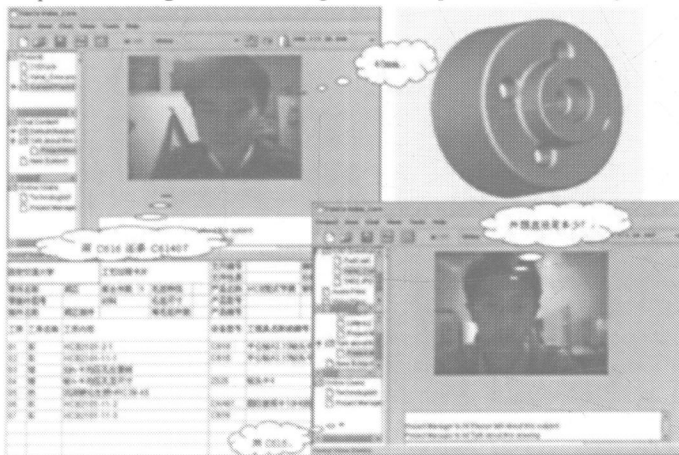
Figure 5 A Screen Snapshot of Synchronously Collaborative Manufacturing Applet-Servlet Pair

4 DISCUSSIONS

One of the motivations to do this research is to let the customer and the supplier join the manufacturing process in the mode of online. The contents described in this paper only focus on the part level in which the problem solving becomes simple. Under this case, the role of suppliers is little bit small. If we extend the framework into the product level, the suppliers will be involved in depth.

Another motivation to do this research is to hope to look for a kind of mode so as to integrate the distributive manufacturing resources for rapid, agile, and customer-driven manufacturing. So this approach can actually be concerned with a kind of initial implementation of e-manufacturing. However, there are still a lot of researches that need to do, like online manufacturing process control, digital manufacturing facilities, etc.

5 CONCLUSIONS

To sum up, we put forward an e-service platform in order to implement the web-based online part manufacturing in the form of synchronously collaborative work and give the correspondent working logic. One of the contributions in this paper is to clarify some of the key issues, including selecting a feasible manufacturer, queuing a manufacturing task, visualizing the manufacturing process, and using the synchronously collaborative work mechanism. In addition, making customers and suppliers involve in the manufacturing process in the mode of online service is another key point of our research. It must be mentioned that this research still needs to improve in many aspects.

ACKNOWLEDGEMENTS

REFERENCES

1. **Sun, Huibin and Jiang, Pingyu** (2002) A Project-based Synchronously Collaborative Tool for Networked Manufacturing. Journal of Xi'an Jiaotong University. Vol.36, No.7, pp726- 730. (in Chinese)
2. **Tso, S.K., et al.** (1999) A Framework for developing an agent-based collaborative service-support system in a manufacturing information network. Engineering Application of Artificial Intelligence. (12): pp43-57.
3. **Tso, S.K., et al.** (2000) Coordinating and Monitoring in an intelligent global manufacturing service system. Computers in Industry. 43(2000): 83-95.
4. **Slansky, Dick** (2001) Java Technology Powers e-Manufacturing. White Paper from: http://java.sun.com, Dec.2001, 15pp.
5. **Zhang, Yingfeng** (2002) Research on e-Service Platform for Web-based Part Manufacturing. Master Thesis. Mar. 2002, 81pp. (in Chinese)

Collaborative integrated planning for managing product rollovers in Internet-enabled supply chains

R GAONKAR and N VISWANADHAM
The Logistics Institute – Asia Pacific, National University of Singapore, Singapore

ABSTRACT

The proper management of multiple product generations overlapping each other is of critical importance in the high-tech industry. In this paper, we develop a mixed integer-programming model for integrated planning across the supply chain and show how such a model may be used for making decisions related to introduction and rollovers of products across multiple generations. We assume that all stakeholders in the supply chain share information on their capacities, schedules and cost structures. Based on this information the model addresses the issue of partner selection and planning for optimal profit.

1. INTRODUCTION

The dynamic nature of business today coupled with increasing customer expectations has increased manifold the competition in almost all industries. Increased competition has seen companies attempting to differentiate themselves by constantly innovating and introducing new products that cater to the ever-changing needs of the customers. For example in the PC industry, new product models appear every 6 to 9 months. Given the short product lifecycles, the ability to develop, produce and sell new products faster than the rest, and hence fully exploit the market opportunity, is a key competitive asset. This creates challenges of dealing with mass customization, rapid inventory depreciation and handling complex multi-sourced supply chains, when planning for production of multiple generations of products.

To successfully manage these issues it is imperative for businesses to develop systematic tools and methodologies for planning, based on the market demand and the production and distribution capabilities of the supply chains. Furthermore, given the highly distributed nature of manufacturing today, collaboration is a very essential aspect of supply chain planning. Hence, integrated planning on global information and close collaboration between supply chain partners are the two key critical success factors for successfully introducing and managing multiple generations of products through their lifecycles. In this regard, the ubiquitous nature of the Internet provides the ideal tool to not only collect and share operational information across the supply chain but also to manage and coordinate the

activities of all the supply chain participants. In particular, Internet-based public exchanges and private marketplaces have played an important role in redefining the nature and scope of supply chain interactions and decision-making.

In this paper, we focus our efforts on addressing a new area in integrated production and logistics planning for new product introductions and product rollovers in supply chains managed through private marketplaces, especially within the discrete manufacturing industry. A private marketplace is usually driven by a channel master with enough bargaining power to force its supply chain partners to participate in the marketplace. A description of a collaborative private marketplace implemented by Hewlett Packard is provided by Hammer [1]. In a collaborative private marketplace all the supply chain partners openly share information on their capacities, inventories, schedules, costs and lead-times in the marketplace. Sharing of such detailed operational information requires a high-level of trust between the various supply chain participants and a sense of a collective common destiny in the pursuit of common market objectives. The mission of the channel master then is to optimally plan, in collaboration with its partners, new product introductions and rollovers for maximum profit, subject to the production and logistics capacity constraints of its supply chain partners.

The issue of new product development has been widely studied in the marketing, operations management and engineering design literature. Krishnan and Ulrich [1] present a comprehensive review of the literature in this field in their review paper. Under their classification our problem in general arises under the study of supply chain design and specifically under the topic of production ramp-up and launch. They present a review of number of papers in both areas. However, the literature in the arena of planning for new product introduction and launch is mostly in the area of marketing with very little quantitative analysis with respect to production planning. Terwiesch and Bohn [2] try to quantitatively model the process improvement and learning and the resultant gradual increase in production yield during the introduction of a new product. In this respect our paper is an initial attempt to explore the application of some supply chain planning models for simultaneous production of multiple generations of products, each of them in different phases of their lifecycles.

2. PROBLEM FORMULATION

We consider the problem of manufacturing and logistics planning for managing product introductions and rollovers across multiple generations of products in a web-based collaborative environment. We assume that there are a number of component suppliers, sub-assembly manufacturers, contract manufacturers and logistics service providers in the supply chain owned by a powerful channel master. These supply chain participants may be geographically distributed in different parts of the globe. Each of them shares information on their production schedules, capacity, cost, quality, etc through the private marketplace. The logistics providers also share information on their costs and capacities for transporting various goods between the supply chain participants. We also assume that the demands over the entire life cycles of the various products in the various geographical market areas are known, through some forecasting model. These demands can be fulfilled by different sets of manufacturers and suppliers at different costs and in different lead times with the support of the logistics service providers. With access to complete visibility into its supply chain,

afforded by its private marketplace, the channel master needs to plan how best to manage rollovers between products and introduce a new generation of a product into the market, using a team of suppliers, contract manufacturers and logistics service providers to meet the market demand and maximize profit over the entire product lifecycle. Hence, a collaborative approach in product development and supply chain management is required.

To ease the understanding of our problem, we present an illustrative scenario for our model. Consider the case of a large PC manufacturer such as Dell. It introduces a new model of desktop PCs and laptops every 5-6 months. Each model of desktop PCs and laptops goes through the product lifecycle. When it is introduced into the market the model contains the latest features and commands a premium amongst the few pioneering buyers. With time the demand for the model goes up and it enters the mass market. Subsequently, other models of desktop PCs and laptops, with newer features are introduced into the market and the demand for the older model drops until it is taken out of the market. The drop in demand for the older models coincides with the increasing demand for the newer models and hence there is a rollover from one product generation to another. However, it very often happens that a significant part of desktop PCs and laptops, from both the newer and older generations, are made of the same components. For example a newer PC model might be running on a 1 GHz processor as compared to an older PC model running on a 667 MHz processor, but the hard disks, disk drives, monitors and other components in both models might in fact be exactly the same. Additionally, the duration and the apex of the product lifecycles for various product generations may vary across various geographical and customer market segments, requiring more detailed planning with consideration for issues in product rollovers across both market segments and product generations.

2.1 Notation

Identifiers

$r \in R$: Component type identifier.

$v \in V$: Component supplier identifier.

$i \in I$: Sub-assembly type identifier.

$j \in J$: Sub-assembly supplier identifier.

$t \in T$: Time Period identifier.

$k \in K$: Contract Manufacturer Identifier.

$m \in M$: Market Area Identifier.

$l \in L$: Brand Identifier.

$d \in D$: Transportation Mode (Sea/Air) Identifier

Parameters

PCap : Maximum production capacity.

PC : Unit cost price of production.

PFC : Production fixed cost.

TCap: Maximum transportation capacity.

TC : Unit transportation cost.

TFC : Fixed cost for procuring capacity.

R_{ab} : Units of component a required for the production of 1 unit of sub-assembly b.

M_{ab} : Units of sub-assembly a required for the production of 1 unit of model b.

WC : Unit inventory cost incurred.

TL : Transportation lead-time.

BD : Quantity required by the market.

BSL : Service level.

LSC : Cost of lost sales.

Variables

Q : Quantity produced.

S : Quantity shipped.

S' : Quantity received.

I : Amount of Inventory held.

F : Fixed cost applies $\{\in 0,1\}$.

BS: Quantity sold.

2.1 Mixed Integer Programming (MIP) Model

We now develop a MIP model for a dynamic manufacturing network for product rollovers. The model attempts to maximize the profit of the network subject to various operational constraints.

<u>Objective function</u>

$$
\begin{aligned}
MaxPROFIT \quad = \quad & \sum_{l=1}^{L}\sum_{m=1}^{M}\sum_{t=1}^{T} P_{lm}\, BS_{lmt} - \left[\sum_{l=1}^{L}\sum_{m=1}^{M}\sum_{t=1}^{T} \left(BD_{lmt} - BS_{lmt} \right) LSC_{lmt} \right] \\
& - \left[\begin{aligned} & \sum_{r=1}^{R}\sum_{v=1}^{V}\sum_{t=1}^{T}\left(PFC_{rv}\, F_{rvt} + PC_{rv}\, Q_{rvt} \right) \\ & + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\left(PFC_{ij}\, F_{ijt} + PC_{ij}\, Q_{ijt} \right) \\ & + \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{t=1}^{T}\left(PFC_{lk}\, F_{lkt} + PC_{lk}\, Q_{lkt} \right) \end{aligned} \right] \\
& - \left[\begin{aligned} & \sum_{r=1}^{R}\sum_{v=1}^{V}\sum_{j=1}^{J}\sum_{d=1}^{D}\sum_{t=1}^{T}\left(TFC_{rvjd}\, F_{rvjdt} + TC_{rvjd}\, S_{rvjdt} \right) \\ & + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{d=1}^{D}\sum_{t=1}^{T}\left(TFC_{ijkd}\, F_{ijkdt} + TC_{ijkd}\, S_{ijkdt} \right) \\ & + \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{d=1}^{D}\sum_{t=1}^{T}\left(TFC_{lkmd}\, F_{lkmdt} + TC_{lkmd}\, S_{lkmdt} \right) \end{aligned} \right] \\
& - \left[\sum_{i=1}^{T} \left(\begin{aligned} & \sum_{r=1}^{R}\sum_{v=1}^{V} WC_{rv}\, I_{rvt} + \sum_{r=1}^{R}\sum_{j=1}^{J} WC_{rj}\, I_{rjt} \\ & + \sum_{i=1}^{I}\sum_{j=1}^{J} WC_{ij}\, I_{ijt} + \sum_{i=1}^{I}\sum_{k=1}^{K} WC_{ik}\, I_{ikt} \\ & + \sum_{l=1}^{L}\sum_{k=1}^{K} WC_{lk}\, I_{lkt} + \sum_{l=1}^{L}\sum_{m=1}^{M} WC_{lm}\, I_{lmt} \end{aligned} \right) \right]
\end{aligned}
$$

$$\text{... (1)}$$

<u>Component supplier constraints</u>

$$Q_{rvt} \le PCap_{rvt} F_{rvt} \qquad forall \qquad r \in R, v \in V \;\&\; t \in T \qquad \text{... (2)}$$

$$I_{rv(t-1)} + Q_{rvt} = \sum_{j=1}^{J}\sum_{d=1}^{D} S_{rvjdt} + I_{rvt} \qquad forall \qquad r \in R, v \in V \;\&\; t \in T \qquad \text{... (3)}$$

$$S_{rvjdt} \le TCap_{rvjdt} F_{rvjdt} \qquad forall \qquad r \in R, v \in V, j \in J, d \in D \;\&\; t \in T \qquad \text{... (4)}$$

<u>Sub-assembly supplier constraints</u>

$$S'_{rvjdt} = S_{rvjd(t-TL_{vjd})} \qquad forall \qquad r \in R, v \in V, j \in J, d \in D \ \& \ t \in T \qquad \dots (5)$$

$$I_{rj(t-1)} \geq \sum_{i=1}^{I} R_{ir} Q_{ijt} \qquad forall \qquad r \in R, j \in J, t \in T \qquad \dots (6)$$

$$I_{rj(t-1)} + \sum_{v=1}^{V} \sum_{d=1}^{D} S'_{rvjdt} = \sum_{i=1}^{I} R_{ir} Q_{ijt} + I_{rjt} \qquad forall \qquad r \in R, j \in J \ \& \ t \in T \qquad \dots (7)$$

$$Q_{ijt} \leq PCap_{ijt} F_{ijt} \qquad forall \qquad i \in I, j \in J \ \& \ t \in T \qquad \dots (8)$$

$$I_{ij(t-1)} + Q_{ijt} = \sum_{k=1}^{K} \sum_{d=1}^{D} S_{ijkdt} + I_{ijt} \qquad forall \qquad i \in I, j \in J \ \& \ t \in T \qquad \dots (9)$$

$$S_{ijkdt} \leq TCap_{ijkdt} F_{ijkdt} \qquad forall \qquad i \in I, j \in J, k \in K, d \in D \ \& \ t \in T \qquad \dots (10)$$

<u>Contract Manufacturer Constraints</u>

$$S'_{ijkdt} = S_{ijkd(t-TL_{jkd})} \qquad forall \qquad i \in I, j \in J, k \in K, d \in D \ \& \ t \in T \qquad \dots (11)$$

$$I_{ik(t-1)} \geq \sum_{l=1}^{L} M_{li} Q_{lkt} \qquad forall \qquad i \in I, k \in K \ \& \ t \in T \qquad \dots (12)$$

$$I_{ik(t-1)} + \sum_{j=1}^{J} \sum_{d=1}^{D} S'_{ijkdt} = \sum_{l=1}^{L} M_{li} Q_{lkt} + I_{lkt} \qquad forall \quad i \in I, k \in K \ \& \ t \in T \qquad \dots (13)$$

$$Q_{lkt} \leq PCap_{lkt} F_{lkt} \qquad forall \qquad l \in L, k \in K \ \& \ t \in T \qquad \dots (14)$$

$$I_{lk(t-1)} + Q_{lkt} = \sum_{m=1}^{M} \sum_{d=1}^{D} S_{lkmdt} + I_{lkt} \qquad forall \qquad l \in L, k \in K \ \& \ t \in T \qquad \dots (15)$$

$$S_{lkmdt} \leq TCap_{lkmdt} F_{lkmdt} \qquad forall \qquad l \in L, k \in K, m \in M, d \in D \ \& \ t \in T \qquad \dots (16)$$

<u>Buyer Constraints</u>

$$S'_{lkmdt} = S_{lkmd(t-TL_{kmd})} \qquad forall \qquad l \in L, k \in K, m \in M, d \in D \ \& \ t \in T \qquad \dots (17)$$

$$I_{lm(t-1)} + \sum_{k=1}^{K} \sum_{d=1}^{D} S'_{lkmdt} = I_{lmt} + BS_{lmt} \qquad forall \qquad l \in L, m \in M \ \& \ t \in T \qquad \ldots (18)$$

$$BSL_{lm} BD_{lmt} \le BS_{lmt} \le BD_{lmt} \qquad forall \qquad l \in L, m \in M \ \& \ t \in T \qquad \ldots (19)$$

Constraints 2,4,8,10,14 and 16 state production and transportation capacity limitations. Constraints 3,7,9,13,15 and 18 model the flow balancing constraints for the various inventories in the supply chain. Constraints 5,11 and 17 model the deterministic transshipment lead-time between various locations through different transportation modes. Constraints 6 and 12 check for availability of all required parts before production begins. Constraint 19 is the demand-pull on the supply chain. The solution of this model determines the selection of suitable partners who can help the channel master best meet the market opportunity in a cost effective manner, and also provides a schedule for production and assembly activities within the supply chain. Any of the available optimization toolkits might be used to solve the above mathematical model.

3. COMPUTATIONAL RESULTS

In order to verify the optimised nature of the model that was developed in earlier sections, the model was solved for a known product lifecycle demand curve and a given supply chain network environment. The choice of supply chain partners and the scheduling of activities in the supply chain were observed. The demand curves as given in Fig. 1 for the two products in two market areas were assumed. The product lifecycle durations and the uptake in the two market areas are also different. Hence, as may be noticed there is a rollover period in between when both products are being sold in the market. Also, the products are assumed to share certain components, and procurement of components may be done keeping in mind the demand for both the models.
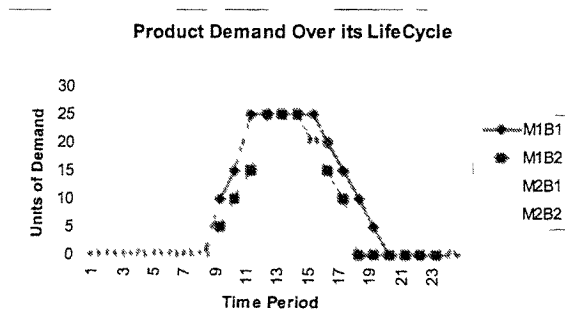
Figure 1: Product Demands over their lifecycles.
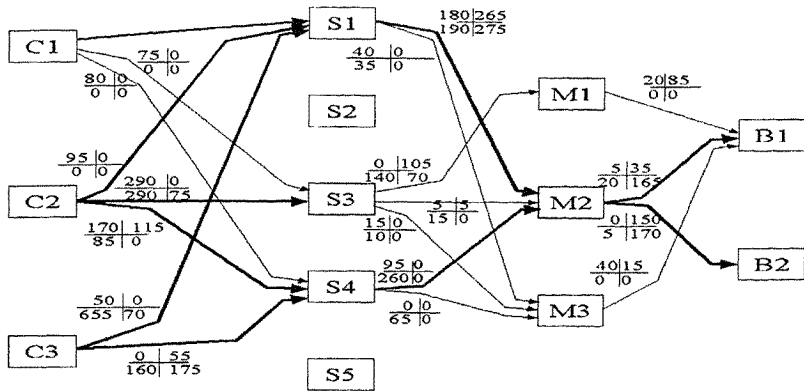
Figure 2: Supply Chain Configuration with Integrated Planning.
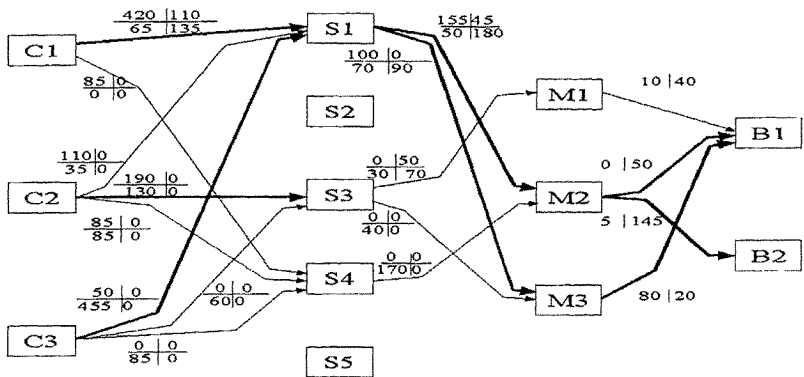
Figure 3: Supply Chain Configuration for Product Introduction of Brand M1.
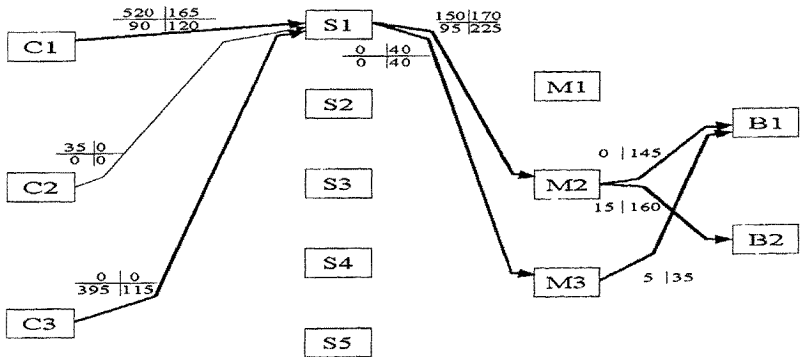
Figure 4: Supply Chain Configuration for Product Introduction of Brand M2.

For a given supply chain network, the following supply chain configuration with integrated planning for two new product introductions was obtained as given in Fig 2. The solution with profit of $6,875,650 was obtained in 27 hrs 23 minutes within 3.91% of optimality. For the same supply chain the optimal configuration when planning independently for individual product introductions was obtained as given in Fig 3 & 4. The profit for M1 introduction was optimally obtained as $ 3,973,350 and for M2 introduction as $ 3,348,205. The profit expected from individual planning for M1 and M2 is greater by around 6.5% than the profit from integrated planning for both M1 and M2 together.

Planning for individual product introductions can over-estimate or under-estimate the profit expected. When planning for two product introductions simultaneously, there can be a significant benefit in terms of securing lower costs for components and transportation costs, by leveraging upon greater volumes over both products. This is especially true for components that are common to both brands. In terms of procurement the costs may be very low, however the lowest cost supplier and transportation provider might not have adequate capacity to meet the needs of both the product introductions together. This will necessitate a need to deal with more expensive suppliers and transportation providers leading to higher costs and lower profits. Therefore, in integrated planning for new product development the trade-off between the cost efficiencies from joint procurement and the cost of dealing with more expensive suppliers needs to be well managed. In industries where there is excess capacity to be able to meet the needs of multiple product introductions significant savings can be expected from joint planning and procurement.

4. CONCLUSION

In this paper we have formulated and solved a integrated model for new product introductions in a web-based collaborative environment. Our formulation here, which is a mixed integer linear programming model, provides a good planning tool to schedule production and shipment activities down the supply chain in line with the demands over the products life cycle. We have assumed the availability of operational information in each stage of the supply chain to all the supply chain partners, which might not be the case in the real world.

5. REFERENCES

[1] **Hammer, Michael,** "The Superefficient Company", Harvard Business Review.
[2] **Krishnan, V.** and Karl T. Ulrich, "Product Development Decisions: A Review of the Literature", *Management Science*, Vol 47. No.1, January 2001, 1-21.
[3] **Terwiesch, Christian** and Roger E. Bohn, "Learning and Process Improvement during Production Ramp-up", *Intl. Journal of Prod. Economics*, Vol. 70 No. 1, March 2001, 1-19.
[4] **Alain, Guinet,** "A Primal-Dual Approach for Capacity-constrained Production Planning with Variable and Fixed Costs", *Computers & Industrial Engineering*, Vol 37, 1999, 93-96.

Development of a dynamic web-based graphing tool

P LIN and **R EAPPEN**
Mechanical Engineering Department, Cleveland State University, USA

SYNOPSIS

Although there are many web-based graphing tools available, they are mostly static in nature. Even the ones that are dynamic usually update and re-graph the data with a preset time increment. It is desired that web-based graphs can be automatically updated and displayed as soon as the data has been modified. This paper describes how Java Applet, Java Servlets and their communication in conjunction with a data transfer program can be used for dynamic data graphing in the web. To make the graphing tool user friendly and interactive, a user interface was designed and built into the graphing tool.

1 INTRODUCTION

In terms of dynamic web content updating, very few literature are available. Most of them are commercially available products. One that is worth of noting is the work done by Kumar (1). He used an instance of a class to accept socket connections and to spin off a new thread for each new connection request. To keep the computer code simple, each thread checks whether the data file has been modified. In this paper, we present the techniques for dynamic web-based graphing using Java Applet, Java Servlets and their communication. The Java Applet is considered the best choice for drawing and displaying graphs on a web browser. The Java Servlets, running in the Server, retrieve the data and check if the data have been modified. The Servlets also act as a means of communication between the Server and the Applet. The detailed descriptions of Applet and Servlets are described below:

1.1 Java Applet

A Java Applet is a small program written in Java that is embedded in an application (2). The Java Applet class is a user interface component. When an Applet instance is first instantiated,

Java invokes the Applet's initialization method. When the Applet containing web page is about to appear, Java invokes the Applet's start method. When the Applet containing a web page is about to be updated to by a new one, Java invokes the stop method. When the web page is removed from the browser's cache and the Applet is about to go away, Java invokes the Applet's destroy method. After a Java Applet is embedded in a web page, a browser can download it and execute it automatically. However, this makes the Internet browser insecure. Therefore, jdk1.0 assumes that all Java Applet is not to be trusted and should be under the watch of the security manager. By dong so, it severely limits what the Applet could do. For example, the security manager makes sure that the Applet could not write to the user file system, read certain system properties, accept incoming socket connections, or establish outgoing socket connections to any host but the origin Server (3). Fig.1 shows the relationship between Applet and web page.

The user's computer requests a web page.

Responds

Web Server

The user's computer displays the graph in the browser.

The Applet used for graph display is embedded in the web page.
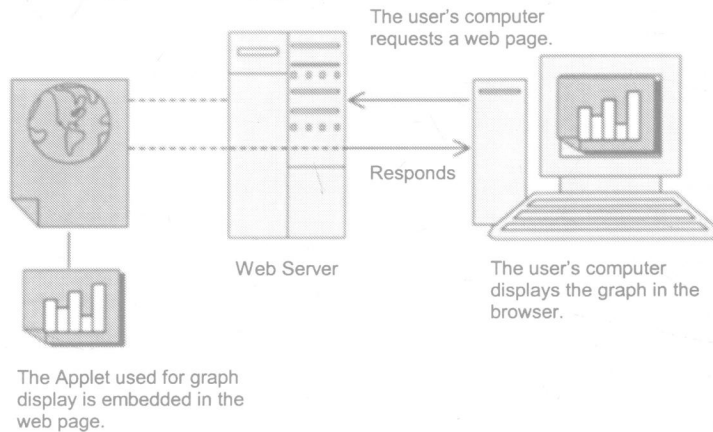
Fig. 1 Java Applet and Web Page

1.2 Java Servlets

Java Servlets is a generic server extension - a class that can be loaded dynamically to expand the functionality of the Server (4). The Servlets are commonly used with a web Server. The Servlets run inside the Java virtual machine on the Server, and operate solely in the domain of the Server. Unlike the Applet, they do not require support for Java in the browser. Another advantage of Servlets is that they are independent of the operating system and the web server. The Servlets invocation is highly efficient. Once the Servlets are loaded, they generally remain in the Server's memory as a single instance. Thereafter the Server invokes the Servlets to handle a request using a simple, lightweight method invocation. The Servlets can begin handling the request almost immediately. Multiple or concurrent requests are handled by separate threads.

In addition to using the Java Applet and Servlets, the communication between them has to be established, and data has to be transferred using FTP commands.

1.3 File Transfer Protocol (FTP)

FTP is an acronym for File Transfer Protocol. FTP is a client/Server application that allows file transfer between computers. The transfer can take place between a mainframe and a local terminal or over the Internet between a host computer and a distant server. FTP is a powerful tool, which allows users to access archives that are available on a large number of computer hosts. The key elements of FTP are:

a) Finding FTP sites from a client based system
b) Establishing a connection with the Server
c) Developing an ability to search through "archives" to retrieve information
d) Using FTP commands to facilitate the transfer of information
e) Allowing for the differences in file types and compression's techniques.

The concept of client/server is important - that is, the "local" client is initiating a communication pathway with a remote server that may contain public information of interest to the client.

FTP can be invoked by a command line, such as "ftp.mcgiserver.com". When a connection is established, one will be asked for a log in name and password. Many systems allow anonymous FTP, in which case one should use the log name `anonymous' and email address as password. If no host is specified or the connection is unsuccessful, FTP enters the command interpreter and awaits instructions.

2. THE METHODOLOGY

Generally speaking, the entire process of dynamically displaying and graphing on the web can be divided into the following three steps.

Step 1: Data Checking and Requesting

A host computer generates a set of data and saves it in a data file. Then, a data transfer program (using FTP commands) always running in the background checks if the data file residing in a server has been modified. As soon as the file has been modified, the data transfer program takes over completely and establishes a connection with the FTP Server, and transfer the data file to server. Meanwhile, the Java Applet sends a request for retrieving the updated data. This step is illustrated in Fig.2 below:
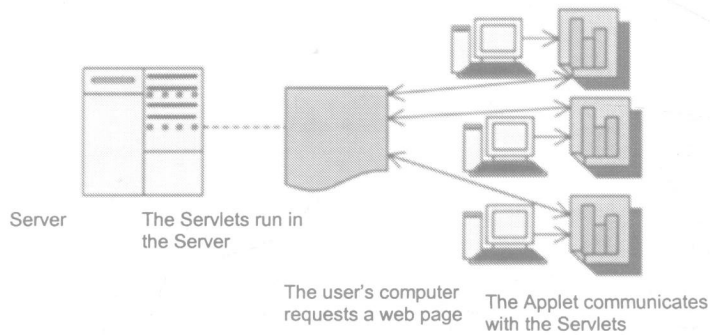
Fig. 2 Java Applet Requesting Data from Servlets

Step 2: Data Updating and Graph Repainting

Upon a request from the Applet, the Servlets receive the updated data from the server and send to the Applet. Afterward, the Applet executes on the user browser and repaints the graph.

Step 3: Applet-Servlets Communication

For the graph to represent the most current data, the Applet has to constantly request for the updated data from the Server. For this purpose, the Servlets are used in conjunction with the Applet to establish the "Applet-Servlets Communication".

The Applet-Servlets communication scheme is shown in Fig. 3, in which the Applet on the user's browser connects to the Servlets in the Server using an URL connection which points to the location of the Servlets. Then, the Applet sets up a socket connection to communicate with the Servlets. In return, the Servlets use the same or different socket connection to communicate with the Applet. The Applet reads data from the Servlets using an Input Stream and repaints the graph, unless the Applet is put into a sleep mode.

The user's computer requests the
web page that contains a grap

Requests

Respond

Transfers

Servlets

Serv

Appl

The Applet requests data from

The web page located in
the Server contains the
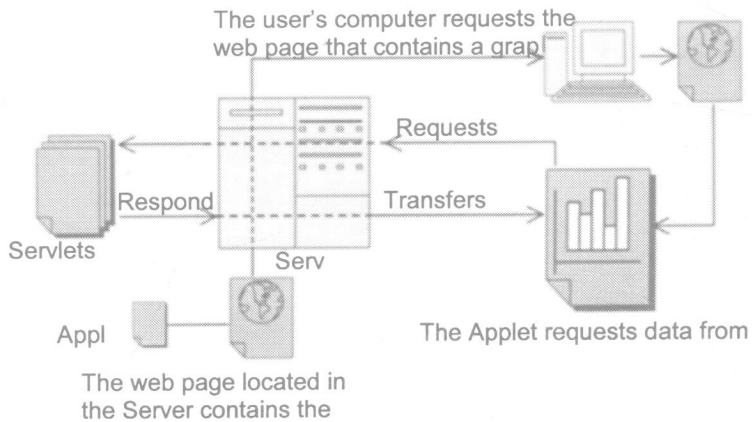
Fig. 3 The Applet-Servlets Communication

3. THE USER INTERFACE

The presented dynamic web-graphing tool was designed to be user friendly and interactive by means of user interface. The interface consists of three modules: Configure Graph, Form HTML File and Graphing Data.

The first screen shot shows the Configuration Panel (Fig. 4). All the information necessary for displaying the graph is entered here. Clicking the "config" button will generate the configuration data, which can be copied and used for graphing in the web.

Fig. 4 The Configuration Panel

The second screen shot shows the "Form HTML File" Panel (Fig. 5). This panel is used to embed the Applet into a web page. This panel allows the user to open any HTML file on his or her system and embed the Applet into it. This panel also allows the user to transfer the new HTML web page to the web Server.
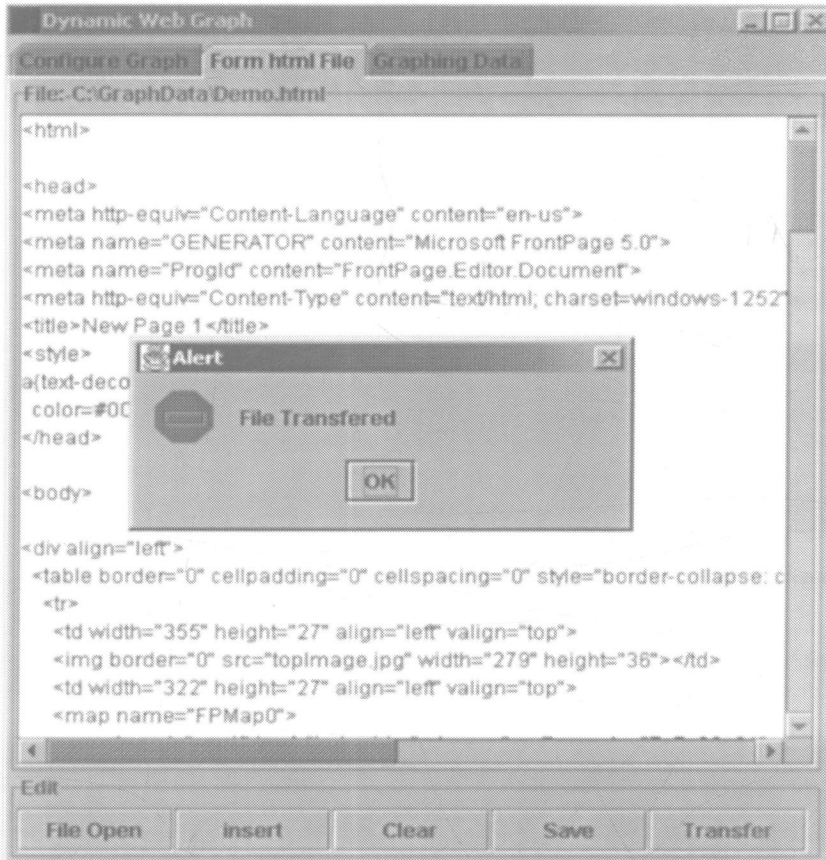
Fig. 5 The Web File Editing and Transferring Panel

The "Graphing Data" Panel as shown in Fig. 6 gives the user an option to manually enter new input values. Note that in addition to this manual mode, the input data saved in a file called "data.txt" can be automatically retrieved without human intervention. Once the new data has been generated, clicking the "Send" button will transfer the data as a data file to the remote web Server and also generate a log file, which can be viewed to check the data transfer process. As a web page is loaded to the user's browser, the Applet (embedded in the web page) executes and paints the web graph. Afterward, the Applet continues to wait for the Servlets to send the new data.
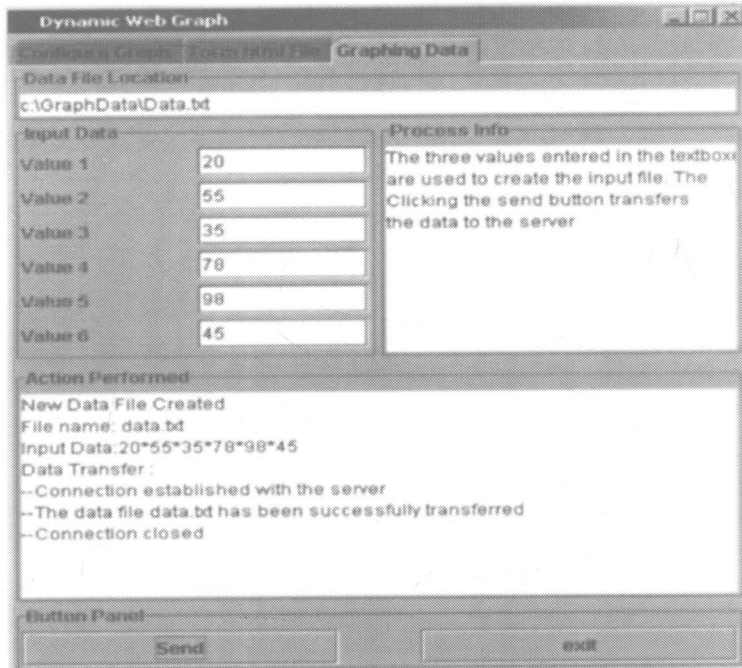
Fig. 6 The Data Creation and Transfer Panel

4. CONCLUSION

The developed tool is written in Java, a platform-independent programming language, which can run on any operating system and is compatible with all the major browsers. Furthermore, it does not require any maintenance once the tool has been installed. This paper showed how Java Applet and Servlets could be used for web-based graphing. The dynamic capability of the web-based graphing is accomplished by constantly checking if data has been modified, and communicating between the Applet and the Servlets. Unlike most commercially available web updating tools, the presented tool does not perform re-graphing if the data remain unchanged.

REFERENCES

1 **Kumar, D. S. R**. (2001) Updating Web Content dynamically with Java, http://devx.com.
2 **Jaworski, J.** (1999), Java Unleashed, TATA McGraw Hill Inc., Second Edition.
3 **Naughton, P.** (1999), Complete Reference for Java 2, TATA McGraw Hill Inc., 3rd Edition.
4 **Hunter, J. and Crawford, W.** (1998) Java Servlet Programming, Orieilly Inc., 1st Edition.

Authors' Index

International Conference on Manufacturing Automation covers all aspects of manufacturing automation popular in research. CAD, CIM, rapid prototyping and tooling, industrial robotics, process planning, geometric modelling, and feature-based technology are all covered here. Other topics include product development, high-efficiency machining, and Internet-based design.

The broad theme of the contributions is rapid product development with the emphasis on software, modelling, and planning. A range of work from international experts is brought together offering insights and developments in the following fields of research:

• Design systems and methodologies

• Machining technology

• Intelligent systems

• Technology management

• Rapid product development

• Internet-based systems

Ideas, research, and technological developments in the areas of manufacturing automation are explored to inform those working in the field and to foster better links between industry and academia.

The editors, Professor S T Tan, Dr I Gibson, and Dr Y H Chen, are all members of the Mechanical Engineering Department of The University of Hong Kong. They are all considered experts in their respective fields and combine to form many years of experience in both the software and hardware aspects of automated manufacture.