Recent Advances in Integrated Design and Manufacturing in Mechanical Engineering

Edited by
Grigore Gogu
Daniel Coutellier
Patrick Chedmail
Pascal Ray

Springer-Science+Business Media, B.V.

Recent Advances in Integrated Design and Manufacturing in Mechanical Engineering

Recent Advances in Integrated Design and Manufacturing in Mechanical Engineering

Edited by

GRIGORE GOGU

Laboratoire de Recherches et Applications en Mécanique Avancé,
Institut Français de Mécanique Avancée et Université Blaise Pascal,
Clermont-Ferrand, France

DANIEL COUTELLIER

Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines,
Ecole Nationale Supérieure d'Ingénieurs en Informatique, Automatique, Mécanique
Énergétique et Électronique, Université de Valenciennes, France

PATRICK CHEDMAIL

Institut de Recherche en Communications et Cybernétique de Nantes,
Ecole Centrale de Nantes, France

and

PASCAL RAY

Laboratoire de Recherches et Applications en Mécanique Avancé,
Institut Français de Mécanique Avancée et Université Blaise Pascal,
Clermont-Ferrand, France

Springer-Science+Business Media, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-6236-9 ISBN 978-94-017-0161-7 (eBook)
DOI 10.1007/978-94-017-0161-7

Printed on acid-free paper

Originally published by Kluwer Academic Publishers in 2003

Softcover reprint of the hardcover 1st edition 2003

TABLE OF CONTENTS

Chapter 3
METHODOLOGICAL ASPECTS OF INTEGRATED DESIGN AND MANUFACTURING

PREFACE

This volume contains selected manuscripts of the papers presented at the 4th International Conference on **I**ntegrated **D**esign and **M**anufacturing in **M**echanical **E**ngineering – **IDMME** 2002 held in Clermont-Ferrand, France, at the Institut Français de Mécanique Avancée (French Institute for Advanced Mechanics), May 14-16, 2002.

Following the first three IDMME conferences (held in Nantes in 1966, in Compiègne in 1998 and in Montréal in 2000), the purpose of IDMME 2002 conference was to present recent developments in the integration and the optimization of product design and manufacturing systems.

The initiative and the organization of the conference exists thanks to the French national network **AIP-PRIMECA** (**AIP** - **A**teliers **I**nter-Etablissements de **P**roductique / Inter-Institute Productics Workshops and **PRIMECA** - **P**ôle de **R**essources **I**nformatiques pour la **MECA**nique / Computer Resource Pole for Mechanics). The conference benefited from the scientific sponsoring of **AFM** – **A**ssociation **F**rançaise de **M**écanique (French Association of Mechanical Engineering), **ASME** – **A**merican **S**ociety of **M**echanical **E**ngineering, **CIRP** – Collège International pour l'Etude Scientifique des Techniques de Production Mécanique (International Institution for Production Engineering Research), **CSMA** – **C**omputational **S**tructural **M**echanics **A**ssociation, **IFToMM** – **I**nternational **F**ederation for the **T**heory **o**f **M**achines and **M**echanisms, **SCGM/CSME** – **S**ociété **C**anadienne de **G**énie **M**écanique / **C**anadian **S**ociety of **M**echanical **E**ngineering.

The conference brought together 270 participants from 16 countries. A hundred and fifty five papers were presented in 47 sessions and included in the CD-proceedings given to the participants. Three invited lectures were presented :

1. The stakes of concurrent engineering in the tyre industry
   by Didier MIRATON, Director of the MICHELIN Technology Center, Clermont-Ferrand, France
2. The dual nature of technical artifacts
   by Louis BUCCIARELLI, Professor at MIT, USA
3. Different uses of polyhedron models in a context of digital mock-up applied to aeronautical products engineering
   by Xavier ROGOVITZ, Head of DMU Technical Data Manager A400M programme, EADS CASA, Espagne

This book contains the full versions of 50 papers selected by the International Scientific Committee from the 155 papers presented at the IDMME 2002 Conference:

Chair
GOGU, G. *France*
Co-Chairs
CHEDMAIL, P. *France*
COUTELLIER, D. *France*
DORNFELD, D. *USA*
FORTIN, C. *Canada*
Members
ANGELES, J. *Canada*
BATOZ, J.L. *France*
BILLOET, J.L. *France*
BOCQUET, J.C. *France*
BOURDET, P. *France*
BRISSAUD, D. *France*
CLEMENT, A. *France*
CLOZEL, P. *France*
COGNET, G. *France*
COFFIGNAL, G. *France*
DEGALLAIX, G. *France*
DE PENNINGTON, A. *U.K.*
DUFFY, A.H.B. *U.K.*
FAVREL, J. *France*
GUILLOT, J. *France*
ISPAS, C. *Romania*
JEDRZEJEWSKI, J. *Poland*
JOVANE, F. *Italy*
KIMURA, F. *Japan*
KJELLBERG, T. *Sweden*
KRAUSE, F.-L. *Germany*
KUSIAK, A. *USA*
LAPPALAINEN, K. *Finland*
LEMAIRE, M. *France*
LINDE, H. *Germany*
MARQUIS, D. *France*
MARTIN, P. *France*
MASCLE, C. *Canada*
McMAHON, C. *U.K.*
ORSERO, P. *France*
PEGNA, J. *Canada*
PELLE, J.-P. *France*
PESEUX, B. *France*
PLAY, D. *France*
PROVOST, G. *France*
RAY, P. *France*
RIGAL, J.-F. *France*
RIS, G. *France*
RITCHIE, J.M. *U.K.*
SOHLENIUS, G. *Sweden*
SUH, N.P. *USA*
TICHKIEWITCH, S. *France*
TOLLENAERE, M. *France*
TOMIYAMA, T. *Japan*
TÖNSHOFF, H. K. *Germany*
VAN BRUSSEL, H. *Belgium*
VAN HOUTEN, F.J.A.M. *Netherlands*
VERON, M. *France*
WALDRON, J. K. *USA*

The 5th IDMME Conference will take place in Bath, U.K., in 2004.

The editors
G. Gogu, D. Coutellier, P. Chedmail, P. Ray

ACKNOWLEDGEMENTS

The 4th IDMME Conference, held at the French Institute of Advanced Mechanics in Clermont-Ferrand in May 2002 received the scientific and the financial support of the French Ministry of Higher Education and Research.

AIP-PRIMECA, AFM, CSMA, the Foundation of the French Institute of Advanced Mechanics, Blaise Pascal University– Clermont-Ferrand, Conseil Général du Puy-de-Dôme, AUBERT&DUVAL Holding, EDS, HURON, MICHELIN, Volvic provide financial support for the Conference and for the edition of this book.

All the members of the organizing committee contributed to the success of the IDMME 2002 Conference. They belong to the Laboratory of Research and Applications in Advanced Mechanics (LaRAMA), the French Institute of Advanced Mechanics (IFMA) and Blaise Pascal University (UBP) – Clermont Ferrand. Pascal RAY (LaRAMA, IFMA, UBP) was the chair of the Organizing Committee and Bernard CHENEVEAUX (MICHELIN) acted as co-chair.

AIP-PRIMECA

The 4^{th} edition of the International Conference on Integrated Design and Manufacturuing in Mechanical Engineering (IDMME), held from 14th to 16th May 2002 at the Institut Français de Mécanique Avancée (IFMA) - French Institute of Advanced Mechanics located at Clermont-Ferrand (France), was organized under the scientific responsibility of the recently launched national network AIP-PRIMECA following the implementation of a new structure associating the activities of two previous networks, namely the AIP (Ateliers Inter-Etablissements de Productique/ Inter-Institute Productics Workshops) and PRIMECA (Pôle de Ressources Informatiques pour la Mécanique/Computer Resource Pole for Mechanics). This network is organized along the lines of a joint project : the evolution, in the field of training for Integrated Design in Mechanics and Productics, closely linked to the constantly changing industrial needs of these last 15 years. It has managed to bring together the different participants in the fields of Mechanical Engineering, Industrial Engineering and Productics. The national AIP-PRIMECA network is in charge of promoting both exchanges of experience and know-how capitalisation. It has a paramount mission to fulfil, be it in the field of initial and continuous education, technological transfer and knowledge dissemination through strong links with research labs. The network dynamics resulting from the merger of previous structures has made it possible to generate strong synergies within the regional poles/institutes thus facilitating the medium/long-term consolidation of a development policy for the resource centers many of which are increasingly turned into high-tech and high specificity competence centers. The amalgamation process has allowed the development of an open institute in the field of integrated design as applied to Mechanics and Productics. In addition, it facilitates the already existing programmes as regards pedagogical experiments, thus allowing the promotion of research programmes being carried out by the various network partners in their own research labs.

Among the programmes promoted, the network regularly organizes special events such as scientific day conferences, national colloquia, fall conferences and international congresses. IDMME, as part of this programme, has enabled an assessment of the state of the art as regards tool evolution, methodologies, organizations in the field of integrated design and manufacture. The industrial problematic is currently slanted towards the improvement of the design processes now perceived as an inverse approach with the different products to manufacture as the starting point. Needless to say, both methodological improvement, as a blanket term, and model optimization are still topical issues quite in line with the organization of an optimal design. This organization radical change on how to apprehend design is an on-going process, which must not neglect the improvement of already existing tools thus allowing a product validation in the final phase, aiming at being more performing.

A selection of the best papers from this conference has therefore been made which has led to the publication of this book whose ambition is to be considered as a work of reference both for researchers in this particular field and for teaching staff confronted with training methodology in integrated design and productics. It should allow assessment of the scope of development prospects in an extremely wide-ranging field.

Professor Daniel Coutellier
Director of AIP-PRIMECA network

INTRODUCTION

This book presents recent advances in the integration and the optimization of product design and manufacturing systems. It contains 50 selected papers from the 155 papers presented in the 4th International Conference on **I**ntegrated **D**esign and **M**anufacturing in **M**echanical **E**ngineering – **IDMME** 2002 held in Clermont-Ferrand, France, at Institut Français de Mécanique Avancée (IFMA) – French Institute of Advanced Mechanics in May 2002. The conference highlighted three main topics:

a) optimization of product design process (design for manufacturing and global optimization, product and process configuration management, representation and integration of design constraints including greening design, integration and coupling of models, integration if CAD and CAE models),

b) optimization of manufacturing systems (modeling of processes, rapid prototyping and tooling, modeling and design for manufacturing, modeling for control and inspection, process planning, CAM and off line programming, CAM, and optimal parameters in manufacturing processes),

c) methodological aspects of integrated design and manufacturing (computational geometry in design and manufacturing, concurrent engineering, integrated design and CAD/CAM systems, object modeling, feature-based modeling, design and communication, design and innovation, robust and reliable design, interfaces, technical data exchanges, distributed CAD, cooperative systems, intelligent agents).

The book is divided into 3 chapters corresponding to the three main topics mentioned above.

In the first chapter, various approaches related to optimization of product design process are presented to evaluate the mechanical design process and mass customization (pp.3, 33 and 53), modeling the product representation (pp.23 and 73), computer support for engineering design (pp.13 and 43), support systems for tolerancing (pp.63, 85 and 95), simulation and optimization tools for structures (pp.105, 149, 159 and169) and for mechanisms (pp. 117, 129 and 139).

The second chapter is dedicated to the optimization of manufacturing systems using : multi-criteria optimization and fuzzy volumes (pp.181 and 191), tooth path generation (p.221), machine-tool behavior (pp. 201, 231, 271, 281 and 313), surface integrity and precision (pp.251, 291, 301), process simulation (pp. 241, 261).

The third chapter presents methodological aspects of integrated design and manufacturing related to solid modeling (pp.325 and 337), collaborative tools and knowledge formalization (pp. 347, 391, 463, 483, 493 and 503), integrating product and process design and innovation (pp. 359, 369, 379, 401 and 473), robust and reliable design (pp. 411, 421, 431, 443 and 453), multi-agent approach in a VR environment (p.513).

The present book is of interest to engineers, researchers, academic staff, and postgraduate students specializing in integrated design and manufacturing in mechanical engineering.

The editors
G. Gogu, D. Coutellier, P. Chedmail, P. Ray

Chapter 1

OPTIMIZATION OF PRODUCT DESIGN PROCESS

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 1–2.

THE NOVEL USE OF A CAPABILITY MATURITY MODEL TO EVALUATE THE MECHANICAL DESIGN PROCESS

I. Egan[1], J. Ritchie[2] and P. D. Gardiner[3]

[1] *UK ATC, Royal Observatory, Edinburgh*

[2] *School Of Engineering and Physical Sciences, Heriot Watt University, Edinburgh*

[3] *School of Management and Languages, Heriot-Watt University, Edinburgh*

Abstract: *This paper describes a process improvement study undertaken at three sites of UK electromechanical engineering companies, using a derivation of the Carnegie-Mellon/SEI Systems Engineering Capability Maturity Model® (SE-CMM^{SM}) called the Process Capability Model – Mechanical Design (PCM-MD). The model was applied within a traditional engineering discipline domain, namely mechanical design. The new assessment tool was piloted on a sample of nine mechanical engineers and eight design engineers as well as some ancillary functions, such as stress and thermal analysis. This was expanded to take into account the views of the downstream manufacturing disciplines and was then subsequently rolled-out into the other companies. The results from these studies support the view that the SE-CMM can be adapted and used in CMM-type assessments for mechanical design process benchmarking.*

Key words: Benchmarking, CMM, design, mechanical engineering, process improvement, project management.

1. INTRODUCTION

The Software Capability Maturity Model®[1] (SW-CMM^{SM} [2])[13], created under the auspices of the Software Engineering Institute (SEI) at Carnegie

[1] *Capability Maturity Model* and CMM® *are registered trademarks in the US Patent and Trademark Office.*

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 3–12.

Mellon University Pittsburgh, has percolated the software world as a methodology with which to understand and help improve the software development process and hence related business capabilities of companies. Its origins are with the United States Department of Defense (DoD) which faced serious cost and schedule problems on several large, software-intensive contracts during the 1980s. A key decision was made to tackle these problems by addressing the software process capability of the organisations awarded contracts by the DoD. Benefits gained from software process improvement using the Software CMM for thirteen companies studied by Herbsleb [8] have included a return on investment ranging from 4:1 to 8.8:1. The yearly reduction in the time to market for the same companies ranged from 15%-23%.

The aim of the research described here is to establish whether or not a CMM-type model can be used to identify areas of strength and weakness in a mechanical design, project-driven environment in order to allow targeted improvements.

2. BACKGROUND

By understanding where an organisation's true capabilities lie, it is possible to make better decisions on how to ensure that customers' needs continue to be met in response to changes in the business environment. Several techniques can be used to measure business efficacy. Collectively these frameworks, methods and tools cover a broad range of approaches including:

- quality awards such as the European Foundation for Quality Management (EFQM) Excellence Model [7] and the Malcolm Baldridge Award [2];
- frameworks such as the ISO 9000 series of quality standards[10];
- methods that people in engineering companies use in their day-to-day work such as Concurrent Engineering practices, Just-in-Time techniques (JIT), Quality Function Deployment (QFD) and Six-Sigma;
- tools for measurement and change such as benchmarking and business process reengineering.

Bartholomew [4] highlights the need to address processes before buying new technology when solving difficulties in business. By optimising the process, the introduction of new technology can be built on a more robust platform. Sinclair and Zairi [16] draw attention to the use of non-financial data at board-room level and point out that this has led to new performance measures at lower levels. Significantly, they claim that, '*once a performance measurement system has been developed, it then becomes a key agent of managing change within the organisation.*' Neely [12] identifies the shift in emphasis due to Total Quality Management (TQM) from '*conformance-to-specification*' towards '*customer satisfaction*' and that the introduction of Six-Sigma and Statistical Process Control has brought focus to the measurement of the process rather than the output. Moreover, Voss [18] states that no matter how good the focus and commitment of the company to meet a particular goal is, it will fail if there are inappropriate processes, or a misaligned infrastructure.

Povey [14] states that, '*the early approach of benchmarking everything*' is now being questioned by managers who are becoming more aware that they should benchmark the things that are strategically important to their organisation. Verma [17] highlights from the Theory of Constraints perspective that, '*every organisation in a real environment is overwhelmed with problems and/or opportunities which need the managers' attention and/or corrective actions. However, limited time, effort and resources make it difficult to act on all such problems or opportunities. Hence, the manager has to find what should be changed, to effectively improve the performance.*' In the same vein, Humphrys [1] has stressed '*the importance of an accurate assessment of the current situation before undertaking any efforts to improve processes.*' The need to think through and target a company's resources on the priority areas is clear from the above observations. The CMM approach addresses these points by providing a means to map the existing engineering, project and support capabilities and identify the areas where change can be made to provide the most help towards achieving business needs. It draws on the TQM philosophy advocated by Deming, Juran and others [9] and provides a structure with which to carry out repeatable assessments.

3. METHODOLOGY

The initial step taken was to build up knowledge of CMM material published to date. As expected there were significant secondary literature sources in the software and systems engineering disciplines [15] but very limited information in other sectors. In order to fully understand the mechanical environment of the company concerned, a period of process data collection with key practitioners was carried out on-site. By understanding the scope of the work carried out in the mechanical design environment, it was possible to identify the most appropriate CMM model to use as a reference tool. Accordingly, the SE-CMM, which uses a continuous representation, was selected as the starting point from which to work [3]. The advantage of the SE-CMM over the SW-CMM is its use of a continuous representation rather than a staged representation. The staged representation provides a blanket maturity level over several process areas whereas the continuous representation provides an individual capability level for each of the eighteen process areas in the model. The continuous representation facilitates the linking of business needs with capability which becomes quite pertinent when deriving recommendations from the assessment results.

After selecting the SE-CMM as the starting point for the mechanical design model, the questions asked, the structure and the language used were carefully and critically assessed for relevance. In general, the structure remained intact although a significant amount of tailoring was required to adapt the language used in the model from a systems to a mechanical engineering environment. This was done with great care and sensitivity to avoid altering the intended meaning of the questions with respect to the design lifecycle. An initial pilot study using questionnaire and interview sessions was conducted at Site 1 using the derived PCM-MD. The questionnaire data provided the assessment team with clear pointers for appropriate questions to be asked in structured interviews. Initial interviews took place with lead practitioners, due to their overall knowledge and experience within the company. This was followed by further interviews with groups of the lead practitioners' co-workers, to ensure that a balanced set of evidence was obtained. This helped ensure that the full views and opinions of the people actively involved with the mechanical design process were taken into account and guarded against bias in the results due to a deliberate or sub-conscious desire to score well.

The pilot study was followed by further assessments at two other locations (Sites 2 & 3). Details of the three sites are given at Table 1.

Table -1. Site Assessment Details

Site	Number of Staff in Mechanical Discipline	Total Number of Staff (Approx)	Date of Assessment
Site 1	42	1000	July 1999
Site 2	32	900	June 2001
Site 3	11	93	November 2001

All of the companies involved are active in the full product lifecycle from concept identification, through design, build and test, to providing on-going customer support. Sites 1 and 2 are involved in batch production of several different product lines whilst Site 3 produces one-off, custom-built products to a small, generally fixed customer base. The process areas used in the pilot questionnaire are shown in Table 2 whilst Fig. 1 shows the model architecture used [3].

Table -2. Process Areas

	ENGINEERING
PA 1	Understand Customer's Needs and Expectations
PA 2	Derive Specification and Allocate Requirements
PA 3	Analyse Possible Solutions
PA 4	Generate Design Solutions
PA 5	Integrate Product Modules
PA 6	Verify and Validate Product
PA 7	Integrate Disciplines
	PROJECT MANAGEMENT
PA 8	Ensure Quality
PA 9	Manage Risk
PA 10	Plan Product Development Effort
PA 11	Monitor and Control Product Development
PA 12	Manage Configurations
	SUPPORT
PA 13	Define Organisation's Mechanical Design Process
PA 14	Improve Organisation's Mechanical Design Process
PA 15	Manage Mechanical Design Support Environment
PA 16	Provide Ongoing Skills and Knowledge
PA 17	Manage Evolution of Product
PA 18	Co-ordination With Suppliers

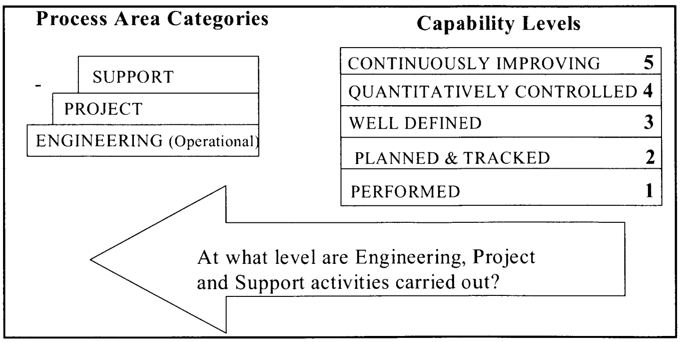


Figure -1. Model Architecture

The process areas relate to the work that is carried out by the practitioners; the capability portion addresses the level to which this work is carried out on a scale ranging from Level 1 to Level 5. Further details on the methodology used can be found at [5].

4. RESULTS AND ANALYSIS

Quantitative data from each site was collected in the form of "Yes", "No", "N/A" (not applicable), "D/K" (do not know) answers to the questions in each questionnaire. Additional qualitative data in the form of written comments made by the participants was also collected. The scoring method used in the initial analysis followed the principles established in the reference model [11], i.e. a minimum of 90% of positive replies is required before a work practice can be regarded as being carried out when the answers to the questionnaires are collated.

The questionnaire results from the pilot study at Site 1 indicated the main hurdles to overcome were in the areas of communication ‘PA 7 Integrate Disciplines’ and quality of processes ‘PA 8 Ensure Quality’. This can be seen in Fig. 2 where these two areas show a sharp contrast to the other engineering and project areas (PA 1-12). The other low responses from the organisational support areas (PA 13-18) increased in level after interviews. The questionnaires were given to a wide range of engineers not all of whom knew that support activities at the organisation level were actually going on; this lack of communication is reflected in the low scores for ‘PA 7 Integrate Disciplines’. Fig. 2 shows the questionnaire only assessment results for Site 2 at the 90% level. This contrasts to Site 1 where several areas, especially in the Engineering segments were seen to fall between Levels 1 and 2.

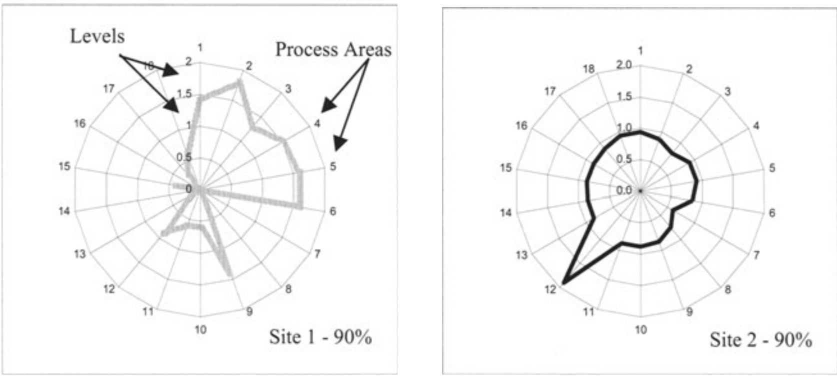


Figure -2. Results for Site 1 and Site 2 at 90% Level

However, by reducing the questionnaire pass level down from 90% it is possible to see the weaker areas emerge as relatively stronger areas move outwards from the centre of the diagram. For example, Fig. 3 shows the effect of setting a pass level of 80% and 50% respectively for Site 2.

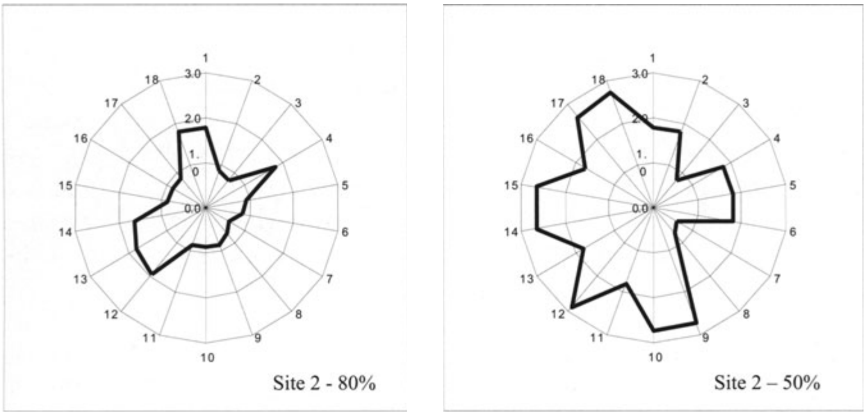


Figure -3. Site 2 Results at 80% and 50% Level

By doing this, it emerges that Process Areas 3, 7 and 8 have particularly low positive support from the questionnaire respondents. This provides a good guide for the structured interviews in terms of selecting the follow-up areas to concentrate on. Further details of this type of approach and detailed results from the pilot study can be found at [6].

On initial review of the questionnaire results, Site 2 gives the appearance of a Level 1 company, with the exception of a strong showing on '*PA 12 Manage Configurations'* (Fig 2). However, on further analysis, using the technique of lowering the pass level described above, problems are indicated with effective communication and control of process quality compared to the other areas (Fig 3). This is expected to change; part of the company improvement strategy is to use tools such as PCM-MD to establish metrics for their mechanical design processes in order to target improvements. By comparison, Site 3 has relatively better communications (it is a much smaller company) and a good interface with its customers. However, it is seen to lag in terms of configuration control and its general approach to ensuring the design progresses in a logical, systematic manner (Fig.4).

When the spider graphs from sites 1, 2 and 3 are combined on the same diagram (Fig. 4) it can be seen that the companies differ in their areas of strength. Subsequent interviews highlighted a common problem area in terms of communication and quality (Process Areas 7&8). Inspection of the quality process area answers in the questionnaires indicates that this problem is biased towards the quality of the site processes that are used rather than the actual quality of the products.

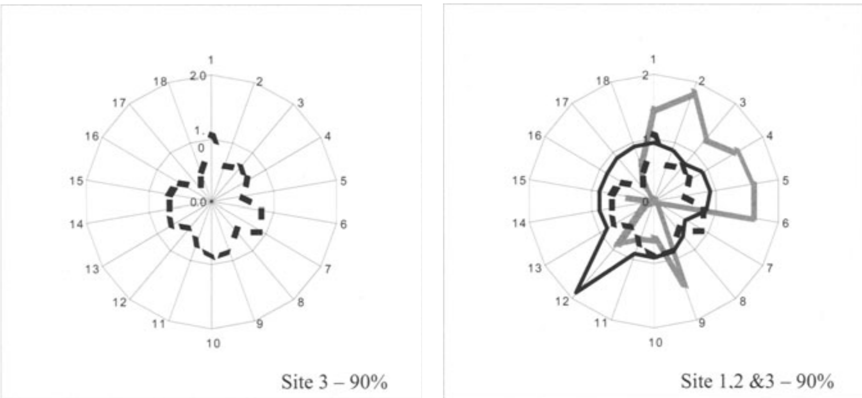


Figure -4. Results at 90% for site 3 and results at 90% for sites 1, 2 and 3

5. DISCUSSIONS AND CONCLUSIONS

One of the key findings of the research to date is the importance of communication and process quality issues. All three assessments have revealed problems in the integration of disciplines, with a particular focus on the communication issues therein. The people who participated in the pilot study interviews and questionnaire sessions (Site 1) recognised the overall picture painted by the assessment results and agreed that they were representative of the company's current state with regard to the mechanical design process. This is an important indication of the model's ability to characterise the activities on the sites on which it is used.

An interesting difference between sites 1 & 2 is that whilst both sites engage in large, multidisciplinary projects and are of a similar size, one maintains a functional approach towards the physical location of their mechanical engineering members whilst the other follows a project team orientation. (Both sites need a multi-discipline input to achieve their product design, manufacture and test tasks.) Although a functional approach was believed to bring advantages to internal discipline activities by the mechanical design management of Site 2, these initial results have highlighted some of the disadvantages.

A key feature of the model is its ability to be used as a reference to help determine the necessary engineering infrastructure that must be in place to deliver a company’s business objectives. By developing a repeatable process assessment tool, engineering management have access to a valuable gauge with which to measure the current and future capability levels of the company's design processes and associated support areas. Although the ability to take a tactical viewpoint to increase capability in specific targeted areas is possible, the results from this type of assessment are typically more strategic in terms of implementation and timescale of business returns.

Overall, the research has generated output that suggests that the adaptation of the SE-CMM for use in the mechanical engineering environment is both viable and of use to the companies that have participated in the assessments. Feedback from carrying out the first three studies suggests that the reports provided to the companies based on the assessment process accurately captured the strengths and weaknesses of the areas assessed. This allowed action plans to address areas of weakness to be drawn up by the companies involved. The only negative comment was a tendency for the questionnaire sessions to be tedious to complete at times due to the number of questions (twelve) that were repeated for each individual process area to establish its capability level. The approach towards gathering questionnaire data must be tuned to the level of the company if participants are not to be alienated from taking part.

Of note, the information gathered was focussed more on the actual work carried out by engineers. This differs from the more strategically aimed performance assessment models that address higher-level aspects of a company's business. The PCM-MD can thus be seen to be a complementary rather than alternative tool for a business keen to understand what areas of its mechanical design process are effective.

References

[1] A. April, A. Abran, E. Merlo, "Process assurance audits: lessons learned" International Conference on Software Engineering, Kyoto, Japan, Apr 19-25 1998.

[2] "BaldridgeQuality Award", http://www.quality.nist.gov

[3] R. Bate, D Kuhn, C. Wells, "A Systems Engineering Capability Maturity Model V 1.1" *CMU/SEI-95-MM-003*, Carnegie Mellon University. Pittsburgh, Nov 1995.

[4] D. Bartholomew, "Process is back" *Industry Week*, Vol 248, Iss 20, pp31-35, 1999.

[5] I. Egan, P.D. Gardiner, I. Black, J.M. Ritchie, "Business Process Measurement - Adapting and Using a Capability Maturity Model in a New Domain," *15th World Congress on Project Management*, London, UK, May 2000.

[6] I. Egan, I. Black, J.M. Ritchie, P.D. Gardiner, "Process improvement in mechanical design using a derivation of the systems engineering capability maturity model." *33rd International MATADOR Conference*, UMIST, Manchester, UK, July 2000.

[7] "European Foundation Quality Model," www.efqm.org

[8] J. Herbsleb, "Benefits of CMM-based software process improvement: initial results," *CMU/SEI-94-TR-013*, Software Engineering Institute, Carnegie Mellon University. Pittsburgh, 1994.

[9] K. Ishikawa, *"What is Total Quality Control?"* pp14-20, Prentice Hall, 1985.

[10] "ISO 9000-9004 Quality Management and Quality Assurance Standards, 1987" http://www.iso.ch/

[11] D. A. Kuhn, et al. "A description of the Systems Engineering Capability Maturity Model Appraisal Method V 1.1' *CMU/SEI-96-HB-004*, 3-21 Carnegie Mellon University. Pittsburgh (March 1996).

[12] A. Neely, M. Gregory, K. Platts, "Performance measurement system design a literature review and research agenda," *International Journal of Operations and Production Management*, Vol 15, No 4, pp80-116, 1995.

[13] M.C. Paulk, B. Curtis, M.B. Chrissis, C.V. Weber, "Capability Maturity Model for Software V1.1" *CMU/SEI-93-TR-024*, Carnegie Mellon University. Pittsburgh, Feb 1993.

[14] B. Povey, "The development of a best practice business process improvement methodology." *Benchmarking for Quality Management and Technology*, Vol 5, No 1, pp27-44 1998.

[15] Proceedings of Ninth Annual International Symposium of the International Council on Systems Engineering, Brighton, UK June 1999.

[16] D Sinclair, M. Zairi. "Effective process management through performance measurement," *Business Process Re-engineering & Management Journal*, Vol 1, No1, pp75-88, 1995.

[17] R. Verma, "Management science theory of constraints/optimized production technology and local optimization," *International Journal of Management Science*, Vol 25, No 2, pp 189-200, 1997.

[18] C.A. Voss, "Alternative paradigms for manufacturing strategy," *International Journal of Operations & Production Management*, Vol 15, No 4, pp 5-16, 1995.

COMPUTER SUPPORT FOR INSIGHTFUL ENGINEERING DESIGN BASED ON GENERIC AND RIGOROUS PRINCIPLES OF SYMBOLIC ALGEBRA

H. Sawada[1] and X-T. Yan[2]
[1] National Insttitute of Advanced Industrial Science and Technology, Japan
[2] CAD Centre, University of Strathclyde, UK

Abstract: Engineering designers make many design decisions for selecting optimal options based on insights into underlying relationships among design parameters under consideration. As design proceeds, more and more design parameters are introduced and concretized. The increase of design parameters poses challenges and difficulties for designers to gain deep and rich insights, and to make informed design decisions. Designers therefore need to be adequately supported to explore design alternatives to search for the optimal one. This research aims to provide design support tools for gaining better insights, and to integrate them into a design support system that facilitates designers' exploration of design solutions. To achieve this aim, new constraint solving methods based on generic and rigorous principles of symbolic algebra have been derived. These methods overcome difficulties of conventional methods, and provide necessary and sufficient information for making insightful design decisions. The advantages of this approach will be shown through a case study.

Key words: insightful design, multidisciplinary design, non-linear constraints

1. INTRODUCTION

Designers often use a solution-focused approach[1]. During a design process, designers propose possible design solutions[2]. Initially, they are often abstract and incomplete.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 13–22.

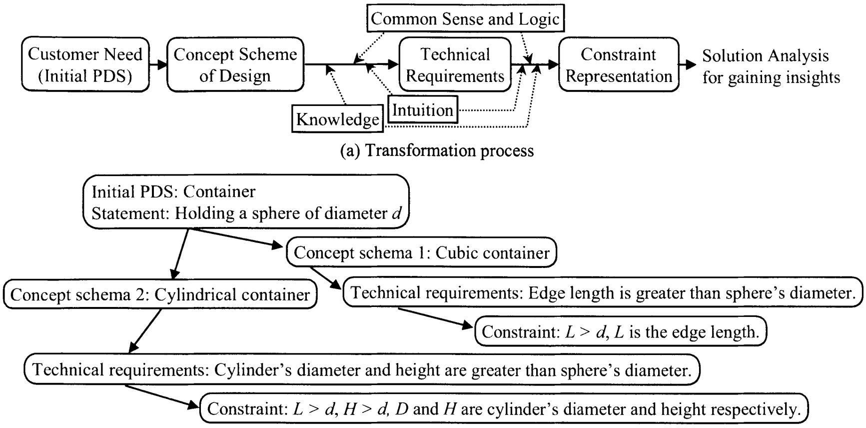
Through investigation and elaboration of them, designers gain insights into the proposed design solutions, customers' requirements, their relationships, and the design problem. Based on the insights obtained, designers can make informed design decisions, and explore design alternatives to progress towards an optimal design solution.

At the beginning of engineering design, a design solution is represented as a qualitative concept. Designers need to transform such an abstract expression into a certain quantitative representation by introducing design parameters. The qualitative relationships between design parameters are usually complex and highly coupled, and there are too many unknown design parameters. An in-depth analysis of these incomplete design solutions gives designers insights including underlying relationships between design parameters. However, it is difficult to analyze an incomplete design solution using conventional analysis tools[3]. Thus, designers often have to use a trial and error method to determine the design parameter values. Any design decision made under these circumstances results in a less optimal solution. This research aims at overcoming these difficulties by providing designers with informative and insightful design support tools based on generic and rigorous principles of symbolic algebra, and integrating them into a design support system that facilitates designers' exploration of design alternatives.

2. ALGEBRAIC CONSTRAINT REPRESENTATION

Design commences with a statement of customer need. This is formally translated into an initial Product Design Specification (PDS), which includes short statements about customers' requirements[2]. Based on it, several design concepts are generated. Following the concept generation process, technical requirements of the product are established by using designers' common sense and logic, intuition, different domain knowledge and reasoning process. These technical requirements are represented as quantitative constraints thereafter.

The purpose of such transformation is to provide a scientific basis for product modeling by introducing definitive and precise descriptions of the product used to gain insights into the design problem and design solutions through solution analysis. Figure-1 shows this transformation process and an example. It is known that most technical requirements can be converted into algebraic expressions. This research uses this observation as an axiom. Since algebraic constraint representation is generic, rigorous and domain independent, it is ideally suitable for multidisciplinary and multi-level complexity engineering design and modeling.



(a) Transformation process

(b) Example of the transformation

Figure -1. Transformation of customer need to constraint representation

3. ALGEBRAIC CONSTRAINT SOLVER

Solution analysis is a key activity in gaining insights into a design solution space. These insights include possible design parameter values[4], optimized numerical solutions[5], fundamental relationships between design parameters[6], and conflicts in a design solution[7]. Since conventional design support tools are limited in providing the above information, designers are obliged to solve a design problem by trial and error[3]. To overcome such difficulties, new algebraic constraint solving methods have been derived to provide a comprehensive set of support tools[8]. They have been integrated into the algebraic constraint solver consisting of four modules: parameter value module, solution space module, conflict detection module, and optimization module.

3.1 Parameter value module

Design parameter values are not always determined based on rational decisions. Rather, several design parameters are assigned numerical values considered to be appropriate based on designers' knowledge and experience. However, since design requirements are highly coupled, designers often encounter difficulty in finding possible assignment of numerical values. The parameter value module provides the following functions[8].

1. to make clear whether a design solution exists;
2. to compute numerical solutions to a given design problem;

3.2 Solution space module

Designers need to understand the underlying relationships among design parameters, and often try to ignore unimportant information and focus on fundamental relationships[6]. However, since relationships between design parameters are highly coupled, it is difficult to extract such fundamental relationships. This module establishes explicit relationships among design parameters, and shows them as a partial design solution space in the form of a two-dimensional graph[8].

3.3 Conflict detection module

Several constraints may be potentially inconsistent with each other due to possible conflicts in design requirements. Detection of such a conflict raises a designer's understanding of and a need for a trade-off between design requirements. This module reveals conflicts by identifying inconsistent constraints[8].

3.4 Optimization module

Designers need to conduct optimization based on an incomplete design solution to get an idea of design parameter values that make the design solution optimal. The optimization module provides the following functions[8].

1. to investigate whether the given objective function has an optimal value;
2. to determine the design parameter values to optimize the given objective function, when the optimal value actually exists;

3.5 Interaction with a designer

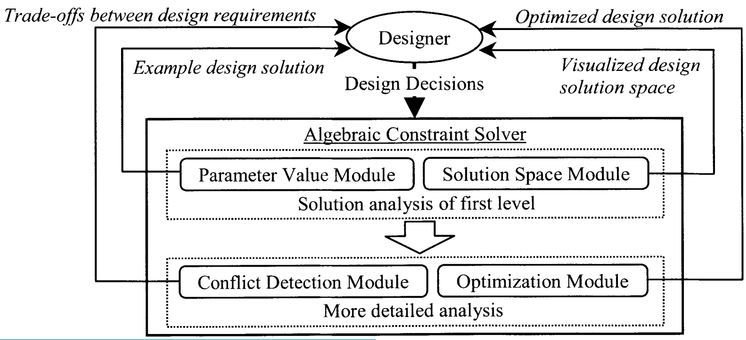


Figure -2. Interaction between a designer and the algebraic constraint solver

Figure-2 shows the interaction between a designer and the algebraic constraint solver. The parameter value module and the solution space module give a designer an intuitive and rough idea of a design solution. On the other hand, the conflict detection module and the optimization module provide more detailed information. The designer can get the feedback and make use of it for further design decision-making. Since all the functions of the algebraic constraint solver are based on generic and rigorous principles of symbolic algebra, all the analysis results are guaranteed to be correct. This implies that a computational mistake due to rounding errors, cancellation of digits, and loss of trailing of digits, is completely eliminated.

4. DESIGN PROCESS MODEL

The authors regard a design as a process of defining constraints and solving a design problem using constraints. Figure-3 shows the proposed design process model. It defines the framework of the prototype system.

4.1 Design concept description

When a designer works on a design problem, he/she tries to decompose the design problem into smaller sub-concepts. That is, a design concept has a *part-of* hierarchy. For each sub-concept, design is conducted separately. Thus, a designer requires a design support system to record the part-of hierarchy, and to provide multiple design environments each of which enables the designer to concentrate on the sub-concept design.

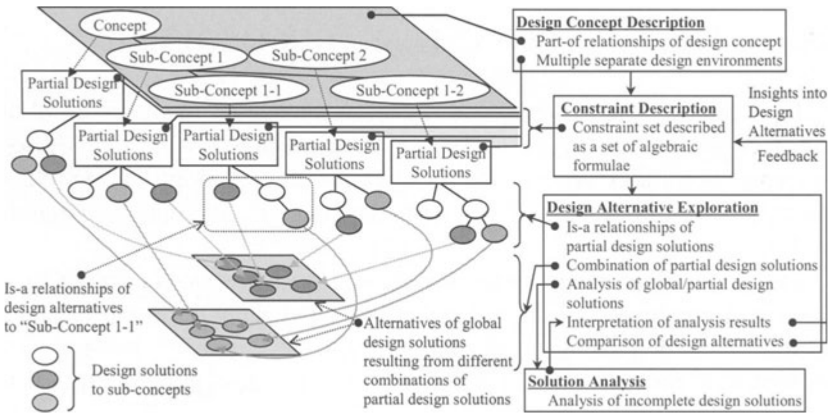


Figure -3. Design process model

4.2 Constraint description

A design solution is concretized by introducing design parameters and assigning their values. It is argued that most constraints can be converted into algebraic expressions. It takes an extremely long time for a designer to describe all the constraints by him/herself. This implies that designers need a support for describing algebraic constraints. Since it is useful to provide a database storing commonly used components[4,9], such a library representing a database has been created.

4.3 Design alternative exploration

For each sub-concept, design alternatives are generated as a result of different options of design decisions. These design alternatives can be classified in the tree structure representing *is-a* relationships between them.

The partial design solutions to sub-concepts are analyzed, compared, and combined into a global design solution. These partial/global design solutions are usually incomplete. Analysis of the incomplete design solutions is conducted to gain insights for further design decision-making. The insights gained are fed back to each partial design process. Each partial design solution should be investigated more precisely based on the insights.

The above discussion leads to the necessity of the following design support functions. Firstly, all the generated design alternatives are recorded as well as the *is-a* relationships between them so that the designer can explore them. Secondly, it is necessary to store and present all the results of analysis associated with the corresponding global/partial design solutions.

5. PROTOTYPE SYSTEM – DECOSOLVER

Figure-4 shows the overview of the prototype system named DeCoSolver (Design Constraint Solver). It consists of three modules in addition to the algebraic constraint solver: the design concept description module, the constraint description module, and the design alternative exploration module.

Design concept description module consists of two sub-modules: *the concept decomposer* and *the concept-tree*. The concept decomposer provides facilities for describing a design concept with part-of hierarchy. The concept-tree provides a facility for switching design sub-concepts that designers work on.

Constraint description module provides facilities for describing algebraic constraints. All the constraints are described via *the constraint editor* with

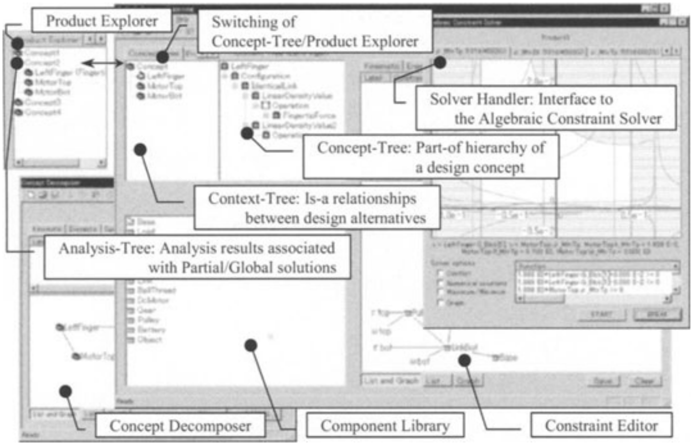


Figure -4. Overview of DeCoSolver

the support of *the component library*, which is a database of the commonly used components. Designers can select necessary components, and construct a product model by combining them. The database allows designers to add new components.

Design alternative exploration module provides an environment in which designers are encouraged to explore design alternatives. It consists of two sub-modules: *the context-tree* and *the product explorer*. The context-tree holds all the created design solutions in the form of an *is-a* hierarchy. The product explorer provides facilities for integrating partial design solutions into a global design solution. It also provides the designer with *the solver handler* – the interface to the algebraic constraint solver. Designers can analyze partial/global design solutions via the solver handler. All the results of analysis are held in the form *the analysis-tree* associated with the corresponding design solution.

6. DESIGN EXAMPLE

This section demonstrates the application of DeCoSolver through a design example of a heat pump system, which involves thermodynamic and hydrodynamic issues. Figure-5 shows the heat pump system used here. It consists of a compressor, a condenser, an expansion valve and an evaporator. Arrows "⟶" and "→" show the flow of refrigerant and water respectively. The refrigerant evaporates at the evaporator and collects heat from the water. Then, it is adiabatically compressed by the compressor. In the condenser, it condenses and releases heat to heat up the water.

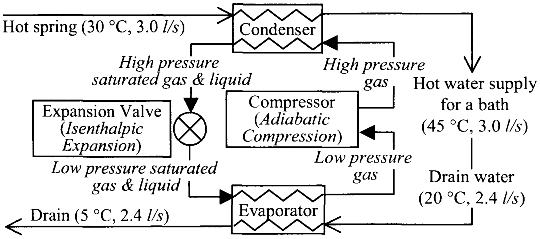


Figure -5. A heat pump system

Design task

A designer should determine the following design parameter values:
Tc, Te: Temperature of condensation and evaporation [K]
Ac, Ae: Heat transfer area of the condenser and evaporator [m^2]
Pd: Discharging pressure of the compressor [Pa]
κ: Compression ratio of the compressor
Qr: Mass flow rate of the refrigerant [kg/s]

Due to the loop structure of the heat pump system and non-linearity of thermodynamic properties, these design parameters are coupled in a complex manner. Conventionally, a designer has to determine the design parameter values by trial and error. DeCoSolver can help a designer to gain insights into the underlying relationships and to determine the design parameter values without any trial and error.

Design procedure

1. Constructing the product model

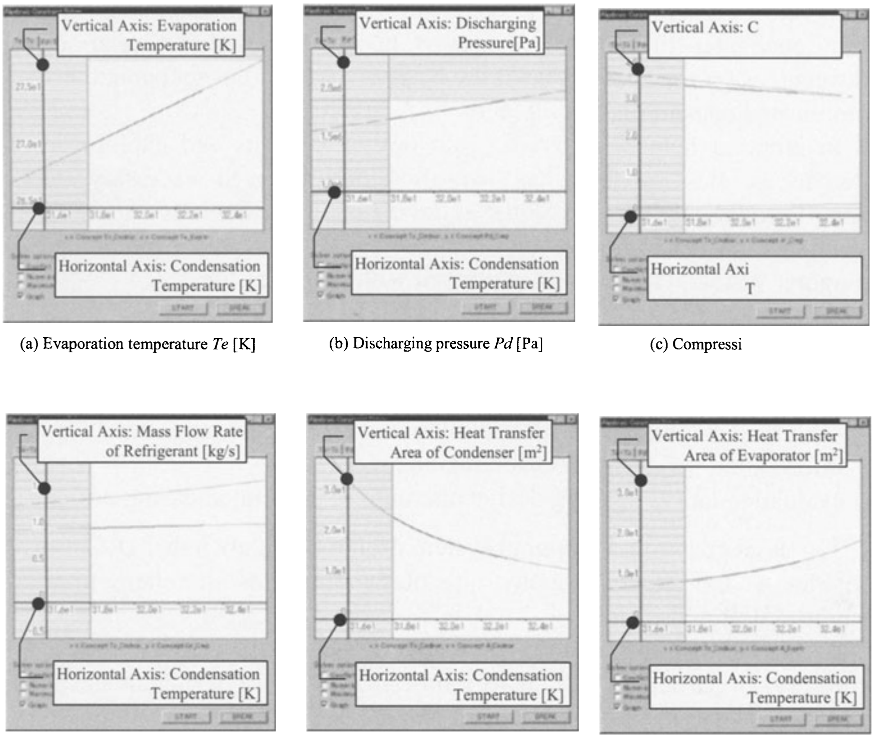
Necessary components are selected from the component library and configured into a product model via the constraint editor. DeCoSolver automatically generates a global constraint set of the heat pump system.

2. Gaining insights

Graphs representing the explicit relationships between *Tc* [K] and the other design parameters are drawn (Figure-6). In these graphs, while points on black lines are valid, points on gray lines or in the hatched area violate some constraints. From these graphs, the following insights can be gained.

- As *Tc* increases, *Te* and *Pd* also increase almost linearly.
- *Qr* and κ are almost unchanged.
- As *Tc* increases, *Ac* decreases and *Ae* increases non-linearly.

The above information implies that the designer can exclude *Qr* and κ from consideration at the moment. Suppose the designer focuses on *Ac* and *Ae*. Because a small surface area results in small equipment, it is generally preferable to have a small value for the summation of *Ac* and *Ae*.



(a) Evaporation temperature *Te* [K] (b) Discharging pressure *Pd* [Pa] (c) Compressi

(d) Mass flow rate of refrigerant *Qr* [kg/s] (e) Heat transfer area of condenser *Ac* [m^2] (f) Heat transfer area of evaporator *Ae* [m^2]

Figure -6. Relationships between the design parameters

3. Determining the design parameter values
   With the support of DeCoSolver, *Tc* can be determined so that the total heat transfer area is minimized. In this case, *Tc* is determined as 323 [K], giving a total heat transfer area of 29.76 [m^2]. The numerical value of *Tc* is used to determine the other design parameter values.

7. CONCLUSION

Engineering design is an intensive decision making process. Informed and insightful design decisions lead to a high quality product design with less or no rework. Therefore, it is necessary for designers to gain deep insights into both the given design problem and incomplete design solutions generated during a design process.

Usually, these insights can be gained through experience by understanding underlying relationships between design parameters quantitatively. These insights can then be used to explore design alternatives.

However, designers, especially those who work in multidisciplinary design, often encounter difficulties in extracting such underlying relationships between design parameters due to the large number of design parameters and complicated constraints among them.

In order to help designers to gain design insights and explore design alternatives, this research has identified and derived necessary design support tools, which overcome conventional difficulties in handling incomplete design solutions, and then integrated them into a design support computer system, DeCoSolver, equipped with facilities for:

- supporting the description of a design concept with a part-of hierarchy;
- supporting the establishment of quantitative relationships between design parameters as algebraic formulae;
- analyzing incomplete design solutions and extracting insightful information for design decision making;
- evaluating and comparing design alternatives generated during design.

The case study of a heat pump system design has shown that DeCoSolver provides a designer with many quantitative insights that help to make insightful design decisions.

Before putting DeCoSolver to practical use, further research work is necessary. In particular, linking with other software tools such as CAD and databases of commercial components is indispensable. It may improve the product modeling ability and the ability to explore design alternatives.

REFERENCES

1. Cross, Nigel, *Engineering Design Methods (2nd Ed)*. John Wiley & Sons, 1994.
2. Pugh, Stuart, *Total Design*. Addison-Wesley Publishers, 1990.
3. Sawada H, Yan XT. Applying generic constraint solving techniques in providing insights into engineering design. Proceedings of ICED 01, Design Methods for Performance and Sustainability; Professional Engineering Publishing Ltd, 2001; 123-130
4. Thornton A.C., Johnson A.L. CADET: a software support tool for constraint processes in embodiment design. Research in Engineering Design 1996; 8: 1-13
5. Thompson, David, *Design Analysis*. Cambridge University Press, 1999.
6. Hoover SP, Rinderle JR. Abstractions, design views and focusing. Proceedings of DTM'94; ASME, 1994; 68: 115-129
7. Oh, V., Sharpe J.E.E. "Conflict Management in an Interdisciplinary Design Environment." In *AI System Support for Conceptual Design*, J.E.E. Sharpe ed. Springer-Verlag, 1996; 298-318
8. Sawada H, Yan XT. Application of Gröbner basis and quantifier elimination in engineering design: an introduction for engineers. Proceedings of ASCM2001; World Scientific, 2001: 141-150
9. Park M.-W., Cha J.-H., Park J.-H., Kang M. Development of an intelligent design system for embodiment design of machine tools. Annals of the CIRP 1999; 48: 329-332

MODELLING PRODUCT REPRESENTATIONS USED IN AUTOMOBILE BODY DESIGN

C. A McMahon and P. Kugathasan
University of Bristol, Faculty of Engineering, England

Abstract: Engineers use many different models in the course of their work, varying from physical models through graphical to mathematical and computational models. This paper describes a study of the automobile body-in-white design process to identify the models used in the course of this process. It explains the experimental method of using Petri net diagrams to describe the design process in an automotive company, and then using these models in conjunction with the company's engineers to identify the models used by the engineers at each stage in the process. It will describe the overall use of models in the process, and also one particular aspect of the process in some detail, to illustrate the application of the experimental technique. Representations involve the initial use of sketches and stylists' renderings, then the development of the body form through physical models. Specialist models are used for formability and structural analysis, and for the development of tooling. The paper concludes with some suggestions on the lessons to be learned for process modelling and computer support tools.

Key words: information modelling, design process, Petri nets, automobile engineering.

1. INTRODUCTION

Successful implementation of concurrent engineering and the development of computer techniques for its support depend upon a detailed understanding of the processes and activities involved in product introduction, and of the product representations used by the participants. This paper will describe a study of a particular design process - that for the design of automobile bodies - to identify the product representations (which

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 23–32.

will be called design models) used in the course of this process. The paper will explain the experimental method used, which was to develop Petri net diagrams to describe the design process in an automotive company, and then to use these diagrams in conjunction with the company's engineers to identify the design representations used by the engineers at each stage in the process. It will describe the overall use of models in the process, and also one particular aspect of the process in some detail, to illustrate the application of the experimental technique.

2. MODELLING PRODUCTS AND THE PRODUCT DESIGN PROCESS

Three aspects of the design/manufacturing process were considered in the study. These are the information entities describing the artefact and the associated design and manufacturing information, the information flow and the participant activities or tasks (activities which are carried out on information entities to modify, develop or apply them in some way).

Information entities describe information such as documents, drawings and computer-aided design models. Understanding these entities allows engineers and designers to identify where particular information is available and to be precise about the context, for example to assist automatic data exchange. Modelling of information flows in the design process enables engineers and designers to identify where information they have generated will be used and by whom, and what are the constraints in the activities. Modelling participant activities allows engineers to understand what use will be made of the information that is passed on to later activities. They are then able to pass on information in a usable form.

A number of approaches have been applied in the modelling of product and process in engineering design. These include the ICAM Definition Method (IDEF) series of modelling methods, [1], those connected with the international STEP standard for the representation of product data, ISO-10303 [2], data flow diagrams [3], and the Fusion and UML modelling languages [4][5] among others [6]. In the present work, the Petri net approach was used, because of its ability to clearly distinguish between resource and activity in a simple, easy to understand diagram, and because of its capability (not used in the work described here) for modelling dynamic processes [7]. The Petri net method was incorporated into a hypertext system in order to provide a mechanism for presenting a structured map of tasks, information resources, relationships and flows in an easily accessible form to act as a discussion support tool for designers and engineers. The system provides, as shown in Figure 1, a high level block diagram model of the overall product introduction process activities and the relationships between the major groups involved, together with a Petri net description of the relationships and data flows between the activities of the product introduction process, the product models and other resources required for their operation. Information frames provide descriptive information or more formal models of activities or product models referenced from the Petri net.

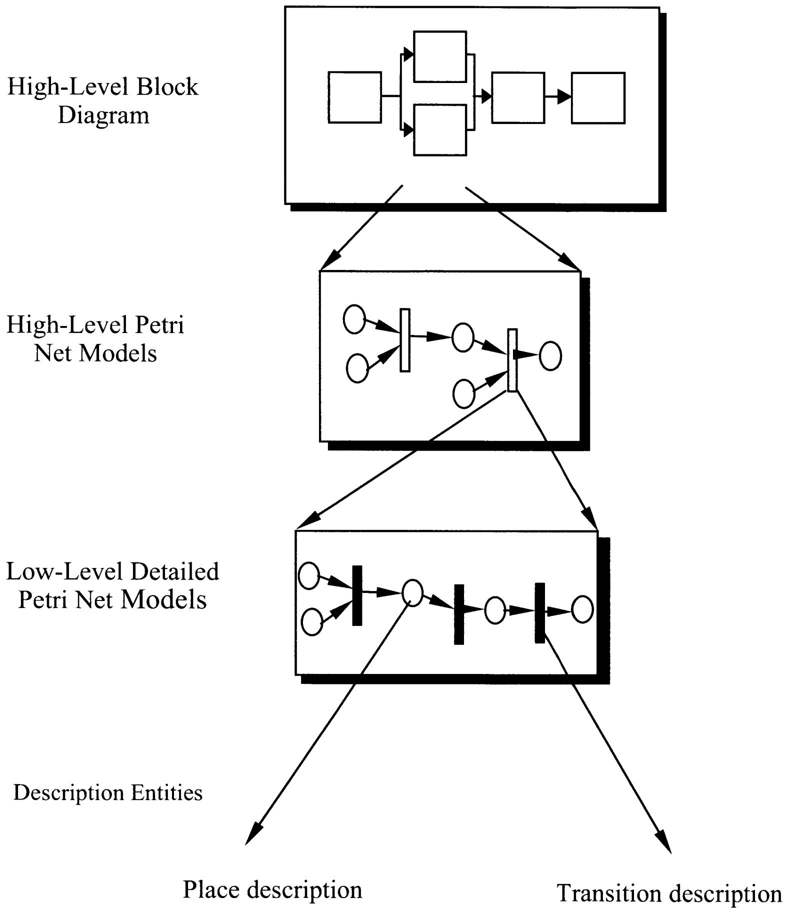


Figure –1. The overall relationship between model elements

2.1 Petri Net Models.

Petri nets provide a method of modelling and representing processes and their relationship to the resources required for them to take place. A simple Petri net graph (Figure 2) is composed of a set of *places* represented by circles, a set of *transitions* represented by bars, a set of *directed arcs* connecting places and transitions, an *input function* and an *output function*. In the present use of Petri nets, the transitions are used to represent activities in the overall process, and the places are used to indicate the information and other resources required for the activity to take place. The input and output functions relate respectively to the resources that are precedent to transitions, and to those generated by the transitions. The direction of flow of the net is indicated by the directed arcs linking places and transitions. The availability of resource at a place can be indicated by a token, shown as a dot, drawn at a place. When the transition fires, as shown in Figure 2, tokens are removed from the input places and added to the output places, thus enabling downstream activities to fire if appropriate. Tokens *mark* the Petri net, which can then be used to model the dynamic state of the network, but this capability is not used in the present application.

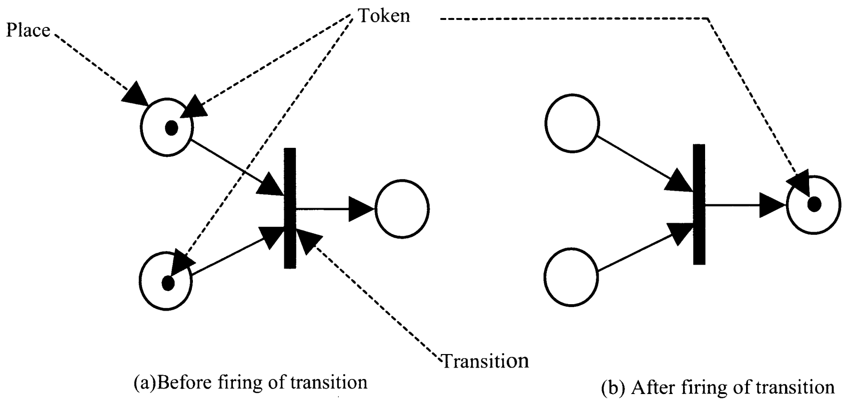


Figure -2. A Simple Petri Net graph

3. THE BODY-IN-WHITE DESIGN PROCESS

The design and development of an automobile is very complex, involving several internal departments of a company and many external contractors. The body-in-white (BIW), the basic structure of the vehicle, is a key element of the design and development process. It is that part of an automobile comprising the pressed sheet metal body panels joined together by welding or some other means. Its purpose is to provide the external, visible part of the body, the passenger shell and structure onto which other components of the automobile may be attached.

The principal elements of the BIW development process were identified, along with some of the main design participants, through discussions and interviews with engineers in an automotive manufacturing company (AM).

In order to support discussions, first of all a high level block diagram of the participant groups was developed and incorporated in the hypertext system. This is shown in Figure 3. Detailed Petri Net descriptions of the participant activities represented by each of the eight blocks in the diagram were then developed. The description for one aspect, packaging design, is shown in Figure 4. This shows the location of the packaging activity within the overall information flow and also information about the resources and processes carried out within the packaging activity.

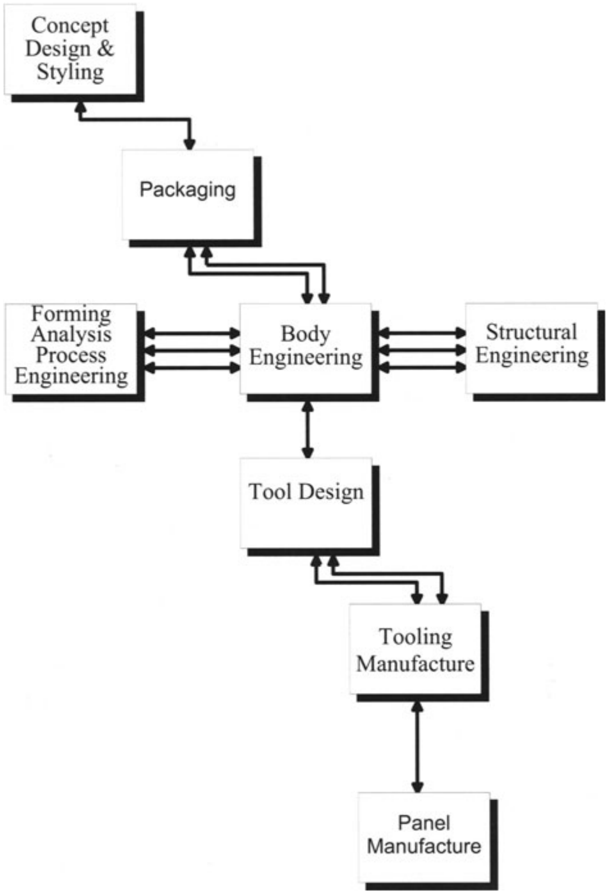


Figure –3. High level block diagram of the automobile body-in-white design process

The "features list" in Figure 4 describes the product features that the body needs to accommodate, such as a sun-roof. Information describing objects and transitions was attached to the Petri net diagram using hypertext links. The whole diagram was developed over a period of weeks in which a number of discussions took place with engineers, and the diagram, together with the descriptive information, was used as the basis for discussions.

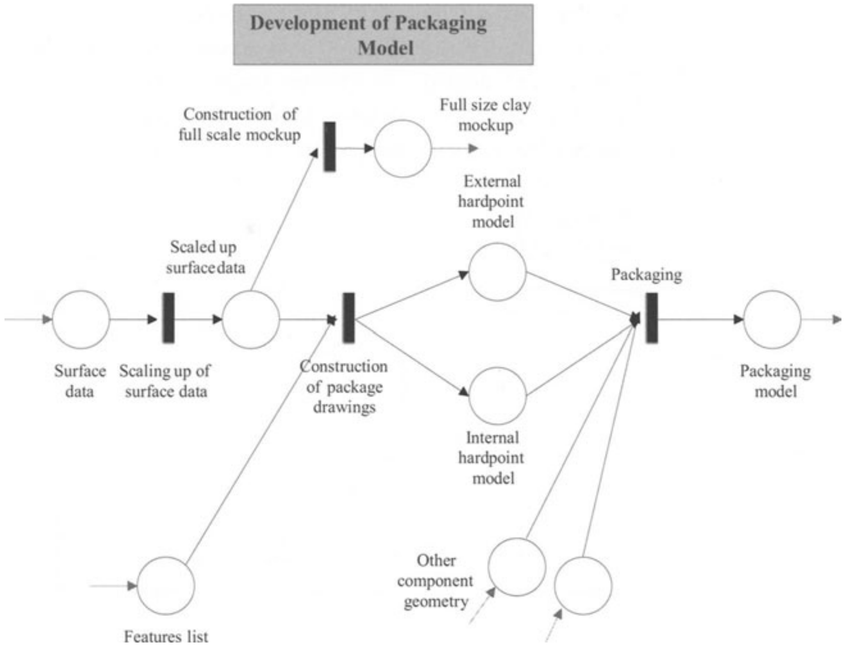


Figure –4. Top-level description for packaging design

3.1 The Process and Participant Groups

Figure 3 shows the overall schematic of the eight participant groups within the design and manufacture of the BIW in the AM. This is the high-level model presented in the hypertext system. The arrows indicate the direction of the overall information flow. In more detail, the activities shown in the figure are as follows:

– **Concept Design and Styling**: The concept design and styling team develop the exterior style of the vehicle. A number of different styles will be developed, and only a small number of these will be developed in detail for which clay models are made.
– **Packaging**: The packaging team ensure that all components of the vehicle, such as engine and drivetrain, fit within the envelope defined by the design team.
– **Body Engineering**: The body engineering team develop the structural sections as well as the individual body panels.

- **Structural Engineering**: The structural engineering team carry out simulations, using various load conditions, of the vehicle structure to ensure that the sections and panels satisfy legal and crashworthiness requirements and also maintain satisfactory stiffness.
- **Forming Analysis**: Forming analysis investigates the formability of the body panels which are developed. They ensure that panels may be manufactured economically and without problems, such as tearing.
- **Process engineering**: In BIW development, process engineering is the development of a set of plans that define the pressing operations that are necessary to form a panel.
- **Press tooling design and manufacture**: Press tooling design is based on the process plans that are developed by the process engineering team. A separate press tool is required for each pressing operation, where each press tool is manufactured through a process of lost polystyrene casting.
- **Panel manufacturing**: The finished press tools are initially tested to ensure that the panels are formed correctly. Rework of the press tool may have to be carried out at this stage. Once the press tool has been checked it can be used in full-scale production.

3.2 Design and Manufacturing Models

In order to carry out the various assessments, analyses, simulations and planning activities of BIW engineering, a number of model representations are used, each being required for a specialist application to be carried out. Figure 4 shows the relationship between model representation and process for the development of packaging models. This shows that the packaging team works principally with 2D drawings and computer-based surface models. The packaging models are essential to ensure that all the components of the vehicle fit within the desired envelope. This means that several departments, including the concept design department and body engineering, must work together to develop a feasible packaging model. Development of packaging models and drawings begins with receipt of scaled up surface data, which has been taken from the selected clay model. A features list is used to identify the important attributes of the vehicle that must be included, for example safety equipment and carry over items from previous models. Construction drawings, basic drawings indicating the spatial arrangement of the basic elements of the vehicle, are generated. Objects, representing passengers and the engine, are added to the drawing. Internal and external hardpoint models, defining critical dimensions and locations of parts on the vehicle such as the engine and passenger head clearance, are used at this stage. Surface data from the concept design and styling team is used to generate the computer-based surface packaging models. Many legal and safety requirements for the vehicle are satisfied during this stage. The outputs from this stage in the design process are three-dimensional surface-based packaging models as well as two dimensional full sized drawings.

Similar diagrams to that shown in Figure 4 were developed for each of the other activities in Figure 3. Where necessary, lower level diagrams were used to provide more detail in particular areas, and to stop the diagram becoming too complex. These diagrams showed the following additional use of design representations:

- **Concept Design and Styling**: this area uses 2D sketches and a variety of sizes of 3D clay models of the concept vehicles. When a suitable concept clay model has been chosen, its surface is digitised to generate the computer-based surface model which is used in subsequent activities.
- **Body Engineering**: Surface models are used to develop the individual panels. Details such as flanges are added to the panel surface model.
- **Structural Engineering**: Structural development of the vehicle begins with simple beam finite element models, developing into more refined models of the entire BIW as further information is obtained.
- **Forming Analysis**: The forming analysis team use finite element models of panels to carry out the simulations of the panel being formed.
- **Process Engineering**: Initial concept process plans are created from isometric drawings of the panels. These become more refined as more information is known about the panel design. These process flow diagrams use information taken from the surface model of the panel.
- **Press Tooling Design and Manufacture**: Surface models of the panels are used to define the press tool faces, which are used to manufacture the 3D polystyrene models used in lost pattern casting operations.
- **Panel Manufacturing**: The actual press tools are placed on a production line ready for manufacturing the panels.

4. DISCUSSION

Throughout the design of the BIW a variety of model representations are used, each required to carry out a specialist application. Often, the same information is being modelled but is presented in different ways. Shah and Mäntylä [8] use the term viewpoint dependent models to describe these various specialist representations of the same product or part. The knowledge about the nature of the viewpoint dependent models is critical to the project timescale in terms of ensuring that the correct information is available at the required time. Viewpoint dependent models are often a simplification, embellishment or modification of another model. At present, models are transferred between departments and manually modified, for example by removing detail, to obtain the appropriate viewpoint dependent models. Reducing the time taken to generate these models would reduce the design time and relieve the engineer of some of the mundane tasks. To achieve this, it is important to identify the variety of viewpoint models that are used along with the processes that are carried out to generate them. As more and more of the models that are being used during the design and manufacture of the BIW are computer-based it may be important to begin to automate the generation of some of these viewpoint models. The key factor in understanding viewpoint dependency is to understand the inter-relationships between the models, which is why it is important to understand tasks and information flows as well as information entities

The application of the Petri net modelling approach to the BIW design process has resulted in a model that provides an easily accessible design support tool for engineers during the product introduction process and the use of hypertext allows the engineer to navigate through it with relatively little training. The clarity of the simple Petri net models ensures that engineers know at a glance whether it is a process activity or a product information entity which they are looking at. By merely selecting that node of the Petri net model the engineer is presented with detailed information.

The current Petri net model indicates the type of resources that are required to carry out particular activities, such as the format of a component model to be placed in a vehicle envelope. This is an improvement over simple Petri nets without adding greatly to the complexity of the resulting diagrams. However, the type of resource is only indicated in the information frame, and not in the Petri net. In addition, the diagrams contain no indication of the time required for an activity, and are therefore of limited value for management purposes. There are several existing Petri net extensions, such as coloured Petri nets or Predicate/Transition nets [9], which may overcome these problems. The drawback of their use would be greater complexity being introduced in to the model making it difficult for engineers to interpret.

The Petri net model has been demonstrated to engineers and designers from the AM and has been well received. Suggestions for improvements have included the representation of time within the model as noted above, and incorporation of an indication of the availability of information, or that it has been superseded and can no longer be used for a particular process.

The application that has been presented covers a small portion of vehicle product introduction. There will be a number of problems with scaling the model up to cover the entire process, including storage of more information, ensuring that the information is valid and up to date and the added complexity of maintaining a much larger Petri net based model.

5. CONCLUSIONS

In order to develop useful support tools for the product design process a good understanding should be obtained of the activities involved in the process, and of the models and representations used by those activities. This paper has introduced a method for modelling the product design process, based on Petri net modelling and embedded in a hypertext system. It has been shown how this approach has indicated the different representations used in the automotive body-in white design process, and the implications of the wide range of representations have been discussed.

6. ACKNOWLEDGEMENTS

The authors are grateful for the support of a research studentship from the EPSRC and to the engineers of AM company for their contribution of expertise in the development of this research.

REFERENCES

1. IDEFine-0 User Manual, Micromatch Limited, 1989
2. STEP, *Standard for the Exchange of Product Model Data*, ISO TC184/SC4/WG1, DP 10303, October, 1988.
3. Downs, E, Clare, P, and Coe, I., *Structured Systems Analysis and Design Method: Application and Context*, Englewood Cliffs: Prentice Hall, 1988
4. Coleman, D, Arnold, P, Bodoff, S, Dollin, C, Gilchrist, H, Hayes, F, and Jeremes, P, "Object-oriented development - The Fusion method", Englewood Cliffs: Prentice-Hall, 1994
5. Quatrani, T, *Visual Modeling with Rational Rose and UML*, Reading, MA: Addison-Wesley, 1999
6. Eastman, C.M., and Fereshetian, N, Information models for using in product design: a comparison, Computer-Aided Design , 26(7), 1994.
7. David, R., and Hassane, A., Petri nets for modelling of dynamic systems - a survey, Automatica, 30(2), pp 175-202, 1994
8. Shah, J.J., and Mäntylä, M., *Parametric and Feature-based CADCAM: Concepts, Techniques and Applications*, New York: John Wiley & Sons, 1995.
9. Dicesare, F., Harhalakis, G, Proth, J.M., Silva, M., and Vernadt, F.B., *Practice of Petri Nets in Manufacturing*, London: Chapman & Hall, 1993.

SEQUENCING FOR PROCESS PLANNING BY GENETIC ALGORITHM

P. Dépincé [1], N. Lançon [1] and J. Vancza [2]
[1] Institut de Recherche en Communication et Cybernétique de Nantes, France
[2] Laboratory of Engineering and Management Intelligence, Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungar

Abstract: In this paper, we present an application of Genetic Algorithm (GA) in Computer Aided Process Planning (CAPP). Sequencing of machining operations is an important function in process planning elaboration. It conditions the choice of machines, the feasibility and the cost of the process plan. We suggest a GA that allows the handling of conditional precedence constraints, the representation of different process planning solutions and the discovery of alternative optimized solutions. The formulation of planning problem covers resources like machine-tools, set-ups and tools. We assumed that resource information is provided either automatically or interactively. Sequencing is performed using ordinal representation and Tool-Set-up-Machine's triplets using a domain specific integer coding schema. The influence of the different parameters such as the population size, the probabilities of mutation and crossover, local search methods on the evolution of the population are presented.

Keywords: CAPP, Genetic Algorithm, Sequencing, Manufacturing

1. INTRODUCTION

One of the missing links between CAD and CAM is the virtual absence of any systematic methodology for generating and evaluating alternative ways to manufacture an intended design. Most integrated CAD/CAM systems try to generate a single process plan for a given design. The task of manufacturing process planning is to determined an ordered set of operations

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 33–42.

that, at the end of the manufacturing process, leads to the desired part: one can say that it is the systematic determination of the methods which allows a product to be manufactured economically and competitively [1].

Over the last three decades, many researchers have addressed the area of computer aided process planning (CAPP), computers have been implemented widely in manufacturing industry and have become more and more powerful. However even if both design department and workshop have received help from this technology through CAD/CAM systems or NC machines, the manufacturing process is still not fully automated and the benefits within the industrial environment are not as satisfactory as expected. The process planning problem is difficult to handle in totality because it is knowledge intensive, highly dependent on the background of user and workshop, the information is often imprecise and vague and process planning problems have alternative solutions.

Many different activities take place in process planning among which one can distinguish: CAD model specification and interpretation, possible process determination, resource selection (machines, tools, fixtures, …), operation sequencing, determination of tool paths and cutting conditions and choice of the best process planning model. In manual process planning all these tasks are done by the human process planner, in an arbitrary order. The resulting plan is highly dependent on individual skills, human memory and the workshop's background. So different process planners will propose different process plans for the same part. Moreover the same planner can propose different process plans for the same product: a process planning problem has alternative solutions. The need to speed up the generation time of process plans has led to the development of CAPP systems. Research has been conducted in order to formalize and systemize process planning. The first research handled specific problems such as cutting parameters and tool life. Now, as automated planning functions have been introduced, traditionally two types of automated process planning are recognized: variant type and generative type [2]. The emergence of Artificial Intelligence tools such as neural networks, fuzzy logic, genetic algorithms, etc, gives us new opportunities to solve the complex problem of an automated elaboration of process plans. The most used methods in CAPP researches are fuzzy logic or mixed methods using neural networks, fuzzy logic and expert systems [3]. One can find in [3] and [4] good summaries of AI applications in CAPP. Actually AI techniques are used for specific functions such as cutting tool selection, machine operation sequence, pattern recognition, ….

The use of AI techniques in process planning has highlighted the need for the right tool with respect to the problem domain. Some AI techniques such as genetic algorithms or fuzzy logic can deal with alternative manufacturing routes and produce non linear process plans.

In this paper we deal with the problem of operation sequencing and we present a method for generating optimized sequence of process planning operations based on a genetic algorithm. A population of feasible solutions is generated and bred using reproduction, crossover and mutations operators. We are interested in the best solution in terms of cost that emerge after breeding the population over a number of generations. Before the presentation of the GA implementation of the sequencing, we will discuss the nature of the planning problem. Then we present the influence of the GA parameters on the results according to the statistical study of the program's performances for different parameters and parts.

2. PROCESS PLANNING MODELS

In this section we describe the planning problem and more precisely how we deal with the sequencing.

2.1 Part model

A part is described through its machining features [5]. We have to identify the set of all potential machining features which can be used to create the part. In some cases, some similar features can be grouped: for example, three identical holes in the same orientation. This also allows us to group some features in one block for which the machining operations are already known.

2.2 Resources models

The resource models are composed on the one hand with the machines and tools which are available to manufacture a part and, on the other hand with the description of the set-up which allows the machining. The description of the tools, machines and set-ups that fit a part is represented by an integer. To simplify the problem, we have taken into account only conventional milling machines, NC milling machines (2,5 axes) and conventional lathes and NC lathes (2 axes).

Given a part, for each feature we have to specify all combinations of tool, machine and set-up that can be used to machine this feature. Hence, for each operation, we have defined triplets composed of tool, set-up and machine: called TSM triplets.

2.3 Ordering

The sequencing of the operations must satisfy different constraints. The constraints represent either technical constraints (machine-tool and tool capabilities, …) or technological constraints (expert rules and knowledge, tolerance). We presume that technical constraints are taken into account in the determination of the TSM triplets. Technological constraints as precedence constraints between operations are described in a precedence matrix, PM (Figure 1), [6].

	1	2	3	4	5	6	7	8	9	10	A	B	C
1											0		
2											0		
3	1	1									2	0	
4											0		
5				1							1	0	
6											0		
7											0		
8											0		
9							1	1			2	0	
10							2	2	2		3	1	1
											1,2,4,6,7,8	3,5,9	10

Figure -1. Matrix of precedence constraints of the "Socle"

PM(i,j) = 1 or 2 means that operation i has to be machined after operation j,

We compute the sum for each row (column A). All the operations which have a zero coefficient in column A can be machined first. Then we reiterate the process for the remaining operations (column B) and so on. This representation allows us to take into account conditionals constraints (coefficient (i,j) are set differently at 0 or1).

2.4 The cost evaluation

The optimisation criteria in CAPP are complex and we can define several cost (total machining cost, machining time, …) which will bring about different consequences in the final process plan. Some researchers use a criteria based on a global cost ([7], [8]). In [5], the criteria is the costs caused by machine tools and/or set-up changes.

Our cost evaluation of each plan is composed of three costs:

1. the cost linked directly to the machines (hours cost),
2. the cost of resource changes: generated by the preparation of the first tool, set-up and machine and their modification during the machining,
3. the cost generated by the operations' machining (hourly cost of the tool, machining time, …).

3. GENETIC ALGORITHMS FOR PROCESS PLANNING

The process-planning problem is a combinatory problem. This means that the number of possible plans increases proportionally to the factorial of the number of features. Genetic Algorithm (GA) is a good optimizing tool for such combinatory problems in which we try to obtain a solution as close to the optimum as possible. We are going to describe how we solved the above stated CAPP problem by the application of GA. We will define the code used, which is composed of two parts: the information on the sequencing and the resources assignment of each operation, then the genetic algorithm process. An overview of GA can be found in [9].

3.1 The coding of the genes

The individuals represent process plans which are defined as a sequence of the different operations. Each plan is complete, i.e. it contains all the operations needed to machine the part. The first operation will be described first and so on. The place of an operation in the process plan is coded in the individual with the other information.

3.1.1 The coding of the tools, set-ups and machines

For coding the assignment of the resources, a code using integers turned out to be more convenient than a binary code. Hence we have decided to choose an integer code for each tool, machine and set-up:

- tools are described with their name, cost and diameter, number of teeth and lengths if required,
- set-ups are described with their name,
- machines are described with their name, power, hourly cost and type.

3.1.2 The coding of the triplets Tool-Set-up-Machine (TSM)

Up to now, an individual would be composed of a succession of genes including the code for the place of operations in the sequence and 3 integers per operation coding the tool, set-up, and machine. To prevent the appearance of wrong individuals during the processes of mutation, crossover or even creation of the first generation, we have used the TSM triplets. They are coded with an integer.

3.1.3 The coding of the operation's sequence

The sequencing of the operations is one of the principal problems of process planning. The representation has to be independent of the GA processes (mutation, crossover and even creation of the first generation): without special precautions, the created sequence can easily be wrong (some operations can be duplicated or missing). Correction of such mistakes leads to a loss of time. An example of this problem is given (Table 1): when using a classical crossover with an integer representation, the new generation is composed of wrong individuals with missing and duplicated operations.

Table -1. Mistakes generated with the classical crossover and integer representation

<table>
<tr><th>Parents:</th><th>offspring</th></tr>
<tr><td>1 4 | 3 2</td><td>1 4 | 1 3</td></tr>
<tr><td>4 2 | 1 3</td><td>4 2 | 3 2</td></tr>
</table>

The "ordinal representation" proposed by [10] for the travelling salesman problem might be a solution in our case. We present this representation by the following example: consider a sequence 1 2 4 3 8 5 9 6 7 where the sequence represent the cities that are visited. This sequence in the ordinal representation is coded with 1 1 2 1 4 1 3 1 1. The first 1 means that the first city is the first city in the list of the available cities (1 2 3 4 5 6 7 8 9), so our journey begins with city number 1. Then the first city is removed from the list which becomes 2 3 4 5 6 7 8 9. The second 1 means the same, we have to go to the first city in our list, which is city number 2. And so on, till we have listed all the cities. This representation prevents the creation of wrong individuals.

For our convenience, we have kept in the program two representations for each individual in the population. The main one describes the population with ordinal representation and the TSM integers: it is used for the operation of crossover and mutation and can be called the genotype. The second represents the real process plans with the operations and the tools, machine and set-up assigned for each operation. It is used to calculate the cost and the final evaluation: it represents the phenotype.

3.2 The genetic process

The processes involved in a genetic algorithm are composed of: initial generation, selection mechanism, crossover, mutation, fitness evaluation and environmental selection [11]. In the following paragraph we give the advantages of our propositions.

3.2.1 The Generation of the initial population

The initial population is created randomly. The operations sequence is created using the ordinal representation. The TSM are chosen randomly in the set of possible triplets described by the user for each operation.

3.2.2 The selection mechanism

In GA, individuals undergo genetic operations (like crossover and mutation). The selection of the individuals for those processes is proportional to their fitness. This selection strategy is implemented by a wheel. In this approach, every individual has a space on the wheel whose size is proportional to its evaluation, or in other words, its fitness. To know which individual is selected, we have to turn the wheel a number of times equal to the number of individuals we want to generate.

3.2.3 Crossover

The main advantage of the ordinal representation and the used of TSM triplets is the simplicity of the crossover process. The crossover process creates new individuals, mixing the genes of the parents. It is applied to the individuals according to the probability of the crossover and the wheel mechanism. The new individuals are added to the population. One TSM will be replaced by one other during the crossover and thanks to the general representation this new TSM will also fit this operation.

3.2.4 Mutation and local search

Mutation is a random local change in the genetic code of the individuals. It is executed after the birth of new individuals according to the mutation probability and the wheel mechanism. The classical mutation process is replaced by a local optimisation process which searches for local improvements. Next to the selection of a gene, the program searches in the set of possible TSM for one which uses the same resources as the preceding operation. If such a TSM exists, it is chosen and as there are less tool, set-up and machine changes, the cost decreases.

These processes do not increase the size of the population, the new individuals replace the old ones. However, the best-ever individual is kept in the new population.

3.2.5 Fitness evaluation and environmental selection

The fitness evaluation includes the cost criteria, we use a constant cost for every violated precedence constraints. All the individuals of the population are ordered and from one generation only the best individuals are kept in the population.

4. RESULTS

All the processes described in the preceding paragraph have been implemented. Some tests have been carried out in order to determine the effects of the random function used in the program, the history of the cost function and the fitness of the population, the influence of the different parameters (probabilities of mutation and crossover and the population size) and the effect of the local search.

4.1 Convergence

4.1.1 Analysis of random factors

For a given problem, we launched the program with the same parameters 200 times and show the variation of the resolution speed.
According to the 5 parts we have tested, the convergence speed is fast in most of the cases: the best solutions are generally found before 1000 generations. However, the problem is not solved before 5000 generations in 12% of cases. One can note that the number of generations can be increased and/or the program can be launched more than once because the execution time is short (nearly 60 sec). We have to take care that the results we obtain are not too specific. So for the study of the influences of the parameters, we use at least the average result obtained through 50 executions.

4.1.2 Influence of the probabilities of mutation and crossover

We naturally looked for a fast convergence towards the optimum plans. Different probabilities of mutation and crossover have been tried in order to determine the best pair. There are few "good" parameters: the probability of crossover should be high and the probability of mutation should be close to 10%. Those parameters are quite good for all our 5 examples.

4.1.3 Influence of the population size

Up to now, we have considered the number of generation as our criterion for the study of the convergence speed because we had decided to keep the population equal to 50 individuals. However the real criterion should be the time it really takes to compute the solution. We have tried to calculate this time because it was not technically possible to measure it. The time needed to reach the solution would be proportional to the size of the population multiplied by the number of generation needed. Increasing the size of the population decreases the number of generations needed to find the solution, even if the real time increases.

4.1.4 Influence of the local search

The aim of the local search is to improve the solution by checking the opportunity to use the same machine, set-up and tool for two consecutive operations. This process is used in alternation with that of the mutation process. We have compared the results obtained with local search or without it. We can see that the local search is useless when the crossover probability is low. When this probability increases, the local search gives better results.

4.1.5 History of the population

It is also interesting sometimes to follow only one population. For example we followed the evolution of one population through the generations. The first improvements of the population come from the suppression of the individuals which do not respect the precedence constraints. So the whole population improves its fitness and then the solution can be constructed from the relatively good solutions in the population. According to this experience and those dealing with the other parts, we can conclude that the constraints in the final solutions are in all cases satisfied (hence all the solutions were executable and the plans were legal) and the cost decreases during the GA process.

5. CONCLUSION AND PERSPECTIVES

We have presented in this paper a method based on genetic algorithms to sequence the operations in process planning. The results we have obtained are promising even if some drawbacks still exist. All individuals created during GA processes are correct thanks to the use of TSM and ordinal representations. In order to take into account precedence constraints we use a matrix description. We have proved that the convergence speed can be improved using a local search algorithm during the mutation process. One of the main problems is the relatively weak population average fitness. If the whole population were fitter, it would greatly increase the convergence speed. Future approaches could consider the modification of mutation and crossover processes in order to improve this convergence.

REFERENCES

1. L. ALTING, H. ZHANG, "*Computer Aided Process Planning: the state of the art survey*", Int. J. Pro. Res., vol 27/4,1989, pp. 553-585.
2. L. SABOURIN, "*L'expertise en conception de gammes d'usinage: approche par entités et propagation de contraintes*", Ph.D Thesis, Laboratoire Universitaire de Recherche en Production Automatisée, ENS Cachan, France, 1995, 242 pages.
3. R. TETI, S.R.T. KUMARA, "*Intelligent computing methods for manufacturing systems*", Annals of CIRP, Vol. 46/ 2, 1997, pp. 629-652.
4. C. L. HORRIS, "*Annotated bibliography on CAPP*", Int. J. of Advanced Manufacturing Technology, Vol. 12, 1996, 309-329.
5. J. VANCZA et A. MARKUS « *Experiments with the integration of reasoning, optimisation and generalisation in process planning* », Advances in Engineering Software, Vol. 25, pp. 29-39, 1996.
6. G. HALEVI, R.D. WEILL « *Principales of Process Planning A logical appr*oach », édition Chapman & Hall, 1995.
7. M.S. SHUNMUGAN, S.V. BHASKRA REDDY et T.T. NARENDRAN « *Selection of optimal conditions in multi-pass face-milling using a genetic algorithm* », International Journal of Machine Tools & Manufacture 40 (2000) pp 401-414, 1999.
8. J. G. TAIBER « *Optimisation of process sequences considering prismatic work-pieces* », Advances in Engineering Software, Vol 25 (1996), pp. 41-50, 1996.
9. S. R. LADD « *Genetic Algorithms in C++* », édition M&T BOOKS, 1996.
10. Z. MICHALEWICZ « *ARTIFICIAL INTELLIGENCE : Genetic Algorithms + Data Structures = Evolution Programs* », édition Springer-Verlag, 1992.
11. B. AWADH, N. SEPEHRI et O. HAWLESHKA « *A computer-aided process planning model based on genetic algorithms* », Computers Ops Res., Vol 22, No. 8, pp. 841-856, 1995.

A TOOL FOR VERIFYING THE LOCAL QUALITY OF A 3D F.E. ANALYSIS IN DESIGN

E. Florentin, L. Gallimard and J-P. Pelle
LMT-Cachan, ENS de Cachan, CNRS, Université Paris 6, France

Abstract: In many practical situations, it is necessary to have reliable evaluations of local quantities. For example, this is the case in design where dimensioning criteria nearly always involve local quantities (stresses in specified zones, Von Mises' stresses, displacements, stress intensity factors, ...). In current industrial practice, these quantities are evaluated using F.E. analysis. Even if the mechanical model chosen is adequate (good description of the geometry, good knowledge of the material's characteristics and of the loading), the F.E. analysis introduces errors in the quantities being calculated. For the engineer, it is essential to study and, if possible, to evaluate the quality of the calculations carried out in order to validate the results. In this work, we are concerned with the quality of a linear F.E. analysis. We propose a tool, based on the concept of error in constitutive relation, which enables one to evaluate the local quality of the stresses. We show that this approach gives good results.

Key words: F.E., local error, quality, dimensioning

1. INTRODUCTION

The study presented here constitutes a new evolution in the estimation of discretization errors in finite element analysis which will finally meet the needs of the engineering community. In many practical situations, reliable estimates of local quantities are required. This, for example, is the case in design, where dimensioning criteria nearly always involve local quantities:

– stress in the most highly-loaded zones,
– Von Mises' equivalent stress in these same zones,
– displacement in a specific part of the structure,

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 43–52.

– stress intensity factors in the vicinity of a cracked zone.

In current industrial practice, these quantities are evaluated on the basis of finite element calculations. Even if the mechanical model chosen is appropriate - proper description of the geometry, good knowledge of the materials' characteristics and of the loading cases - the resulting values are impaired by a certain discretization error. It is essential for the engineer to study and, if possible, to evaluate the quality of the calculations performed in order to validate the results obtained. This is a crucial prerequisite in order to make consistent and sure decisions. The problems which will be discussed here are linear elasticity problems, which are representative of the great majority of 3D finite element analyses performed in an industrial context. First, we will aim to propose a method to evaluate the quality of the local stresses, which constitutes the first dimensioning tool. This method relies on the concept of error in constitutive relation [Ladevèze et al., 1999] and on the construction of an associated admissible field [Ladevèze and Rougeot, 1997]. This method, which is specific of stresses, is a direct method. Other works on this topic [Peraire and Patera, 1998,Prudhomme and Oden, 1999, Rannacher and Suttmeier, 1997, Ladevèze et al., 1999, Strouboulis et al., 2000], which extend the quality evaluation beyond stresses to many other types of quantities (displacements, ...), require a second finite element calculation (extractor calculation). Here, the objective is to provide the engineer, for a zone considered to be critical:

– a reference stress in each element,
– a precise estimate of the quality of this reference stress with respect to the exact stress.

This reference value can initially be the energy norm of the stress over an element. Next, we will examine, with the same goal, the evaluation of the quality of Von Mises' equivalent stress, which is frequently used to test elasticity limit criteria.

2. ERROR IN CONSTITUTIVE RELATION

2.1 The reference problem

Let us consider a structure Ω bounded by $\partial\Omega$ and subjected to:

– a prescribed displacement on a part $\partial_1\Omega$ of the boundary: $\underline{U}_d$
– a surface density of force defined on the part $\partial_2\Omega = \partial\Omega - \partial_1\Omega$ of the boundary: $\underline{F}_d$
– a volume density of internal force given in Ω : $\underline{f}_d$

The problem which enables one to determine the solution over the whole structure can be written as follows:

Find $\underline{U}(M)$ and $\sigma(M)$ defined on Ω which verify:

- the kinematic constraints (KA): $\underline{U}_{|\partial_1\Omega} = \underline{U}_d$
- the equation of equilibrium with external forces (SA):
  $\underline{div}\ \sigma + \underline{f}_d = 0$ in Ω and $\sigma\ \underline{n} = \underline{F}_d$ over $\partial_2\Omega$
- the constitutive relation (CR): $\sigma = \mathbf{K}\ \varepsilon(\underline{U})$

where $\mathbf{K}$ designate the material Hooke's operator.

This problem is solved approximately using the finite element displacement method. This method leads to the pair $(\underline{U}_h, \sigma_h)$, which verifies the constitutive relation and the equilibrium equations exactly, whereas the kinematic constraints equations are verified approximately.

The concept of error in constitutive relation relies on theconstruction of a solution to the reference problem, called the admissible solution, starting from the finite element solution and the problem's data. This solution, designated by $s_{Ad} = (\underline{U}_{KA}, \sigma_{SA})$, verifies the kinematic constraints equations and the equilibrium equations. However, it does not verify the constitutive relation. The quality of this approximate solution is quantified by calculating a residual of the non-verification of the constitutive relation designated by e_h (which, through Hooke's operator, connects a stress σ to the strain $\varepsilon(\underline{U})$ associated with the displacement $\underline{U}$):

$$e_h^2 = \sum_{E \subset \Omega} e_{h,E}^2 \tag{1}$$

with $e_{h,E}^2 = \| \sigma_{SA} - \mathbf{K}\ \varepsilon(\underline{U}_{KA}) \|_E^2$ introducing $\| \cdot \|_E$, the energy norm over element E.

The critical point is the construction of s_{Ad} from the finite element solution $(\underline{U}_h, \sigma_h)$. For the displacement, we simply choose $\underline{U}_{KA} = \underline{U}_h$. For the stress σ_{SA}, we use a method described in [Ladevèze and Rougeot, 1997].

2.2 Effectivity index

One means of quantifying the quality of an estimator is to compare the result of the estimation to the true error. This ratio is called the effectivity index. Thus, one can define a global effectivity over the whole domain Ω:

$$\zeta = \frac{e_h}{\sqrt{\int_\Omega Tr\left| (\sigma_{ex} - \sigma_h)\ \mathbf{K}^{-1}\ (\sigma_{ex} - \sigma_h) \right| d\Omega}} \tag{2}$$

A good global estimate of the error corresponds to $\zeta = 1$. The further the effectivity index from 1, the worse the estimate.

The estimator in constitutive relation guarantees an effectivity index greater than 1, which means that it leads to an estimate of the error which is an upperbound of the actual error.

To quantify the quality of the local contributions to the global error, one can also define, element by element, a local effectivity index ζ_E:

$$\zeta_E = \frac{e_{h,E}}{\sqrt{\int_E Tr\left[(\sigma_{ex} - \sigma_h)\, \mathbf{K}^{-1}\, (\sigma_{ex} - \sigma_h)\right] dE}} \tag{3}$$

Then, σ_{ex} can be obtained either analytically or numerically (one would then seek an estimate of this exact solution, called quasi-exact solution, through refinement of the initial mesh).

2.3 Heuristic property

Using the Prager-Synge theorem [Prager and Synge, 1947], one shows easily that:

$$\left\| \sigma_{ex} - \sigma_h \right\|_\Omega \leq \left\| \sigma_{SA} - \sigma_h \right\|_\Omega = e_h \tag{4}$$

A priori, it is not possible to prove, on the local level, an inequality of the type (4):

$$\left\| \sigma_{ex} - \sigma_h \right\|_E \leq \left\| \sigma_{SA} - \sigma_h \right\|_E = e_{h,E} \tag{5}$$

Nevertheless, if one generates the field σ_{SA} using the method introduced in [Ladevèze and Rougeot, 1997], one observes experimentally that:

$$\left\| \sigma_{ex} - \sigma_h \right\|_E \leq C \left\| \sigma_{SA} - \sigma_h \right\|_E \tag{6}$$

where C is a constant which is numerically close to 1.

3. AN EXAMPLE OF APPLICATION

The property discussed above was verified on all numerical examples considered. The following example is presented as an illustration.

This example concerns a clamping plate, which can be used to fasten a workpiece during machining. The plate is built-in along the lower surface to represent the connection with the machine tool. It is also subjected to a pressure load on the perpendicular face to represent the cutting load. The mesh comprises 2,828 TET4 elements. The estimated global error is 41% while the actual global error is 36%.

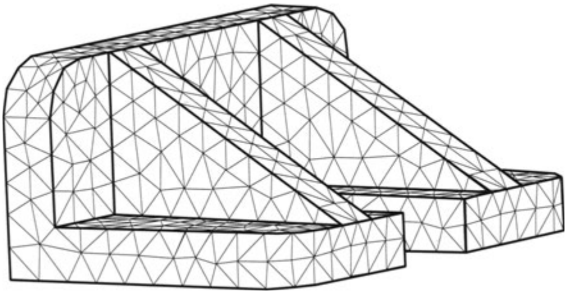


Figure –1. Mesh of the clamping plate

Now, one can plot the local effectivity index of each element. The local effectivity indexes are good because they are close to 1.

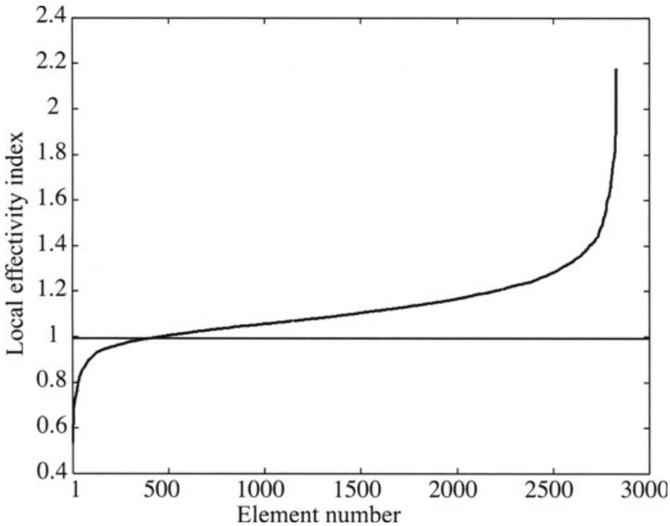


Figure -2. Local effectivity indexes in increasing order

In the zones of interest for the estimates, which are greater than the actual error:

- in the most highly loaded zone, which corresponds to the reinforcements:
  actual error: 7.79 %
  estimate: 8.62 %
  effectivity index: 1.11

- in the zone close to the built-in boundary condition:
  actual error: 27.57 %
  estimate: 29.70 %
  effectivity index: 1.08

4. ESTIMATE OF THE ERROR MADE ON DIMENSIONING QUANTITIES

Now we have a precise idea of the local quality of the finite element solution. This error concerns the energy norm of the stress. In practice, a parameter which is often used is Von Mises' equivalent stress.

Here, we propose to use the constructed admissible fields $s_{Ad} = (\underline{u}_{KA}, \sigma_{SA})$ to evaluate the local quality.

4.1 Relation between the error on Von Mises' stress and the energy norm

Let us designate by $J^E_{VM}(\sigma)$ the average Von Mises' stress on an element E:

$$J^E_{VM}(\sigma) = \frac{1}{mes(E)} \int_E \sqrt{\frac{3}{2} Tr[\sigma_D \sigma_D]} \; dE \tag{7}$$

where: $\sigma_D = \sigma - \frac{1}{3} Tr(\sigma) \mathrm{I}_d$

Thus:

$$\begin{aligned} \left| J^E_{VM}(\sigma_{ex}) - J^E_{VM}(\sigma_h) \right| &\leq J^E_{VM}(\sigma_{ex} - \sigma_h) \\ &\leq \frac{1}{mes(E)} \int_E \sqrt{\frac{3}{2} Tr[(\sigma_{ex} - \sigma_h)_D (\sigma_{ex} - \sigma_h)_D]} \; dE \\ &\leq \sqrt{\frac{3k}{2\, mes(E)}} \left\| \sigma_{ex} - \sigma_h \right\|_E \end{aligned} \tag{8}$$

$$\text{where: } k^{-1} = \operatorname*{Inf}_{\sigma \neq 0} \frac{Tr(\sigma\, \mathbf{K}^{-1} \sigma)}{Tr(\sigma\ \sigma)} \tag{9}$$

is the minimum eigenvalue of $\mathbf{K}^{-1}$.

As shown by the previous inequation, the energy norm is an upperbound of the quantity of interest. This shows that if the error in the energy norm is small, so is the local Von Mises' error.

One can obviously deduce from this the rigorous estimate:

$$\left| J^E_{VM}(\sigma_{ex}) - J^E_{VM}(\sigma_h) \right| \leq \sqrt{\frac{3\,k}{2\ mes(E)}}\ \left\| \sigma_{SA} - \sigma_h \right\|_\Omega \tag{10}$$

Of course, this upperbound is of little practical interest numerically.

4.2 Estimate retained to evaluate Von Mises' error

Using (8) and (6), we have:

$$\left| J^E_{VM}(\sigma_{ex}) - J^E_{VM}(\sigma_h) \right| \leq C \sqrt{\frac{3\,k}{2\ mes(E)}}\ \left\| \sigma_{SA} - \sigma_h \right\|_E \tag{11}$$

with C close to 1.

Therefore, in practice, one can estimate the error on Von Mises' stress using:

$$e^E_{VM} \leq \sqrt{\frac{3\,k}{2\ mes(E)}}\ \left\| \sigma_{SA} - \sigma_h \right\|_E \tag{12}$$

Thus, Von Mises' stress can be overestimated significantly, particularly if stresses σ_{ex} and σ_h are spherical.

One can also introduce the following estimate:

$$\tilde{e}^E_{VM} \leq \left| J^E_{VM}(\sigma_{SA}) - J^E_{VM}(\sigma_h) \right| \tag{13}$$

which consists simply of replacing in (8) the exact stress σ_{ex} by the statically admissible stress σ_{SA} constructed in order to obtain the error in constitutive relation.

4.3 Quality of the estimates

The engineer who wishes, for example, to dimension a part with regard to the elastic limit wants to know the maximum value of Von Mises' stress in the structure. The error made defines an interval within which lies the exact stress.

$$J^E_{VM}(\sigma_h) - e^E_{VM} \leq J^E_{VM}(\sigma_{ex}) \leq J^E_{VM}(\sigma_h) + e^E_{VM} \tag{14}$$

Since the maximum stress is a deciding factor in dimensioning, one gets an upperbound of the quantity of interest by using $J^E_{VM}(\sigma_h) + e^E_{VM}$. Then, one can define a quality index of an upperbound uppE:

$$upp_E = \frac{J^E_{VM}(\sigma_h) + e^E_{VM}}{J^E_{VM}(\sigma_{ex})} \tag{15}$$

This quality index must be greater than 1. One can define a quality index of the upperbound uppE for the heuristic estimate:

$$\tilde{upp}_E = \frac{J^E_{VM}(\sigma_h) + \tilde{e}^E_{VM}}{J^E_{VM}(\sigma_{ex})} \tag{16}$$

4.4 Example

On the clamping plate, we plot upp_E and $\tilde{upp}_E$ for all the elements in increasing order of Von Mises' stress. On this example, we notice that the two estimates e^E_{VM} and $\tilde{e}^E_{VM}$ are, indeed, upperbounds of the true values. In both cases, we get a good-quality estimate of the exact Von Mises' stress.

The quality is better for the second estimate $\tilde{e}^E_{VM}$, particularly in the zones where the Von Mises' stress is high, i.e. the dimensioning zones. Indeed, if we look at the 20% most highly stressed elements in Von Mises' sense, which corresponds to Zone A on Fig. 3, we observe that the quality indexes are close to 1.

In the most highly loaded zone corresponding to the reinforcements:

$J^E_{VM}(\sigma_{ex}) = 260\ MPa$

$J^E_{VM}(\sigma_h) + \tilde{e}^E_{VM} = 265\ MPa$ corresponding to $\tilde{upp}_E = 1.02$

$J^E_{VM}(\sigma_h) + e^E_{VM} = 303\ MPa$ corresponding to $upp_E = 1.17$

In the zone close to the clamped boundary condition:

$J^E_{VM}(\sigma_{ex}) = 190\ MPa$

$J^E_{VM}(\sigma_h) + \tilde{e}^E_{VM} = 236\ MPa$ corresponding to $\widetilde{upp}_E = 1.23$

$J^E_{VM}(\sigma_h) + e^E_{VM} = 360\ MPa$ corresponding to $upp_E = 1.89$

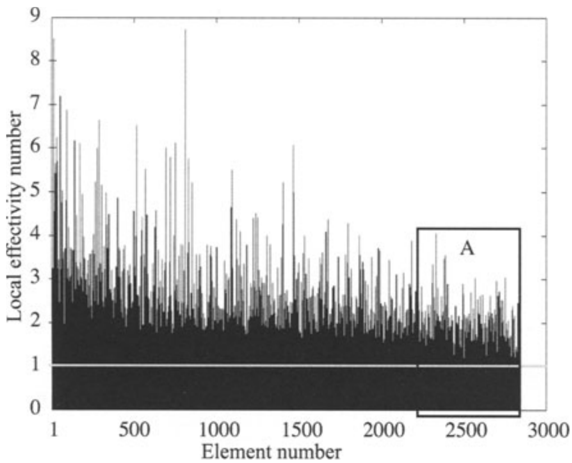


Figure -3. upp_E in increasing order of Von Mises'stress

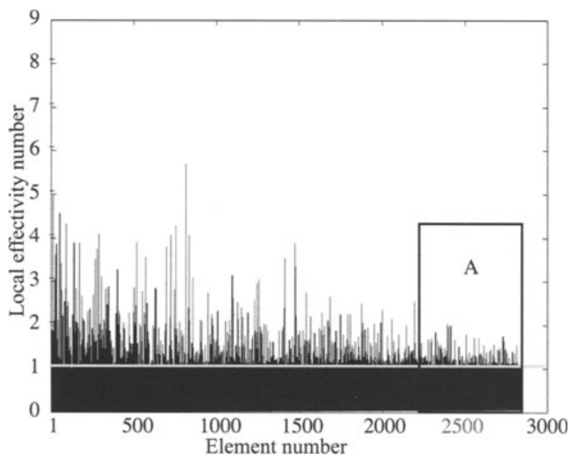


Figure –4. $\widetilde{upp}_E$ in increasing order of Von Mises'stress

Therefore, the designer has two simple quality control tools at his disposal. Very similar results were found on other examples.

5. CONCLUSIONS

The error in constitutive relation provides good local estimates on the stress in terms of energy. The study performed on the Von Mises' stress shows that it is easy to derive realistic estimates of the error made on this quantity in a finite element analysis and, therefore, that this constitutes a usable design tool.

ACKNOWLEDGEMENTS

This work is being sponsored by the French Ministry of National Education in the framework of the RNTL (Network of New Software Technologies). This particular project, "Fiabilité", is a cooperation among LMT-Cachan, EADS, SAMTECH-France and the CEA.

REFERENCES

[Ladevèze and Rougeot, 1997] Ladevèze, P. and Rougeot, P. (1997), New advances on a posteriori error on constitutive relation in f.e. analysis, *Comp. Meth. in Applied Mech. and Engrg.*, 150: 239-249

[Ladevèze et al., 1999] Ladevèze, P., Rougeot, P., Blanchard, P., and Moreau, J. (1999), Local error estimators for finite element analysis, *Comp. Meth. in Applied Mech. and Engrg.*, 176: 231-246

[Peraire and Patera, 1998] Peraire, J. and Patera, A. (1998), Bounds for linear-functional outputs of coercive partial differential equations: local indicators and adaptive refinement. in *Advances in Adaptive Computational Methods*, Ladevèze, P. and Oden, J., editors: 199-216, Elsevier

[Prager and Synge, 1947] Prager, W. and Synge, J. (1947), Approximation in elasticity based on the concept of functions space, *Quart. Appl. Math.*, 5: 261-269

[Prudhomme and Oden, 1999] Prudhomme, S. and Oden, J. (1999), On goal-oriented error estimation for elliptic problems : application to the control of pointwise errors, *Comp. Meth. in Applied Mech. and Engrg.*, 176: 313-331

[Rannacher and Suttmeier, 1997] Rannacher, R. and Suttmeier, F. (1997), A feedback approach to error control in finite element methods: Application to linear elasticity, *Comp. Mech.*, 19: 434-446

[Strouboulis et al., 2000] Strouboulis, T., Babuka, I., Datta, D., Copps, K., and Gangaraj, S. (2000), A posteriori estimation and adaptative control of the error in the quantity of interest. part 1: A posteriori estimation of the error in the von mises stress and the stress intensity factor, *Comp. Meth. in Applied Mech. and Engrg.*, 180: 261-274

DESIGN OF WIRE HARNESSES FOR MASS CUSTOMIZATION

B. Agard and M. Tollenaere
GILCO, Grenoble, France

Abstract: The paper focuses on the design of wire harness assemblies for mass customization by a delayed product differentiation. In order to manufacture broadly diversified products, two algorithms are proposed both using a generic representation of wire harness with all options and variants in order to produce each wire harness in a short period of time. An industrial case study is presented in a contractor/supplier context, where the supplier must respond in a short time and provide a totally diversified product, which is to be delivered according to the specifications provided by the contractor. In the particular case described above, two different algorithms are applied and compared.

Key words: Wire harness, Design of Product families, Product assembly.

1. INTRODUCTION

Wire harnesses are a set of electric cables that are used to connect different elements in electromechanical or electronic systems. The functions of a wire harness are to provide electric power and electronic signals to the different peripheral units. An example of a wire harness in an automobile context is provided in Figure 1.

A wire harness is composed of different kinds of elements:

- A set of cables that are used in order to transmit information and energy.
- Connectors are required to plug the wire harness with the other elements.
- Epissures are soldered joints between cables.
- Derivations are places on the wire harness where some cables change direction.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 53–62.

- Shafts are sometimes installed on zones of the wire harness when it is necessary to resist certain constraints such as vibrations, shocks, friction, waterproofs, …
- Clips are in different places on the wire harness to fix the wire harness on to the final product.

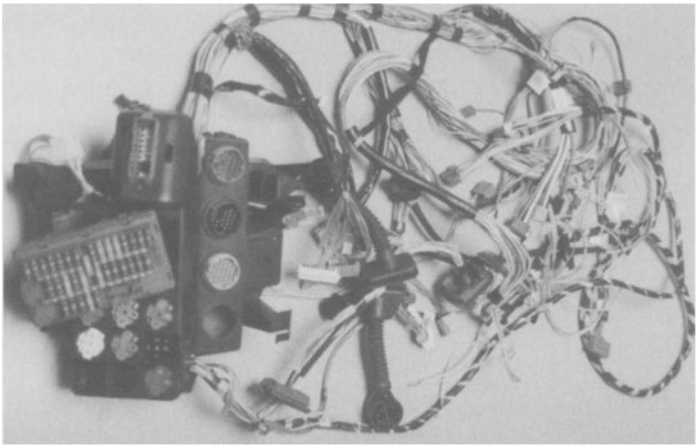


Figure -1. An automobile wire harness

All these elements are here to answer a lot of individual functions. In a medium wire harness in an automobile context, the family can be made of 400 references of cable, 120 Connectors, 50 Derivations and 30 Episures, in order to realize approximately 15 different functions with a maximum of 9 versions for several of these functions.

A wire harness is a component that is rather difficult to manufacture with economic constraints. Complexity comes from diversity; many components of the final product need to be connected to the wire harness in order to receive energy and/or information. Nevertheless, some components could be optional, many of them have different variants, and moreover, these components may evolve towards different versions. Usually different variants and different versions of a component do not have the same requirements with the wire harness. Each time these requirements change (intensity of current, type of connector, number of cables) the wire harnesses have to be adapted.

Let's consider a vehicle for which there are 6 versions of transmission, 7 versions of engine cooler and 9 versions of engine, moreover, there are 5 versions of ABS on option and 2 versions of cruise control on option. For that particular vehicle, one should be able to produce 6*7*9*6*3=972 different combinations of wire harnesses!

Actually the number of wire harnesses to produce is lower than that because some constraints between the different components do exist. For example one can install only one version of engine in each vehicle, there is no ABS on small vehicles hence the existence of exclusive and inclusive constraints such as *If "sunroof" then no "roof antenna"*. Difficulties also arise from manufacturing; the contractor wants the exact wire harness (without unnecessary components that he refuses to pay for) and he wants it to be delivered in a specific order in a short period of time.

To realize all these wire harnesses, the provider has to take into account both the wide diversity and the short delivery time. The supplier then decides to build a certain quantity of sub-assembled modules that he will assemble specifically for each final product during the time interval at his disposal.

These sub-assembled modules can be made anywhere, far from the final vehicle assembly line (where the labor costs are cheaper), while the final assembly will be carried out in factories that are local. Then the most important lead time for the synchronization will be the final assembly in the nearer factories. The diversity will be supported by the proper selection of the sub-assembled modules built in the distant factories; the manufacturing time of these modules then does not enter into the "time of final assembly" which is the real center of interest of this study. The contractor has an estimation of his average sales; the supplier then has an estimation of his needs in each module per period to size each buffer for the same periods.

This results in an increasingly significant number of product alternatives in order to answer diversified functional requirements. For producers, this commercial diversity must be controlled; otherwise an expensive diversification process could result [6]. Necessary commercial diversity can be durably assumed only if it is supported by a low technical diversity, which guarantees acceptable management and development costs [1].

In the context of product design, this research focuses on the description of a product design problem for wide diversity with modular components and product-delayed differentiation for the realization of wire harnesses. This paper is broken up in the following way: section 2 describes the context of design for diversity. Section 3 provides a description of the problem that research addresses and how it could be solved. Finally Section 4 sets out the results provided by the algorithms described in section 3.

2. DESIGN FOR DIVERSITY

According to the wide number of different wire harnesses to realize, interest has been oriented toward the design of products in a context of wide diversity.

To meet diversified needs, several solutions exist from standardized design that offers the possibility to satisfy a set of needs with a single product, to specific design aiming at the strict satisfaction of each need. Most industrial products are at an intermediate level between these two extremes. There are, at the same time, standard elements and specific elements that could be assembled in a specific way in the final product delivered to the consumer.

It is possible to distinguish two product design policies that make it possible to produce different products starting from standardized elements. These policies are modular design and product delayed differentiation.

Modular design enables one to realize a great number of different products using a limited number of modular components. Kusiak [4] worked on modular design with the intention of producing a large variety of products at lower cost. He uses a matrix representation to model interactions between parts and functions, then he breaks down the matrix to extract elements which are interchangeable, standardized, and independent.

Research has also provided methods to design product families. For further information, it is possible to refer to the work of Newcomb *et al.* [7] and Gonzalez *et al.* [2]. Jiao and Tseng [3] provided a methodology to develop an architecture of product families to rationalize the development of products for mass customization.

Delayed differentiation (or postponement) [5,8] consists in delaying the point of differentiation in the product or the process in order to store semi-finished products instead of finished products. The goal is to produce a maximum of standard elements and to push back to the latest moment the point at which each product is differentiated from the others and needs to be identified as such. Different techniques are available to carry out postponement:

- Standardization that consists in using components or processes that are common to a maximum number of products.
- Modular Design consists in breaking up a product into more or less independent sub-elements called modules. It is then possible to produce each module independently. The differentiation of finished products is manufactured at the assembly operation by the choice of modules and their position in the final product.
- Process restructuring concerns restructuring the operations of the manufacturing process of a product. Lee and Tang [5] present examples where the operation that causes differentiation is delayed or where the order of two operations is inverting.

3. DESIGN OF A WIRE HARNESS FAMILY

Currently the provider produces standard wire harnesses. That means he designs a limited number of wire harnesses that can be assembled in the totality of the finished products. Then he has to produce a buffer for each family of wire harness, the delivery time is then no longer a problem, and the synchronization is largely simplified for the provider. On the other hand, standardization causes an envelope effect (some functions are unused) that will have a cost. The interest of such a kind of approach depends on the balance between the additional costs related to the unused elements and the benefits coming from the reduction in diversity.

This approach could be used until now, but in order to decrease the costs, the contractor has asked the provider to make an effort. The contractor decided to pay only for the functions that are required for each end product. The provider will bear the costs of the unused components himself.

The manufacturer is then concerned with in the following problem: which modules should be produced in pre-assembly? Knowing that there is a fixed time for the final assembly (a time lower than the total manufacturing time of the component), and that the contractor requires exact components for a minimum cost.

One of the key points in the modular approach is the total product division in modules. The efficiency of the whole approach depends on the proper compilation of the modules that enable the product to be assembled within the time that the provider has available, that cover all the diversity, and that minimize the number of expensive references.

The goal of this project consists in determining these modules.

Two strategies are presented. The first one will be called "structural strategy" and will consider the family of wire harness to produce according to its physical description; the second one is called "functional strategy" and considers the wire harnesses according to the functions it has to provide.

3.1 Structural strategy

All the different wire harnesses that could be produced are modeled as a generic wire harness with options, variants and versions plus a set of rules to describe the constraints between the components.

The generic wire harness is described as a set of cables. From this point of view, a sub-assembled module, called industrial module (IM) is then a sub set of cables that will be produced at the distant manufacturing sites. The generic wire harness is modeled as a tree. In each branch the cables that pass inside are depicted (see Figure 2).

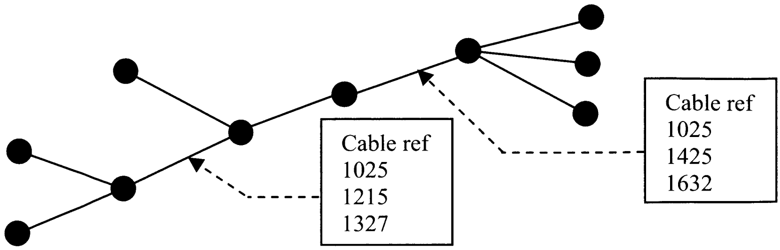


Figure -2. Structural description of a wire harness

The idea is to split the generic wire harness into two independent sub generic wire harnesses, one of them will be the IM produced at the distant manufacturing sites, the other will be the specialization of the IMs in order to create a specific wire harness for the specific end product (See Figure 3).

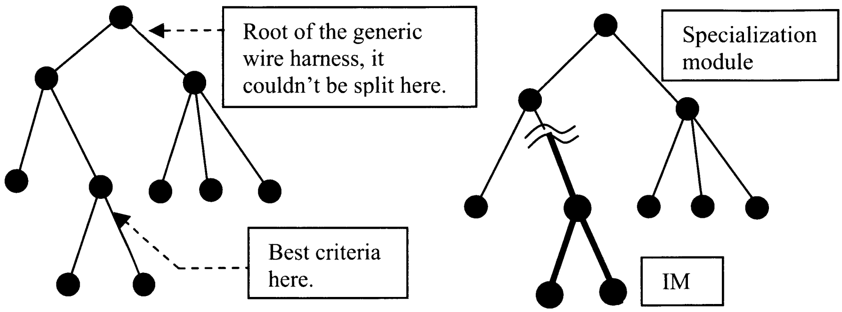


Figure -3. Split of a generic wire harness

One important criterion that one has to keep in focus is the time of specialization; this will be called "time of final assembly". That time has to be less than the time that the provider has available to produce each specific wire harness. Once the specific wire harness has been split into different modules, if the time of final assembly is greater than the limit, the specialization module will then be split another time into two modules and so on until the time of final assembly is under the limit.

In practice, the professional has to select the root of the tree by the choice of a node where the generic wire harness will not be split. Each node is evaluated from the point of view of decreasing the time of final assembly and the cost of creation of IM.

The cost of creation of IM is directly proportional to the number of modules that support all the diversity. In fact one IM is in reality several sub modules. If an IM is created for cables 1, 2 and 3, we will have to create 7 modules minus the constraints in order to produce each specific final product.

The Decreasing Time of Final Assembly (DTFA) is the time to produce the IM because they will be produced for buffers, and that time does not enter into the final assembly. These DTFA are calculated as following:

$$DTFA = \sum_{i=0}^{nb\_of\_branch} time\_to\_realize\_branch(i) + \sum_{j=0}^{nb\_of\_node} time\_to\_realize\_node(j)$$

$$time\_to\_realize\_branch(i) = type\_of\_assembly(i) \times length\_of\_branch(i)$$

$$time\_to\_realize\_node(j) = type\_of\_node(j)$$

Then a criterion is evaluated on each node; that criterion takes into account the DTFA, the cost of the references of modules and the labor cost; the best alternative is selected. The generic wire harness is then split into two parts (see Figure 3) and some branches on the specialization module are forbidden for a future split because some cables with an extremity in the IM have the other extremity in the other part.

3.2 Functional strategy

As above, all the different wire harnesses that could be produced are modeled as a generic wire harness with options, variants and versions plus a set of rules to describe the constraints between the functions. The generic wire harness is described as a set of functions.

In that strategy an IM will be a set of functions that will be realized in the distant factories.

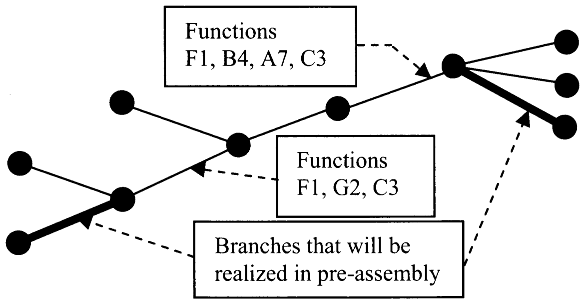


Figure -4. Functional breakdown of a wire harness

All functions are extracted from the generic wire harness; and for each set of functions that appears in one branch an evaluation of the time of final assembly that could be saved and the cost generated by the creation of IMs are calculated.

Then with the same criterion as above, the selection of the best alternative is made. The generic wire harness is then separated into two modules, and if the time of final assembly is greater than the limit, another separation is produced on the specialization module.

In the functional strategy:

$$DTFA = \sum_{i=0}^{nb\_of\_branch} time\_to\_realize\_branch(i) + \sum_{j=0}^{nb\_of\_node} time\_to\_realize\_node(j)$$

$$time\_to\_realize\_node(j) = \left| \begin{array}{l} \bullet\ type\_of\_assembly(i) \times length\_of\_branch(i) \\ \text{if all functions into branch (i) belong to the module} \\ \bullet\ \text{0 otherwise} \end{array} \right.$$

$$time\_to\_realize\_branch(i) = \left| \begin{array}{l} \bullet\ type\_of\_node(j) \\ \text{if all branch from j belong to the module} \\ \bullet\ \text{0 otherwise} \end{array} \right.$$

3.3 First results

Both algorithms have been applied on a representative wire harness. For the structural strategy the results are presented in Figure 5. The results obtained with the same wire harness and with the functional strategy are presented in Figure 6.

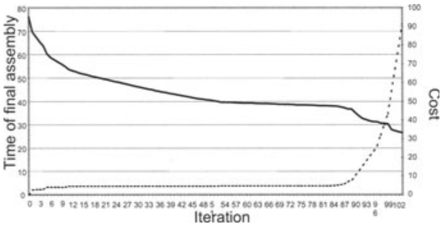


Figure -5. Evolution of the structural strategy

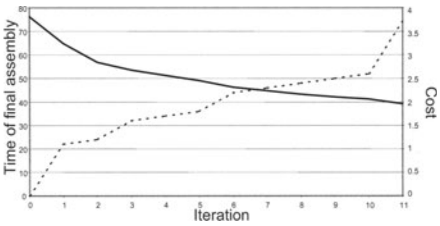


Figure -6. Evolution of the functional strategy

The full line represents the time of final assembly, the dotted line represents the costs of producing the new modules and the X-axis is the number of iterations for each algorithm.

One can observe in both cases a significant decrease of the time of final assembly linked to an increase in the cost of references to manage. Moreover, the time of final assembly could be less for the structural strategy than for the functional strategy. Also in the structural strategy the cost explodes at the end to decrease the time.

The following curves (Figure 7 and Figure 8) show how much it costs for the provider to produce a wire harness under a certain limit time with both strategies. With that kind of representation, the provider can easily sign contracts with his contractor to sell the set of wire harnesses. They can discuss the cost and time they project. Moreover, the contractor can modify his process to increase the time delay for his provider, and then decrease his contract costs.

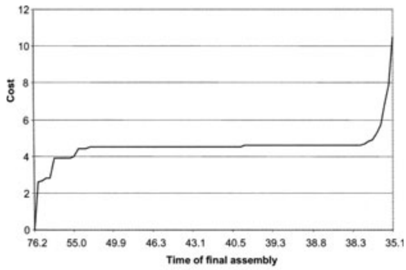


Figure -7. Average Cost and time for structural strategy

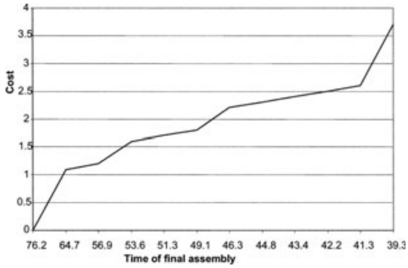


Figure -8. Average Cost and time for functional strategy

In the following representation (Figure 9), both results have been set in the same range, the parts that explode in the structural strategy have been deleted. The full line represents the structural strategy, the dotted line the functional strategy. We can note that for the high limit of time the functional strategy is always less expensive than the structural strategy. However in case the contractor wants to decrease the time of final assembly below the limit of the functional strategy, he must adopt a structural strategy.

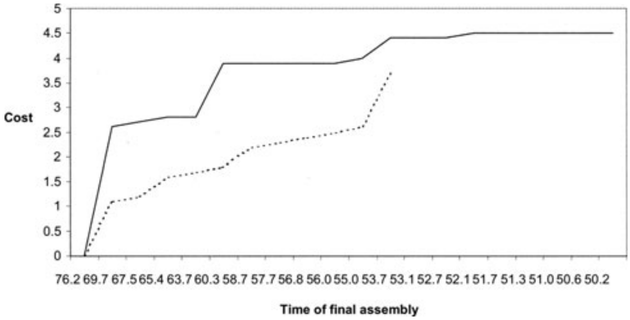


Figure -9. Comparison of both algorithms before the cost explosion

4. CONCLUSIONS AND PROSPECTS

For commercial purposes, it does not appear desirable to reduce the diversity of products perceived by customers with an aim toward marketing strategy. However we have to take into account the fact that the explosion of product variety has a cost and that for the company there is an optimal internal diversity which minimizes costs. We provide a decision-making tool that may help designers in their choices regarding harness subassembly design in order to cover all the commercial diversity.

Two algorithms have been presented that enable the production of subassemblies for a wire harness family. The algorithms use a physical or a functional description of the wire harness family and provide a set of modules to produce in order to decrease the time of final assembly, knowing that it must provide all the diversity in a short time and for a minimum cost. The results allow both the provider and contractor to discuss these relationships; they can adjust the time of synchronization and the number of versions to be produces.

This research leaves open issues: the first development to be considered is an evaluation of the envelope effect. It could be cheaper to add standard elements to all wire harnesses, a tolerance of envelope effect will be integrated to try to decrease the very significant number of references due to the total differentiation. Development of the characteristics as a function of that tolerance will give the best price for the percentage that is permissible.

The actual representation of the wire harness is binary, which means that each option and variant is or is not in the final product. Development of the model may involve changing elements. Instead of regarding that one version of the air conditioning as incompatible with another one, it could be represented for the wire harness as a cable with a variant section that depends on the version.

REFERENCES

[1] P. Child, R. Diederichs, F.-H. Sanders, S. Wisniowski. *The Management of Complexity*, Sloan Management Review, 1991, pp. 73-80.

[2] J. Gonzalez-Zugasti, K. Otto, J. Baker. *Assessing value for product family design and selection*, 25th Design Automation Conference. ASME Design Engineering Technical Conferences, Las Vegas, 1999, Sept. 12-15.

[3] J. Jiao, M. Tseng. *A methodology of developing product family architecture for mass customization*, Journal of Intelligent Manufacturing, Vol. 10, 1999, pp. 3-20.

[4] A. Kusiak. *Engineering Design : Products, Processes, and Systems*, Academic Press, San Diego, CA, 1999.

[5] H.L. Lee, C.S. Tang. *Modeling the costs and benefits of delayed product differentiation*, Management Science, 43(1), 1997, p. 40-53.

[6] M. Martin, I. Kosuke. *Design for Variety: Development of Complexity Indices and Design Charts*, 1997 ASME Design Engineering Technical Conferences, Sept. 14-17, Sacramento, 1997, pp.1-9.

[7] JP.J. Newcomb, B. Bras, D.W. Rosen. *Implications of Modularity on Product Design for the Life Cycle*, Journal of Mechanical Design, 120, 1998, pp. 483-490.

[8] W. Zinn. *Should we assemble products before an order is receive?*, Business Horizons, March-April, 1990, pp. 70-73.

A UNIFIED JACOBIAN-TORSOR MODEL FOR ANALYSIS IN COMPUTER AIDED TOLERANCING

W. Ghie*, L. Laperrière* and A. Desrochers**

*Université du Québec à Trois-Rivières, Department of Mechanical Engineering, Canada

**Université de Sherbrooke, Department of Mechanical Engineering, Canada

Abstract: This paper presents a unified model for computer-aided tolerancing. It combines the benefits of the Jacobian matrix model and the torsor (or screw) model. The former is based on small displacements modeling of points using 6x6 transformation matrices of open kinematic chains in robotics. The latter models the boundaries of 3-D tolerance zones resulting from a feature's small displacements using a torsor representation with constraints. The proposed unified model expands the functionalities of the Jacobian model under two important aspects. First, the punctual small displacement variables of the former Jacobian formulation are now considered as intervals formulated and solved using interval-based arithmetic. The equations describing the bounds within which the feature is permitted to move, which are the constraint equations of the torsor formulation, are applied on the unified model. Second, some of the small displacement variables used in the model are eliminated due to the invariant nature of the movements they generate with respect to the toleranced feature. This standard result of the torsor formulation is applied to the unified model. The effect of this is to significantly reduce the unified model size. A example application is also presented.

Key words: tolerances, deviations, torsor, Jacobian, modeling

1. INTRODUCTION

Manufactured parts are seldom used as single parts. They are rather used in assemblies of different parts. Tolerances on single parts of an assembly are therefore likely to accumulate statically and propagate kinematically, causing the overall assembly dimensions to vary according to the number of

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 63–72.

contribution sources of variation. The critical clearances and fits which affect the whole assembly performance and function are thus subject to variations due to the tolerance stackups.

It is the goal of tolerancing to specify individual part variations that will not compromise a product's ease of assembly and proper function. In tolerance synthesis, one starts with a specification of the assembly's functional requirements in early stages of the design process and transforms them into suitable individual parts tolerances that will meet the functional requirements. In tolerance analysis, the task is reversed, i.e. one adds up the individual part tolerances kinematically and verifies their cumulative effects and validity with respect to some intended functional requirement. Clearly, tolerance synthesis is a more powerful tool as it enables direct specification of part tolerances rather then verifying their effects.

Currently very few CAD systems propose functionalities for assistance in the difficult task of tolerancing. It is the objective of this paper to present an innovative tolerancing model which mathematical formulation is compatible with current CAD systems and which lends itself to solving the dual problems of tolerance analysis and synthesis.

The next section of this paper describes some prior work in computer-aided tolerancing, including the two approaches which have been unified in the proposed formulation. Section 3 describes the unification process using the concepts of intervals, invariant degrees and constraint equations applied to the new model. Section 4 presents a solution of the equations in the unified model as applied to an example application. The last section discusses the results and concludes the paper.

2. BACKGROUND AND PRIOR WORK

A lot of work has been carried out in the past few years to develop a mathematical formulation of tolerances. Two main approaches can be distinguished: one that models the space allowed around theoretic (or nominal) geometric entities [1,2] and one that uses parameterization of the deviation from the theoretic geometric entities [2,3,4].

Wirtz [5] proposes a vectorial approach to tolerancing. Four vectors are used which describe the position, orientation, form and dimension of a part's feature. To each of these vectors are associated two parameters: one for the nominal state and the other for the deviation. This model describes a real surface by vectorial addition of nominal characteristics and their deviations. One advantage of this model is that it allows a link between the surface defects and their manufacturing methods [6]. Another vectorial tolerance approach, relatively different from that of Wirtz, is that of Gaunet [7]. The approach uses the concept of Technologically and Topologically Related

Surfaces (TTRS) introduced by Desrochers [8] in addition to their torsors of small displacements, also called small displacement screw. Small displacements are used to simulate variations between the surface of substitution and the nominal surface according to the tolerance specifications. The tolerances are modeled by torsors or screws. This model also introduces the concept of point sampling on the nominal surfaces.

Several researchers use kinematic concepts to study tolerancing problems. In [9,10], Whitney and al. propose two representations of tolerances to perform tolerance analysis. The former is based on the concepts of small displacements screws and the latter is based on the concepts of homogeneous transforms used in robotics. Chase [11] represents the contacts within a mechanism using kinematic connections. Then a matrix of connectivity is obtained which establishes the vector loops around a functional requirement. The resulting mathematical model can be used for the study of tolerance analysis or synthesis. This system is limited to two dimensions. In [12], Chase, Gao and Magleby generalize this 2-D model to 3-D applications using Hessian matrices.

Another kinematic formulation is proposed by Rivest, Fortin and Desrochers [13]. The description of tolerance zones is encapsulated within the kinematic structure. The model takes into account the type and position of a tolerance zone included in a generic kinematic chain. Each connection in the chain is associated with 14 pivot-type or slide-type connections. Each of the 14 connections can be blocked, left free or limited, according to the specification considered [14].

Laperrière and Lafond [15,16] proposed another kinematics-based method which uses Jacobian transforms and which integrates both tolerance analysis and synthesis in the same model. Modeling tolerances consists in applying small displacements to geometric features which are known to affect some functional requirements. The Jacobian enables transformation of locally expressed parts deviations to globally expressed functional requirements. Each deviation is said to be punctual as it applies to a single point on the part, corresponding to the origin of the local coordinate frame attached to it.

Desrochers [17] proposes a 3-D representation of the tolerance zones using small displacements screws to which the parameters of invariance defined by Bourdet and Clément are also applied [18]. The proposed representation makes use of mathematical constraints which define the extreme limits of a tolerance zone in 3-D.

Some researchers call upon operational research to solve problems in tolerancing. One such approach was introduced by Turner [19] and then improved by Porchet [20]. Zhang [21] proposes an approach to take the cost of manufacturing into consideration during both analysis and synthesis.

3. DESCRIPTION OF THE UNIFIED MODEL

3.1 Jacobian with intervals

The Jacobian approach described in [15,16] requires the prior identification of the tolerance chain, which passes through every functional element (or feature) dimension that is known to affect a functional requirement. A method for automating this identification process has been presented earlier [22]. In [16], the mathematical model describing the effects that small displacements of functional element (FE) have on some remote functional requirement (FR) is represented by:

$$\begin{bmatrix} d\vec{s} \\ d\vec{\alpha} \end{bmatrix}_{FR} = \left[\left[J_1 J_2 J_3 J_4 J_5 J_6 \right]_{FE1} \quad \cdots \quad \left[J_1 J_2 J_3 J_4 J_5 J_6 \right]_{FEn} \right] \bullet \begin{bmatrix} d\vec{q}_{FE1} \\ \cdots \\ d\vec{q}_{FEn} \end{bmatrix} \qquad (1)$$

Where:

$d\vec{s}$: 3-vector of the point of interest's (FR) small translations;
$d\vec{\alpha}$: 3-vector of the point of interest's (FR) small rotations;
$[J1..J6]_{FEi}$: i^{th} 6x6 Jacobian matrix associated with the toleranced FE of the i^{th} FE pair;
$d\vec{q}_{FE_i}$: i^{th} 6-vector of small dispersions associated with the i^{th} FE pair, i = 1 to n.

The resulting model takes the form of six equations with "$6\times N$" unknown factors, where "N" is the number of pairs of functional elements in the identified tolerance chain. The Jacobian approach has no notion of a tolerance zone, unless many different such small displacements are applied successively to the feature, in order to cover the zone in which it is permitted to move. To solve this problem, the punctual-based formulation of the former model in equation (2) must be changed to the interval-based formulation in equation (3). Compared to equation (2), the intervals used in equation (3) enable modeling of tolerance zones instead of single points.

3.2 Eliminating invariant degrees

It was shown in [17] that among the six possible small displacements in 3-D that can be applied to a feature, some may leave the resulting surface invariant. Hence, the variables corresponding to such displacements should be cancelled out by setting them to zero in the right hand side column of the model in equation (3). For example, in figure 2 the small displacements of surface "0" in "y" and "z" and the small rotation around "x", respectively "v", "w" and "α", leave the planar surface invariant. Each such cancelled invariant variable will also eliminate a corresponding column of the Jacobian, leading to substantial simplification of the computations involved.

$$\begin{bmatrix} u \\ v \\ w \\ \alpha \\ \beta \\ \delta \end{bmatrix}_{FR} = \Big[[J_1 J_2 J_3 J_4 J_5 J_6]_{FE_1} \; [\ldots\ldots]_{FE_{n-1}} \; [J_{6n-1} J_{6n-2} J_{6n-3} J_{6n-4} J_{6n-5} J_{6n}]_{FE_n} \Big] \begin{bmatrix} \begin{bmatrix} u \\ v \\ w \\ \alpha \\ \beta \\ \delta \end{bmatrix}_{FE_1} \\ [\ldots]_{FE_{n-1}} \\ \begin{bmatrix} u \\ v \\ w \\ \alpha \\ \beta \\ \delta \end{bmatrix}_{FE_n} \end{bmatrix} \tag{2}$$

Where:

u, v, w : Translation dispersions about X, Y, Z, respectively;

α, β, δ : Rotation dispersions around X, Y, Z respectively.

$$\begin{bmatrix} \underline{u} & \overline{u} \\ \underline{v} & \overline{v} \\ \underline{w} & \overline{w} \\ \underline{\alpha} & \overline{a} \\ \underline{\beta} & \overline{\beta} \\ \underline{\delta} & \overline{\delta} \end{bmatrix}_{FR} = \Big[[J_1 J_2 J_3 J_4 J_5 J_6]_{FE_1} \; [\ldots\ldots]_{FE_{n-1}} \; [J_{6n-1} J_{6n-2} J_{6n-3} J_{6n-4} J_{6n-5} J_{6n}]_{FE_n} \Big] \begin{bmatrix} \begin{bmatrix} \underline{u} & \overline{u} \\ \underline{v} & \overline{v} \\ \underline{w} & \overline{w} \\ \underline{\alpha} & \overline{a} \\ \underline{\beta} & \overline{\beta} \\ \underline{\delta} & \overline{\delta} \end{bmatrix}_{FE_1} \\ [\ldots \quad \ldots]_{FE_{n-1}} \\ \begin{bmatrix} \underline{u} & \overline{u} \\ \underline{v} & \overline{v} \\ \underline{w} & \overline{w} \\ \underline{\alpha} & \overline{a} \\ \underline{\beta} & \overline{\beta} \\ \underline{\delta} & \overline{\delta} \end{bmatrix}_{FE_n} \end{bmatrix} \tag{3}$$

Where:

$\underline{u}, \underline{v}, \underline{w}, \underline{\alpha}, \underline{\beta}, \underline{\delta}$: Lower limit of $u, v, w, \alpha, \beta, \delta$ defining the tolerance zone;

$\overline{u}, \overline{v}, \overline{w}, \overline{\alpha}, \overline{\beta}, \overline{\delta}$: Higher limit of $u, v, w, \alpha, \beta, \delta$ defining the tolerance zone.

3.3 Adding the constraints of the torsor formulation

The intervals expressed in the new model must be bound to the limits of the intended tolerance zone which they represent. These limits are hypersurfaces of the space spanned by the six variables "u, v, w, α, β, δ" and for most common shapes of tolerance zones used in practice, they are not independent. Therefore some inequalities representing such constraints will sometimes involve dependencies between two of the above displacement variables. Moreover, it should be noted that the dependencies will themselves be dependent upon the type of tolerances being represented by the tolerance zones. For instance, for tolerances of location, there should indeed be dependencies between the position and orientation components of the screw while for orientation tolerances, the angular components of the screw should remain independent of those for translation.

4. A TOLERANCE ANALYSIS EXAMPLE

We will use the simple example in figure 1 and 2 to describe the approach.

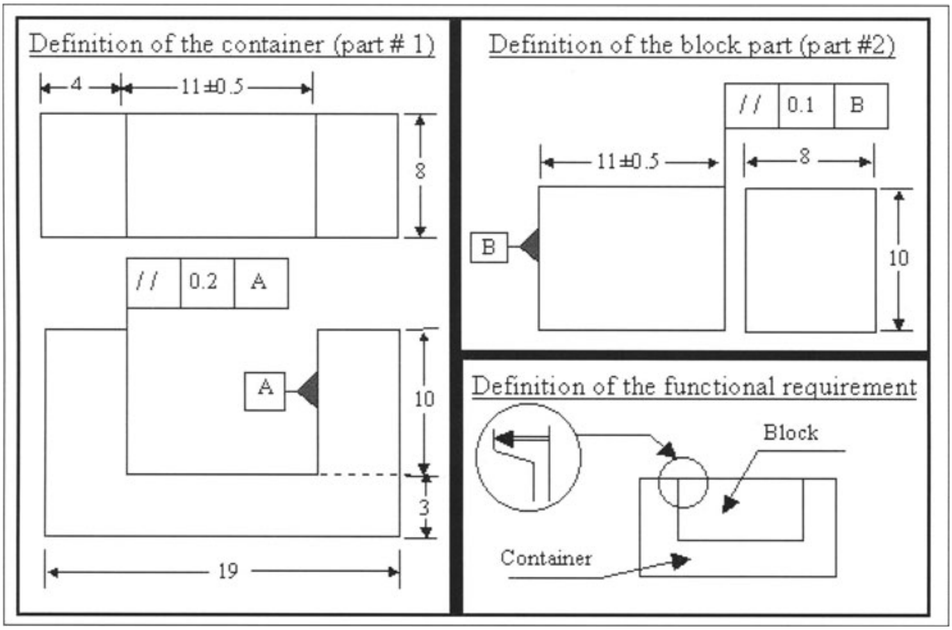


Figure -1 Definition of the block and the container with a functional requirement

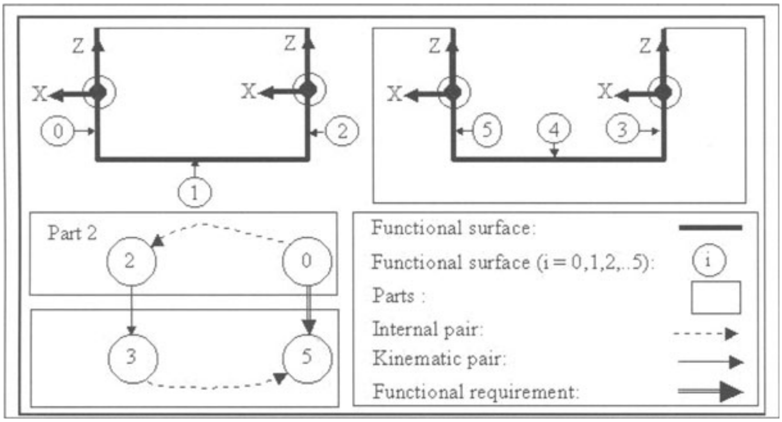


Figure -2. Definition of the part's surfaces and their associated coordinate frames

Using the surface labels in figure 2, we trivially find the tolerance chain that identifies which functional elements of both parts influence the functional requirement in figure 1. They are: internal pair (0,2), internal pair (3,5). Using the frame definitions in figure 2, we also trivially find the Jacobian:

$$J = \left[\begin{array}{cccccc|cccccc} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 11 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -11 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{4}$$

Since the tolerance chain goes through planes only, we need to apply the constraints of only this type of geometry, as follows.

Table -1. Tolerance zones and associated torsor parameters

From plane «0» to plane «2» (Internal pair of block part)	**From plane «3» to plane «5» (Internal pair of container part)**
$-\frac{1}{2} \le u \le +\frac{1}{2}$ $-\frac{0.1}{10} \le \beta \le +\frac{0.1}{10}$ (5) $-\frac{0.1}{8} \le \delta \le +\frac{0.1}{8}$	$-\frac{1}{2} \le u \le +\frac{1}{2}$ $-\frac{0.2}{10} \le \beta \le +\frac{0.2}{10}$ (6) $-\frac{0.2}{8} \le \delta \le +\frac{0.2}{8}$

The final expression becomes:

$$\begin{bmatrix} [-1,+1] \\ [-0.1375,+0.1375] \\ [-0.11,+0.11] \\ [0,0] \\ [-0.03,+0.03] \\ [-0.0375,+0.0375] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 11 & 0 & 0 & 0 \\ 0 & -11 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} \begin{bmatrix} [-0.5,+0.5] \\ [-0.01,+0.01] \\ [-0.0125,+0.0125] \end{bmatrix} \\ \begin{bmatrix} [-0.5,+0.5] \\ [-0.02,+0.02] \\ [-0.025,+0.025] \end{bmatrix} \end{bmatrix} \qquad (7)$$

We first note that the constraints in equations (5) and (6) use the dimensional tolerance 11±0.5 for the small translations and the parallelism tolerances (0.2 and 0.1 for container and block, respectively) for the small rotations. Thus we have a means of predicting the contributions at the functional requirement of both dimensional and geometric tolerances applied on a same feature in the chain.

We see also that for the "x" direction, which is that of the functional requirement in figure 1, we have the standard result in 1-D that the sum of tolerances in the chain equals that of the functional requirement.

We also see the effects of the small rotations permitted by the two parallelism geometric tolerances. They contribute to translations of final surface "5" in both the "y" and "z" directions, by an amount of [-0.1375, 0.1375] and [-0.11, 0.11] respectively. These constitute invariant degrees for surface "5" so it seems they are not relevant for this particular example. Still, the three small translations computed at the functional requirement define the zone inside which the final frame on surface "5" will move due to the tolerances assigned, while the three small rotations (the one around "x" being null) define its resulting possible orientations. What remains to be done is to project these displacements of the final frame at the four corners of surface "5" to determine the tolerance zone inside which this surface moves as a result of the tolerances assigned.

Finally, the above example is clearly a worst case application, since the Jacobian multiplies intervals that define the extreme values that each small displacement can take. The interpretation is therefore that 100% of the parts produced (container and block) that satisfy their individual assigned tolerances (figures 1) will necessarily result in 100% of assemblies for which the functional requirement will be bound to the values of the left column in equation (7).

5. CONCLUSION

A method for combining the benefits of the Jacobian and torsor approach in computer-aided tolerancing has been presented. The result is a so called unified Jacobian-torsor model which takes into account the boundaries of the intended tolerance zones as modeled by the mathematical constraints applied to it. It is an innovative tolerancing model in which the mathematical matrix-based formulation is compatible with current CAD systems.

A method for solving the equations of the unified model has been proposed. The limitation is that the solved values are those of the tolerance zone boundaries, therefore restricting this solution approach for worst-case type applications only. Other methods are currently under investigation in order to make the model also applicable in statistical situations, by considering the intervals as the limits of some statistical distributions with possible percentages of reject.

The underlying interval arithmetic behind the approach makes it mathematically robust and generic [23].

The tool is compatible with the standards in the sense that the dimensional or geometric tolerances of a drawing can be mapped into the variables of the model for performing tolerance analysis.

Finally, the developed tool is also suitable for performing the dual problem of tolerance synthesis: one simply needs to invert the Jacobian. We have started working in this direction and preliminary results are available [24].

ACKNOWLEDGEMENTS

The authors would like to thank the Natural Science and Engineering Research Council (NSERC) of Canada for its financial support through the individual grant program.

List of references

1. A.A.G. REQUICHA. «*Towards a theory of geometric tolerancing*», the International Journal of Robotics Research, Vol. 2, n4 , 1983, pp. 45-60.
2. E. BALLOT, P. BOURDET, F. THIÉBAUT. «*Determination of relative situations of parts for tolerance computation*», 7th CIRP International Seminar on Computer Aided Tolerancing, ENS de Cachan, France, 2001, April 24-25.
3. L. RIVEST, C. FORTIN, C. MOREL. « *Tolerancing a solid model with a kinematic formulation* », Computer-Aided Design, v 26, n6, 1994, pp. 465-485.
4. A. DESROCHERS, A. CLÉMENT, A. RIVIÈRE. «*Theory and practice of 3D tolerancing for assembly*», CIRP International Working Seminar on Tolerancing, Penn. State Univ., 1991, pp. 25-55.
5. A. WIRTZ. «*Vectoriel tolerancing*», International Conference on CAD/CAM and AMT in Israël, CIRP session on tolerancing for function in a CAD/CAM environment, Vol. 2, dec. 11-14, 1989.
6. E. BALLOT.«*Lois de comportement géométrique des mécanismes pour le tolérancement* », Thèse Ph. D., École Normale Supérieure de Cachan, 1995.
7. D. GAUNET. «*Vectoriel tolerancing model* », 3rd CIRP Seminar on Computer Aided Tolerancing, April 27-28, 1993.

8. A. DESROCHERS. «*Modèle conceptuel du dimensionnement et du tolérancement des mécanismes : Représentation dans les systèmes CFAO*», Thèse Ph.D., École Centrale de Paris, 1991.
9. D.E WHITNEY. , J.D. ADAMS. «*Application of screw theory to motion analysis of assemblies of rigid parts*», Proceedings of IEEE International Symposium on Assembly and Task Planning, Porto, Portugal, July 1999, pp. 75-80
10. D.E WHITNEY., O.L. GILBERT, M. JASTRZEBSKI. «*Representation of geometric variations using matrix transformation for statistical tolerance analysis in assembly*», Research in engineering, Vol. 6, 1994, pp.191-210.
11. K. W CHASE, J. GAO, S. P. MAGLEBY. «*Generalized 2-D Tolerance Analysis of Mechanical Assemblies with Small Kinematic Adjustments*», Journal of Design and Manufacturing, Vol. 5, 1995, No. 2.
12. K. W CHASE, J. GAO, S. P. MAGLEBY. *«General 3-D Tolerance Analysis of Mechanical Assemblies with Small Kinematics Adjustments»*, IIE Transactions, 1998, Vol. 30, , pp. 367-377.
13. A. DESROCHERS, L. RIVEST, C. FORTIN. « *Tolerance modeling for 3D analysis: Presenting a kinematic formulation*», 3^{rd} CIRP Seminar on Computer Aided Tolerancing, April 27-28,1993.
14. L. RIVEST, C. MOREL, C. FORTIN. «*Tolerancing a solid model with a kinematic formulation*», Computer Aided design, Vol. 26, 1994, N 6 , pp. 465-485,.
15. L. LAPERRIÈRE, P. LAFOND. «*Identification of dispersions affecting pre-defined functional requirements of mechanical assemblies »* Proc. Of IDMME Conference, Compiège , France, May27-29,1998, pp. 721-728.
16. L. LAPERRIÈRE, P. LAFOND. «*Tolerances Analysis and synthesis using virtuals joints* », 6^{th} CIRP international seminar on computer aided tolerancing, Enschede, Netherlands, March 22-24, 1999, pp 405-414.
17. A. DESROCHERS. «*Determination of Part Position Uncertainty Within Mechanical Assembly Using Screw Parameters*», CD-ROM Proc. 25^{th} ASME Design Automation Conference, Las Vegas, 1997, DAC-8587.
18. P. BOURDET, E. BALLOT. «*Équations formelles et tridimensionnelles des chaînes de dimensions dans les mécanismes* », Séminaire Tolérancement et chaînes de cotes, fév., 1995, pp. 135-146.
19. J. TURNER, S. SRIKANTH. «*Constraint Representation and Reduction in Assembly Modeling and Analysis*», Rensselear Design Research Centre, Tech Report , 1990, No. 90027.
20. G. ZHANG, M. PORCHET. «*Some new developments in tolerance design in CAD* », Proceedings of the 19^{th} ASME Annual International Design Automation Conference And Exposition, Albuquerque, Vol. 2, 1993, pp.175-185.
21. G. ZHANG. « *Simultaneous tolerancing for design and manufacturing* », Int. J. Prod. Res., Vol. 34, N° 12, 1996, pp. 3361-3382.
22. W. GHIE, L. LAPERRIÈRE. «*Automatic generation of tolerance chains around functional requirements of mechanical assemblies*», CD-ROM Proceeding of 3^{rd} IDMME conference, Montreal, Canada,2000.
23. E. HANSEN. «*Global optimization using interval analysis*», *Marcel Dekker*, ISBN 0-8247- 8696-3, 1992.
24. L. LAPERRIÈRE, T. KABORÉ. «*Monte Carlo simulation of tolerance synthesis equations* », Int. J. Prod. Res., Vol. 39, N° 11, 2001, pp. 2395-2406.

A PROPOSAL OF STRUCTURE FOR A VARIATIONAL MODELER BASED ON FUNCTIONAL SPECIFICATIONS

D. Lesage*, J-C. Léon, P. Sebah* and A. Rivière**
Dassault Systèmes, Suresnes, France
Laboratoire des Sols, Solides, Structures, Grenoble, France
***ISMCM, Saint-Ouen, France*

Abstract: For several years, CAD software have adopted a variational approach for modelling. This choice is justified by users' needs. Indeed, a product is often designed to fulfill requirements that are not expressed as geometric specifications. In addition, the geometric model of the product evolves continuously during the design process. During this process, the software must be able to preserve the coherence of the model specifications. To fit into this context, a variational modeler has been set up which includes new concepts of analysis and solving approaches to identify locally over/under-constrained systems and provide more feed back to the designer. The structure of the modeler is briefly outlined followed by the introduction of the analysis and solving modules that are clustered into three tasks: a pre-solver analysis, a solver, a post-solver analysis in case of resolution failure. Parts of examples are used to illustrate the descriptions of this structure.

Key words: variational geometry, functional specifications, design modeler, constraint analysis, problem decomposition

1. INTRODUCTION

For several years, CAD software has developed a parametric representation scheme. Such a technological choice has been driven by users' needs. Actually, a user designs a product in order to fulfill technical requirements. In fact, geometric constraints are merely a consequence of functional specifications. Some of these specifications will be later referred to as engineering constraints when they have no geometric meaning.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 73–84.

The geometric model of a product is subjected to continuous changes during the design process until the functional requirements are fulfilled. Hence, CAD software must provide means to modify/update a model to reach the user specifications. Within this context, many variational or parametric geometric solvers have been proposed. Even though these solvers have now become fairly robust in 2D, their efficiency in 3D is restricted to simple configurations. In addition, some approaches initially devised for geometric purposes only cannot be extended to handle engineering specifications (e.g. equations relating the dynamic equilibrium of a rigid body to geometric parameters). Other approaches, based on a powerful mathematical solver have been derived from the initial problem stated, i.e. to solve a geometric problem often ill constrained (locally over/under constrained) and to provide the user with an intuitive solution.

This paper shows how, starting from an original geometric modeler, analyzer and solver modules have been set up. Section 2 briefly describes the principle of the variational modeler. Then, section 3 outlines the structure set up for analyzing and solving the functional specifications. Section 4 provides details about the analysis phase and section 5 about the resolution scheme generated. Finally, section 6 illustrates how a post resolution analysis can be performed to provide the user with a significant diagnosis in case of failure of the solving process.

2. PRINCIPLES OF THE GEOMETRIC MODELER

An object geometry, either of volume, surface or line type is composed of a set of elementary objects. All these objects can be defined using the position of a set of characteristic points located in the workspace with the addition, in some cases, of complementary data (e.g. knot sequence and homogeneous coordinates for NURBS entities). In addition, using a B-Rep approach, these objects must be bound together using topological relationships to satisfy the necessary coherence constraints of the object.

Usually, a geometric entity is located in space using the Cartesian coordinates of its characteristic points. The basic idea of this modeler is to describe naturally the relative position of the various entities, i.e. an entity is solely defined by its dimensions and/or its relative position with respect to others. To this end, the various characteristic points of an entity are expressed using geometric constraints with respect to those of other entities. Such a set of points and relationships can be seen as a skeleton of the objects geometry. This approach relying on a volume model of type B-Rep, the deformable body is a set of surface entities bound together using topological relationships. This body defines the skin of the object. Each skin entity is attached to its associated characteristic points which are located in space using its associated skeleton entity. This modeler can be called a declarative modeler.

The skeleton: It must contain the data required to locate the characteristic points in space. Clément et al. [2] demonstrated that a polyhedral object can be specified from 13 elementary geometric constraints: the TTRS (Technologically and Topologically Related Surfaces). These constraints can be generated from a set of two elementary constraints: the distance between two points and the angle between two segments [3]. Using points, bi-points, unit vectors and angles between unit vectors, a data structure describes the skeleton (Fig. 1).

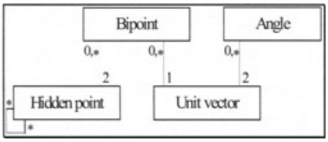


Figure -1. UML of the skeleton

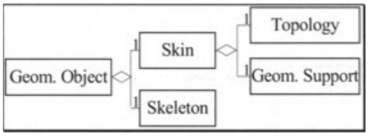


Figure -2. UML of a geometric object

From these data, a pre-solver module allows to generate a necessary and sufficient set of algebraic equations [7]. The user will be able to update this set with a set of engineering constraints defined between the same geometric parameters and some external parameters (e.g. mass, forces, stresses, …). Classically, a well-suited equation system is considered as equivalent to a construction "with the rule and the compass" whereas other configurations lead to a highly intricate system, hence robustness cannot be achieved.

The skin: The data attached to the skin is stored in a data structure of B-Rep type (Fig. 2). The only difference with respect to a B-Rep approach remains the intrinsic properties of the objects solely used to define them. Their position in space is defined by the coincidence relationship between each characteristic point of the object and a point of the skeleton.

3. THE SOLVER APPROACH

The overall behavior of the solver is based on a top-down process where the overall problem is progressively subdivided [1, 5, 8, 11]. To this end, various modules are set up: an analysis module, a resolution module, a post-resolution analysis module in case of resolution failure. Contrary to the previous approaches however, the solver is not bound to a straightforward resolution of the equation system provided by the modeler. Let us assume that this module has as input an equation system representing a construction "with the rule and the compass". Thus, starting from a set of algebraic equations, the solver module has to:

– perform a structural analysis of the problem to reduce its complexity,
– analyze each SIS of the solver roadmap to assess its constraint level and, possibly, to correct the solver roadmap,
– solve each SIS,
– propose other solutions for the current problem,
– generate a meaningful diagnosis for the designer in case of resolution failure.

In order to fulfill these needs, the solver performs a set of successive tasks (Fig.3). In the first place, a first solver roadmap is generated from a structural analysis of the whole set of equations. This analysis being strictly structural, it is possible to generate effective under-constrained SIS. In a second step, a more precise analysis is performed on each SIS. If an under-constrained state has been detected, a rebalancing algorithm is applied to this SIS and the solver roadmap is adequately updated. As soon as this task ends, the solver roadmap is globally as well as locally well constrained and the resolution itself can start. The resolution may behave well. In this case, a first solution can be submitted to the designer. However, the resolution may stop at any point of the roadmap. One cause may be the failure of the resolution algorithm. Hence, a question can be raised: "Does a solution for this SIS exist ?". A negative answer to this question draws the diagnosis to an existency problem of the solution. A second cause may arise from an "over-constraint" which is incompatible with the current value of the partial solution. In this case, the module diagnoses an incoherence problem. Once the failure category has been identified, a final module helps localize the failure to give clear explanations to the designer.

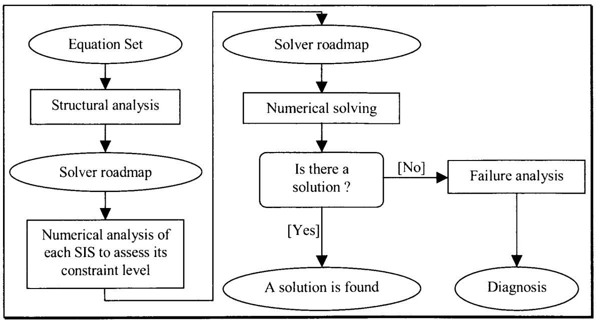


Figure -3. Overall solving algorithm.

4. THE ANALYSIS MODULE

The analysis module is based on a bipartite graph [10]. A bipartite graph is a graph whose set of vertices can be partitioned into two classes such that two vertices of the same class can never be adjacent. Such a graph is useful to analyze an equation system: the first set of vertices aggregate the equations and the second one the variables (Fig. 4). Indeed, the structural properties of a bipartite graph are not sufficient to analyze thoroughly a system. As an example:

- an over-constrained equation system is either redundant or contradictory but the bipartite graph cannot distinguish these configurations,
- generally, a graph structurally under-constrained is represented by an equation system that may provide an infinite number of solutions. Actually, such a system can generate a finite (or empty) set of solutions. As an example, the under-constrained system: “$x^2 + y^2 = 0$” has only one solution: $x = 0$, $y = 0$,
- a system structurally iso-constrained can incorporate redundant or contradictory equations (Fig. 6).

The bipartite graph approach provides a first analysis which subdivides the initial problem into a set of problems of smaller size with some independence between them. In all cases, the subdivision obtained does not modify the effective status of the initial problem. A deeper analysis will be performed later for each SIS using other tools.

The structural analysis performed on the bipartite graph of the equation system, subdivides it into subsystems solvable independently of each other, i.e. SIS. The SIS obtained are further ordered into a directed graph which represents the solver roadmap (Fig. 5). Each vertex of this graph designates a SIS. The directed arcs express the dependencies between the SIS. The use of a directed graph is particularly useful from an algorithmic point of view, i.e. at the stage of the update of the solver roadmap, when solving the problem itself or when diagnozing and analyzing failure of the resolution. Figure 5 provides a solver roadmap with seven SIS with one SIS equivalent to a system of two equations and two unknowns. In this example, the first three SIS (4, 5, 6) can be solved in parallel. The two following ones (3, 7) can be solved only in a second step. In addition, SIS 7 can be solved as soon as SIS 6 is solved. SIS 3 must wait until the first three have been resolved. Finally, solving SIS 1 and 2 depends solely on the resolution of SIS 3.

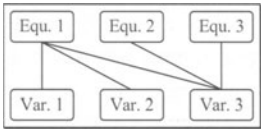


Figure -4. Bipartite graph

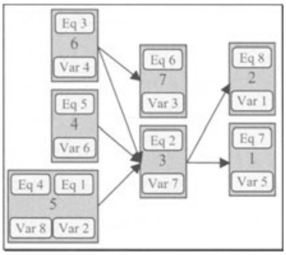


Figure -5. Example of a solver roadmap

$$\begin{cases} X_1 + X_2 = 5 \\ 2X_1 + 2X_2 = 1 \end{cases} \qquad \begin{cases} X_1 + X_2 = 5 \\ 2X_1 + 2X_2 = 10 \end{cases} \qquad \begin{cases} X_1 + X_2 = 10 \\ 2X_1 + 2X_2 = 5 \end{cases}$$

(a) (b) (c)

Figure -6. Structurally identical systems

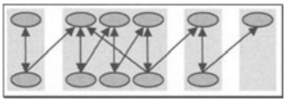


Figure -7. Graph decomposed into SIS

Numerical analysis of each SIS: The bipartite graph only provides a structural analysis of an equation system but it cannot detect configurations where the system is effectively under-constrained, i.e. one equation is a combination of a equation sub-system of the SIS. The example in figure 6 illustrates this problem: the three equation systems are structurally identical but the system (a) is a well constrained system with a finite number of solutions; system (b) is under-constrained, it has an infinite number of solutions and system (c) is incoherent, it has no solution.

A first test is to compute the value of the jacobian with the initial value of the parameters to verify that the system is non-degenerate. The second test is the probabilistic scheme as proposed by Lamure [6]: the value of the jacobian is computed at some random points. If it vanishes each time, then the probability is very close to 1 that the jacobian is identically zero.

Local modification of the solver roadmap: The above numerical analysis has led us to locate a redundant equation into a SIS. From a geometric point of view, it has identified, from the solver roadmap, an under-constrained geometric area. The idea is to check if this under-constraint can be balanced by an existing over-constraint. From a graph point of view, the result of the numerical analysis states that one equation is redundant and, hence, its associated node in the bipartite graph should not have been coupled with this equation. The removal of this couple (equation, variable) frees one variable and hence its node associated that should not be coupled in the graph. An under-constraint graph area appears. If free "equation" nodes exist, it is necessary to search for coupling of higher cardinality. In this case, the variable freed and one node associated with one of the over-constrained equations become coupled. The over-constraint level decreases. The system has been locally balanced. The following scheme describes the algorithms used for such a modification.

Figure 7 represents a bipartite graph decomposed into SIS. This graph is over-constrained, i.e. one equation is not coupled. Assuming that the numerical analysis has identified one of the equations of the second SIS as redundant, this equation is removed from the graph (Fig. 8). The corresponding SIS becomes under-constrained, therefore it is necessary to verify that the global system constraint level could not be decreased. In other words, it is necessary to check that the graph coupling is still of maximum cardinality. If the graph has a uncoupled set of equations (other than the equation just removed) and if there is a path between the newly free variable and one non coupled equation (Fig. 9), then the cardinality of the coupling can be increased to become maximum. Such a path is searched for in the oriented graph where the coupled arcs are bi-directional and the uncoupled ones are oriented in the direction variable-equation.

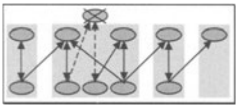


Figure -8. A redundant equation identified

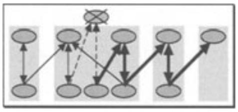


Figure -9. A redundant equation localized

Once such a path has been found, each coupled arc is transformed into an uncoupled one and conversely. Thus, the coupling cardinality has been maximized. To demonstrate this property, one has to consider:

- the graph is initially of maximum cardinality, therefore removing a coupled equation reduces by one the cardinality of the graph coupling,
- if there exists one uncoupled equation, given the variable that has been freed, the cardinality of coupling can potentially be increased by one,
- if there is a path in the oriented graph (described previously), this path will link an uncoupled variable to an uncoupled equation. All the nodes, except the extremities, are attached to one and only one coupled arc. The path is composed of an odd number of arcs ($2n$+1), therefore it contains n coupled arcs. If the arc type is changed (coupled ones become uncoupled ones and conversely), the extreme nodes of the path are therefore linked (like the others) to one and only one coupled arc (definition of the coupling). Hence, the path contains (n+1) coupled arcs,
- if there is no path between the freed variable and one of the uncoupled equations of the graph, then the cardinality of the coupling of this graph cannot be increased,
- if, in the graph, there are other uncoupled variables or equations and if before equation removal the coupling is of maximum cardinality n, then

the maximum cardinality of the coupling after the equation removal is at most *n*. Indeed, if there are other uncoupled variables or equations before equation removal, then there is no path linking one of the free variables to one of the free equations. The removal of the equation leads to the removal of a set of arcs and hence decreases the number of paths between the various nodes without modifying the existing paths. Thus, after removing the equation, it will not be possible to find a path across the other free nodes.

The coupling has now become of maximum cardinality, it is now required to decompose the graph into strongly connexe components. Such a decomposition can be performed on the overall graph but the decomposition into strongly connexe elements with respect to a coupling is unique. The coupling modification is only local, the modification of the decomposition will be only local too. It is therefore justified to restrict this decomposition to the sub-graph where the decomposition is modified. This sub-graph (Fig. 10) is defined by the set of paths derived from the variables of the modified SIS and ending at the newly coupled equation. The decomposition is now performed in this sub-graph to generate the new SIS. The previous SIS are removed from the solver roadmap and the new ones are inserted (Fig.11).

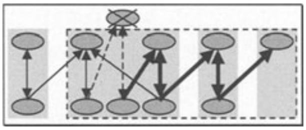


Figure -10. Sub-graph with target path

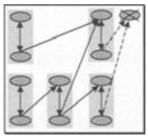


Figure -11. Solver roadmap after modification

5. RESOLUTION ALGORITHM OF A WELL CONSTRAINED PROBLEM

The overall resolution algorithm is rather simple and can be stated as: "while an unresolved SIS exists, try to solve it with the most suited solver; in case of failure, come back to the last SIS solved to try to find another solution".

Difficulties appear as soon as the solver cannot find a solution for a given SIS. Indeed, if there one exists, it should be found. To search without success or announce a wrong diagnosis, the ideal would be to know whether or not the SIS has a solution. This would allow differentiation of numerical failure from the lack of existence of a solution. In the case of numeric failure, another solver can be used and/or other initial conditions can be generated otherwise there is no clear decision to be made. A new question has been raised: "how to know if a solution exists at least for a given SIS and how to find it ?". In fact, if this global problem cannot be answered, the user should be provided with an efficient diagnosis.In this algorithm, one of the steps is the choice of the best suited solver for a given SIS.

Multiplicity of solutions: The roadmap provides, in a first step, a first estimate of solutions. It gives also, in certain conditions, all the problem solutions. The strongest condition is to determine all the solutions of a SIS. This holds if each SIS is dual to a construction "with the rule and the compass", i.e. if all the algebraic solutions of the SIS are known. In this case, all the solutions of the SIS can be given to the user. If this configuration does not hold, it may be possible to provide alternative solutions to the user: either when there is at least one SIS whose algebraic solutions are known and not unique or when multiple solutions of a SIS can be found numerically. The following theorem used by [4] to find the roots of a polynomial system can be stated as an example [9]. Considering a non degenerate polynomial system F(x)=0 for which there is a upper bound *n* of the number of solutions given by the Bézout theorem [4]. F is the target system of an homotopy with complex numbers based on a polynomial system G(x)=0 having also *n* roots. Usually, the homotopy curves originating from each root of G either end at a root of F or diverge to infinity. The key aspect of this theorem lies in the fact that a root of multiplicity *d* of F is generally reached by exactly *d* homotopic curves.

The capability to offer multiple solutions to a configuration is critical especially in 3D. Indeed, geometric solvers cannot always provide the most suited solution. In addition, their local behavior (usually the solution obtained depends upon the initialization conditions) produces only a restricted set of solutions. The user is then bound to reconstruct his/her geometric model to address other solutions.

Over-constrained configurations: There are two classes of over-constrained problems. The first designates a set of redundant but coherent geometric constraints or specifications. The second stands for geometric configurations described by an incoherent set of specifications. The difficulty in the analysis of such configurations relies on the fact that they only differ from a numerical point of view. Thus, a solver must address coherent over-constrained configurations, i.e. hyperstatic configurations.

Using the proposed approach, an over-constraint appears as a supplementary equation. It will be only after the numerical resolution step that the level of over-constraint will be addressed. In the following sections, a redundant equation will be considered equivalent to an over-constraint.

Inequalities: The dimensional constraints are assigned specific values. However, to reduce the number of solutions, the user may want to specify an interval of admissible values of some geometric parameters. P. Serré [12] has proposed a set of constraints expressing such needs (chirality constraints and region constraints). These constraints can be expressed as inequalities. In addition, the inequalities contribute to the extension of the scope of these new constraints since they can be considered as an equation over-constraining the problem. Their management is similar to the management of a redundant equation. To this end, no difference will be made in the following sections between a redundant equation and an inequality.

Solver roadmap, solution tree of an over-constrained problem: Within the structural analysis, a redundant equation is an equation not coupled to a variable. Inside the roadmap, these uncoupled equations will be checked only at a given stage of the resolution process. They can:

- reduce the number of solutions,
- be incoherent with the set of specifications,

but the resolution does not change with respect to the type of over-constraint. It is always necessary to scan the solution tree until a solution is found. If there is no solution, the analysis module will be able to locate the *inconsistency problem* of the specifications.

6. « POST-RESOLUTION » ANALYSIS

This analysis module is set up to bring a maximum amount of relevant information to the user in case of resolution failure. The various configurations of failure have already been described but telling the user that the failure is either a problem of existence or consistency problem among the specifications is not enough. The resolution of a geometric problem can be seen as a sequence of geometric construction and assembly steps. These steps are expressed by the SIS solver roadmap. It is therefore necessary to explain to the user the failure with:

- a localization of the rigid subsets already built,
- an explanation of the objective (assembly and/or construction) of the SIS that fails,
- indications concerning the specifications involved at the current step.

Figure 12 gives a sequential solver roadmap, each step is therefore a geometry construction step. Let us assume that SIS 1 leads to the construction of figure (a) and so on. Let us assume that SIS 3 does not provide a solution. In order to locate the problem for the user, a color code is used as described in figure 13. Then, the user can easily distinguish the part already built and the parameters used for that part from the part under construction and its associated parameters and an irrelevant part.

With this post-resolution analysis, only the sub-graph defined by the set of SIS solved and the SIS lacking a solution are considered. Such a sub-graph will be called *solved sub-graph*. The main feature of the diagnosis algorithm is to determine whether the failing SIS is associated with an assembly step or a construction step. If, within the solved sub-graph, the SIS is a connection point of the roadmap, then it is a SIS of assembly. In this case, it is necessary to extract the connexe sub-graphs attached at this point when scanning the roadmap in the opposite direction starting from the connection point considered. As soon as a sub-graph is identified, a rigid part is identified. Otherwise, the SIS characterizes a construction step.

7. CONCLUSION AND PERSPECTIVES

In this document, the structure of the solving module has been described. This module and its architecture try to provide a maximum amount of information to the user:

– when a resolution failure is faced, where it analyzes the configuration to give the user a clear diagnosis,
– when the resolution succeeds where it can offer alternative solutions if the one provided in first place is not satisfactory for the user.

The proposed solution addresses also the configuration of consistently over-constrained problems. This solution can also be extended to handle inequalities to extend the specification range. The use of SIS decomposition is proved to be efficient, however, the limits of a strictly structural analysis have been demonstrated. A numerical analysis module has been proposed to locate, before the resolution itself, the redundancy problems and to correct them whenever possible. Though the proposed structure enhances the efficiency of the solver, it is necessary to address more deeply the problem of existence and multiplicity of solutions through future developments since the work described covers this aspect only superficially.

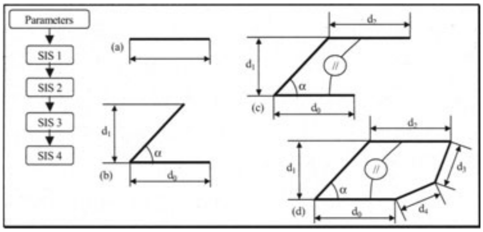


Figure -12. Construction steps of a solver roadmap

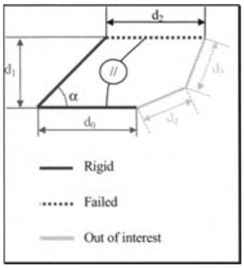


Figure -*13.* Failure diagnosis

REFERENCES

1. S. AIT-AOUDIA, R. JEGOU, D. MICHELUCCI, "Reduction of constraint systems", in Compugraphic, 1993, pp. 83-92.
2. A. CLEMENT, A. RIVIERE, P. SERRE "Geometry and topology declarative: a new paradigm for CAD-CAM systems", proceedings of IDMME 1998, Vol. 2, 1998, pp 587-595.
3. A. CLEMENT, A. RIVIERE, P. SERRE, "Global Consistency of Dimensioning and Tolerancing", 6th CIRP Seminar on Computer Aided Tolerancing, 1999.
4. C. DURAND, "Symbolic and numerical techniques for constaint solving", PhD thesis, Faculty of Purdue University, 1998.
5. I. FUDOS, "Constraint solving for computer aided design", PhD thesis, Faculty of Purdue University, 1995.
6. H. LAMURE H., D. MICHELUCCI "Qualitative Study of Geometric Constraints", In Geometric Constraint Solving and Applications, B. Brüderlin/D. Roller, Springer-Verlag, 1998.
7. D. LESAGE, J-C. LEON, P. SERRE. "A declarative approach to a 2D variational modeler", proceedings of IDMME, Montréal, May, 2000.
8. P. MATHIS, "Constructions géométriques sous contraintes en modélisation à base topologique", PhD thesis, Université Louis Pasteur, France, 1997.
9. A. MORGAN, "Solving polynomial systems using continuation for engineering and scientific problems", Prentice-hall inc, ISBN 0-13-822313-0 025, 1987.
10. K. MUROTA, "Systems Analysis by Graphs and Matroids, Structural Solvability and Controllability", Springer Verlag, 1987.
11. J.C. OWEN, "Algebraic solution for geometry from dimensionel constraints", In proceedings of the Symposium on Solid Modeling Foudations and CAD/CAM Applications, 1991, pp. 397-407.
12. P. SERRE, "Cohérence de la spécification d'un objet de l'espace euclidien à n dimensions", PhD thesis, Ecole centrale de Paris, Laboratoire d'Ingénierie Intégrée des Systèmes Mécaniques et des Matériaux, ISMCM, France, 2000.

INERTIAL TOLERANCING IN THE CASE OF ASSEMBLED PRODUCTS

M. Pillet
LISTIC, Université de Savoie, Annecy, France

Abstract: In the case of assembled products, tolerancing is a key factor for quality and reliability. Traditionally, tolerances are defined by an interval [LSL; USL]. Several approaches have been proposed to determine this tolerance, the main results of which be referred. Inertial Tolerancing consists in tolerancing the mean square deviation from the target rather than the distance. This alternative has numerous advantages over the traditional approach, particularly in the case of product assembly, mixed batches and conformity analysis. In this paper, we will detail "inertial tolerancing" and compare both approaches: traditional and inertial tolerancing.

Key words: Statistical Tolerancing, Worst case Tolerancing, Inertial Tolerancing, Capability, Taguchi Loss Function.

1. INTRODUCTION

In the case of assembled product, tolerancing is a compromise between two situations: worst case tolerancing and statistical tolerancing [Chase 1991]. Worst case tolerancing guarantees the specification at the final level of assembly if all the elementary characteristics satisfy the specification limits. Statistical tolerancing takes into account the low probability of having several characteristics close to the specification limit at the same time [Shewhart 1931] and to increase the specification limits in order to reduce production cost. Both of these situations have some well-known disadvantages (cost of the production for the worst case and risk of non-quality for statistical tolerancing). Several contributions have proposed a

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 85–94.

compromise between worst case and statistical tolerancing such as inflated statistical tolerancing [Graves 1997]. Another response consists in defining the conformity by the capability index Cpm [Chan 1988]. All of these approaches are founded on the traditional representation for tolerance: the interval [LSL; USL]. In this paper, we suggest investigating another way with a new representation for tolerance: inertial tolerance. This representation consists in tolerancing the square deviation from the target and offers numerous advantages. In the case of the assembled product, we will see that this representation makes the best compromise between worst case and statistical tolerancing.

2. THE VARIOUS APPROACHES OF TOLERANCING FOR ASSEMBLED PRODUCTS

The problem consists in determining the tolerances on the elementary characteristics x_i to obtain a final characteristic Y satisfying the need of the customers. As a general rule, when one works on the neighborhood of the target, a first order linear approximation is generally sufficient to study the behavior of the system. One can characterize the behavior of Y by equation 1:

$$Y = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i \qquad (1)$$

α_i represents the coefficient of the influence of x_i in Y. The problem of tolerance consists in trying to reconcile two antagonistic preoccupations:

- Fix tolerances as widely as possible to decrease production costs.
- Obtain an optimal quality level on the characteristic Y.

Let us go over the main methods of tolerancing in the case of a linear relation.

2.1 Worst case tolerancing

In this case, one considers that the final condition Y will be respected in all cases of assembling. One note $t_{xi} \pm \Delta_{xi}$ the worst case tolerancing on x_i. In the case of a linear relation (1) we have:

$$t_y = \sum \alpha_i t_{xi} \quad ; \quad \Delta_y = \sum |\alpha_i| \Delta_{xi} \qquad (2)(3)$$

The tolerance sharing can be carried out using various methods [Graves 2001].

- Uniform tolerance sharing.
- Consideration of standard or design rules.

- Proportional to the square root of the nominal dimension.
- Consideration of the capability experience.

The well-known disadvantage of worst case tolerancing is the high cost associated with this method. Indeed, it leads to very tight tolerances, which are very difficult to obtain during production. Consequently, the cost of inspection, rejects and reworking increases as does the choice of a more sophisticated product method. The main advantage is the guarantee of the specification at the final level of assembly.

2.2 Statistical tolerancing

Statistical tolerancing was developed in order to take into account the low probability of having several characteristics close to the specification at the same time [Shewhart 1931]. From equation 1, where the variables x_i are independent with a standard deviation of σ_i, one has the relationship (4)

$$\sigma_Y = \sqrt{\sum \alpha_i^2 \sigma_i^2} \tag{4}$$

With tolerances proportional to the standard deviation, we obtain [Chase 1991] equation (5):

$$T_Y = \sqrt{\sum_{i=1}^{n} \alpha_i^2 T_i^2} \tag{5}$$

Statistical tolerancing assumes centered processes. Violation of this assumption can result in a defective product. Many authors showed the disadvantages of this type of method. In particular Graves and Bisgaard identified 5 major risks [Graves 2000]:

1. Inadequate functional model.
2. Desire higher capability in assembly than delivered for components.
3. The process mean is off-center.
4. Correlation between component characteristics.
5. Component dimensions are not be normally distributed

Several methods were proposed in order to counter the negative aspects of statistical tolerancing. The main one is inflated statistical tolerancing [Graves 1997][Graves 2000] [Graves 2001]. But, despite the high quality of this tolerancing method, it is still possible to find a situation in which the method will not be appropriate.

3. INERTIAL TOLERANCING

3.1 Définition

The goal of tolerancing consists in determining a criterion of acceptance for the elementary characteristic x_i guaranteeing the acceptance of the resulting characteristic Y. The tolerance limits the cost of non-quality generated by a variation compared to this target. For a good design, when Y is placed on the target, quality is robust compared to the operating and environmental conditions. When Y moves away from the target, the quality will be increasingly sensitive to the conditions, and could lead to the customer being dissatisfied. Taguchi [Taguchi 1987] has showed that the financial loss associated (L) with a shift to the target was proportional (coefficient K) to the square of the variation.

$$L = K(Y_i - cible)^2 \tag{6}$$

For a batch with an average μ, the associated loss is:

$$L = K(\sigma_Y^2 + (\mu - \text{target})^2) = K(\sigma_Y^2 + \delta_Y^2) \tag{7}$$

K is a constant, and $I_Y = (\sigma_Y^2 + \delta_Y^2)$ is the mean square deviation (MSD) of the target. Mechanically speaking, the mean square deviation behaves like inertia. We have the choice to call this term I_Y. In order to have the minimal "inertia" (and the minimal loss), we must have the products as near to the target as possible. As with inertia, the mean square deviation has the very interesting property of additivity in the case of the linear relationship between Y and x_i. To take advantage of this property, we propose to replace the traditional tolerance $Y \pm \Delta Y$ by the tolerance $Y(I_Y)$ in which I_Y represents the maximum mean square deviation accepted on variable Y. By analogy with the MSD with inertia, we have chosen to name this alternative method "Inertial Tolerancing". This way of determining the tolerances has interesting properties that we propose to develop.

3.2 Conformity in inertial tolerancing

With the traditional tolerance representation, the conformity decision for a component is not easy in the case of assembled products. The decision concerns one component, but the product quality will be the combination of several components. When a characteristic is near the specification limit, the decision about conformity must take several known or unknown elements into account:

- Size of the series: Indeed in the case of a unit production the probability of assembling a part to the extremes is more significant than in the case of mass production.
- The distribution of the characteristics: The decision should not be the same if the part nearest the specification limit is alone or if 100% of the production is also near the specification limit.
- The distribution of the other important characteristics of the assembly. In general, these elements are unknown at the time the decision is made.

Inertial tolerancing makes it possible to partly take these various elements into account. Contrary to traditional methods, the goal is not to obtain a level of quality measured by a percentage outside tolerance, but to guarantee low inertia around the target. The decision is not founded on the percentage outside tolerance, but rather on the inertia. Normality is not a necessary criterion.

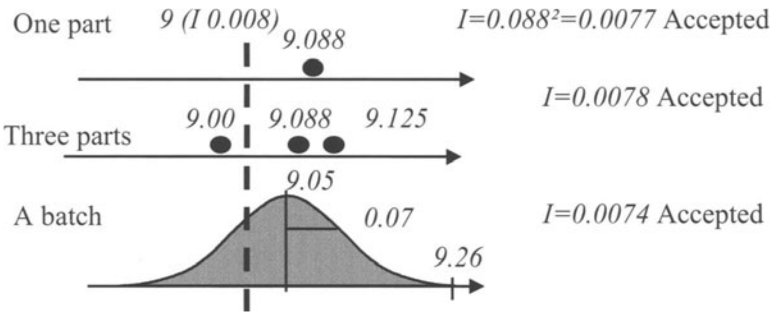


Figure -1. Conformity and inertial tolerancing

The inertial tolerancing leads to "fuzzy tolerances" that vary according to the quantity of parts produced. To illustrate this point, let us imagine a characteristic *9 (I 0.008)* of an academic case Figure 1.

- Case of one part measured at *9.088*. In this case, the inertia for a part is equal to $0.088^2=0.0077$. The part is only just accepted.
- Now let us take the case of 3 parts: *9.00; 9.088; 9.125*. Inertia is then $I=0.0078$, the three parts are accepted. One part has an individual inertia higher than *0.008*.
- Finally let us take the case of a batch of parts averaging *9.05* and with a standard deviation of *0.07*. In this case $I=0.0074$ inertia; the batch is also accepted even though it contains *28.6 %* produced whose dimension is higher than *9.0894* which is the approval limit in the case of unit production.

This "fuzzy tolerance" is a particularly interesting point of inertial tolerance. In the case of unit production, it guarantees the perfect conformity of the production in the worst of cases. In the case of series production, the inertial tolerance takes into account of the weak probability that both extremes of the tolerance limits will be assembled together. Thus, with only one tolerance, one is equally able to respond to cases of unit production as to those in series production.

3.3 Extreme situations of acceptance in the case of inertial tolerance

Let us study the extreme situations of acceptance in the case of inertial tolerancing. Inertia can increase under the influence of two parameters: a shift of the average δ_Y compared to the target or an increase in spread (σ_Y) around the average regardless the form of distribution.

- Centered extreme situation *($\delta_Y = 0$)*

$$I_Y = \sigma_Y^2 \quad \text{one thus obtains: } \sigma_{YMax} = \sqrt{I_Y} \qquad (8)(9)$$

- Extreme situation with a standard deviation equal to zero *($\sigma_Y = 0$)*

$$I_Y = \delta_Y^2 \quad \text{one hus obtains: } \delta_{YMax} = \sqrt{I_Y} \qquad (10)(11)$$

- Extreme µ shift accepted according to standard deviation

In the case of inertial tolerancing, accepted extreme shift will be a function of the standard deviation.

$$\delta / \sqrt{I_Y} = \sqrt{1 - \left(\sigma_Y / \sqrt{I_Y}\right)^2} \qquad (12)$$

Up to $\sigma < 0.6\sqrt{I}$, one can have a relatively significant shift $\delta < 0.8\sqrt{I}$. Similarly, as long as the shift is low, one can accept a relatively significant spread.

3.4 An adapted capability index (Cpi)

In the case of inertial tolerancing, conformity is declared if the inertia I_Y of the characteristic - or the batch – is lower than the desired maximum inertia I_{YMax}. It is possible to define a capability indicator *Cpi* by the equation. (13)

$$Cpi = \frac{I_{YMax}}{I_y} \qquad (13)$$

A manufacturing process must obtain *Cpi* higher than *1*. Later in this paper, we will show that this indicator has many properties, in particular, in the case of mixed batches. The *Cpi* of two mixed batches is equal to the average of both *Cpi*.

4. PROPERTY OF ADDITIVITY OF INERTIAS

4.1 Case of an assembled product

The properties of inertial tolerancing come from the property of additivity of the mean square deviation. To illustrate this property, we will take the case of a resulting characteristic *Y* dependent on a linear function of several elementary characteristics (equation 1)

In this case, it is easy to show that we have:

$$\sigma_Y^2 = \sum \alpha_i^2 \sigma_i^2 \qquad \delta_Y = \sum \alpha_i \delta_i \tag{14)(15}$$

Let us calculate inertia obtained on the characteristic *Y*.

$$I_Y = \sigma_Y^2 + \delta_Y^2 = \sum \alpha_i^2 \sigma_i^2 + \left(\sum \alpha_i \delta_i\right)^2$$

Which is written as follows:

$$I_Y = \sum \alpha_i^2 I_{Xi} + 2\sum \alpha_i \alpha_j \delta_i \delta_j \tag{16}$$

The first part of the equation corresponds to the additivity of different inertia. The double product corresponds to the case where all shifts are on the same side. In the case of random distribution of the averages when the component count is significant, one can consider that this double product is equal to zero. We have a hypothesis here that is close to the assumption of traditional statistical tolerancing. On the other hand, in cases where the number of components is not significant or when the average distribution is not random, it is possible to integrate the double product in the distribution of inertias. From equation (16), we can calculate the distribution of inertias of each elementary characteristic according to the inertia desired on the resulting characteristic. This equation makes it possible - if desired - to integrate the risk of unfavorable systematic decentering of all the elements. This equation easily makes it possible to determine the tolerances on the elementary characteristics according to four hypotheses:

- **Assumption 1:** Worst case

In these conditions, we saw (equation 11) that maximum decentering was $\delta_{YMax} = \sqrt{I_Y}$. In the worst case, inertia will be equal to:

$$I_{YMax} = \sum \alpha_i^2 I_{iMax} + 2\sum \alpha_i \alpha_j \sqrt{I_{iMax}} \sqrt{I_{jMax}} \tag{17}$$

Equation (17) is simplified in the case where $\alpha_i = 1$ and in the uniform distribution of inertia $(I_{iMax} = I_{Max})$ on each characteristic.

$$I_{Max} = I_{YMax} / n^2 \tag{18}$$

- **Assumption 2:** Random distribution of the averages.

In this hypothesis, the double product is equal to zero. Equation 17 is simplified where $\alpha i = 1$ and in the uniform distribution of inertia $(I_{iMax}=I_{Max})$ on each characteristic

$$I_Y = nI_{Max} \qquad I_{Max} = I_Y / n \tag{19}$$

- **Assumption 3:** Unfavorable average shifts of *m* components among *n*.

In this hypothesis, the designer determines the number of characteristics which can have a systematic process bias. Under these conditions the weight of the double product is decreased. Where $\alpha i = 1$ and in the uniform distribution of inertia $(I_{iMax}=I_{Max})$ on each characteristic, we can write:

$$I_X = \frac{I_{YMax}(1+k^2)}{n(1+k^2)+mk^2(m-1)} \tag{20}$$

Among the 3 assumptions, the best choice depends on the result desired on *Y*. To obtain a good inertia on *Y*, assumption 3 offers the best compromise. However, in most cases, one will use assumptions 2 which offers the best compromise between the traditional tolerancing methods (worst of the cases and statistics).

4.2 Application on the ten washers example

As an example [Bisgaard 1997] of tolerance stack-up, suppose we need to make an assembly consisting of 10 steel washers stacked on top of each other. Further, suppose each washer is required to be nominally *1 mm* thick and the assembly $Y=10\pm 0.1$ *mm*. If, as is often assumed the tolerance range is $\pm 3\sigma$, we will require that $\sigma_Y = 0.1/3$. In the case of traditional tolerancing we can write: Worst case tolerancing: $X = 1\pm 0.01$; Statistical tolerancing: $X=1\pm 0.032$.

In the case of inertial tolerancing, the objective is to guarantee production centered on *10* with a standard deviation of *0.1/3*. That corresponds to an inertia of $I_Y=(0.1/3)^2=0.01/9$. In assumption 2 (random distribution of the averages), it is easy to calculate inertia on each washer: $I_X=I_Y/10=0.01/90$ which corresponds to a standard deviation $\sigma=0.0105$ in the case of a centered population. Bisgaard sees what happens if we aim $\delta=0.01mm$ off target for the individual washers in the case of statistical tolerancing. Based on the normal distribution assumption, the probability of being outside the specification limits is *2% [Cpk(X)=0.68]*. After the 10-washer assembly, we find *50%* defective parts on the Y characteristics *[Cpk(Y)=0]*. Two different situations can lead to defective parts on *Y*: an increase in spread and decentering. One can retain two extreme situations:

- The mean is on the target and the spread is maximal

- The standard deviation is minimal and decentering is maximal

Table -1. Extreme situations

	$\delta_{Xi}=0$ situation		$\sigma_{Xi} \approx 0$ situation	
situation	elementary characteristics X_i	dependant characteristic Y	elementary characteristics X_i	Dependant characteristic Y

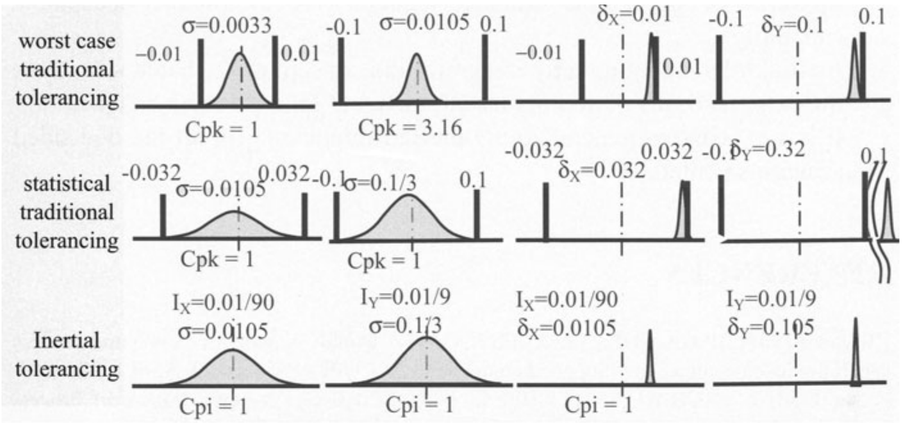


Table 1 illustrates these two extreme cases on x and the consequences on Y in the three cases of tolerancing method: traditional worst case, traditional statistical and inertial with the assumption 2. In this example, inertial tolerance allows to obtain the ideal compromise between both traditional tolerancing methods. If the mean is on the target, inertia means a standard deviation identical to the case of the traditional statistical tolerancing can be accepted. In the case of a standard deviation equal to zero, maximum decentering permitted by inertial tolerancing almost identical to that is allowed by the traditional "worst case" method. In the latter case, the inertia is not respected on the characteristic Y, but the functional condition is satisfied.

5. CONCLUSION

Inertial Tolerancing proposed in this communication offers numerous advantages with regard to the traditional tolerancing methods. The main ones are the following:

- Inertial tolerancing takes into account the combinatory aspect in assembled products. It is rarely one characteristic that is responsible for non-quality, but often the unfavorable combination of several characteristics.

- For an assembled product, the inertial tolerancing leads to a better compromise between the cost of production and quality as opposed to the traditional approaches.
- The properties of additivity make it possible to define an additive process capability index Cpi. That makes it possible to carry out mixtures of acceptable batches without risk of final unacceptable capability.
- Inertial tolerancing directly integrates the concept of the batch size; thus, the extreme limits of acceptance are different for a part or a batch.
- It is very easy to generalize the inertial tolerancing in all the one-sided scenarios [Pillet 2001]

REFERENCES

[Bisgaard 1997] BISGAARD S. GRAVES S. (1997) - Quality quandaries. A negative process capability index from assembling good components? – CQPI Report n°160, April 1997

[Chan 1988] L. J. CHAN, S. W. CHENG, F. A. SPIRING, "A new Measure of Process Capability: Cpm" - Journal of Quality Technology, Vol 20, N° 3, 1988, pp 162-175.

[Chase 1991] K.W. CHASE ; A. R. PRAKINSON- A survey of rechearch in the application of tolerance analysis to the design of mechanical assemblies – Research in Engineering design (1991) 3:23-37

[Graves 1997] S. GRAVES. How to reduce costs using a tolerance analysis formula tailored to your organisation – CQPI Report n° 157 - 1997

[Graves 2000] GRAVES S., BISGAARD S. (2000) - Five ways statistical tolerancing can fail and what to do about them – Quality engineering 13, pp 85-93.

[Graves 2001] GRAVES S. (2001) – Tolerance Analysis Tailored to your organization – Journal of Quality technology – Vol. 33, N°3, 293-303, July 2001

[Pillet 2001] M. PILLET, F. BERNARD, L. AVRILLON(2001) - Le tolérancement inertiel, une autre façon d'intégrer l'aspect combinatoire dans les processus assemblés – Congrès CPI 2001 – Fès Maroc

[Shewhart 1931] W. A. SCHEWART Economic Control of Quality of Manufactured Product. Van Nostrand, New York, 1931

[Taguchi 1987] G. TAGUCHI, System of experimental design, Volumes 1 et 2, American Supplier Institute, INC, 1987

AN INTERACTIVE DECISION SUPPORT SYSTEM FOR UNIDIRECTIONAL TOLERANCING

C. Forster, J–P. Boufflet, S. Hadji and S. Bouabdallah
Université de Technologie de Compiègne, France
Pôle Universitaire Léonard de Vinci, France

Abstract: We propose in this paper the premises of an Interactive Decision Support System dedicated to the unidirectional tolerancing of mechanical assemblies. The foundation of our work is an efficient representation of a mechanical assembly based on graph theory. From this model, we can determine all the configurations of an assembly. An extended syntax for the functional constraints permits then to generate automatically all the tolerance chains. The final result is an automatically generated linear system of equations and inequations expressing the tolerance chains, as well as the existence of the configurations and coherence of the dimensioning scheme. This system can then be solved by an analysis or a synthesis approach.

Key words: Computer-aided tolerancing. Unidirectional tolerancing. Tolerance chains.

1. INTRODUCTION

Tolerancing is a very complex stage of the design process and is subject to intensive research works. Many issues have to be addressed concerning:

a) the representation of the geometrical errors of mechanical parts;
b) the consequences of the adopted tolerances in terms of assembling and functionality (i.e. tolerance analysis);
c) the determination of the geometrical requirements from the functional constraints (i.e. tolerance allocation or synthesis);
d) the link with manufacturing and metrology;
e) automation.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 95–104.

A comprehensive treatment of the tolerancing may be found in the books [1], [2]. For a more complete review on automation, see [3] and [4].

In the current work we focus on the unidirectional tolerancing and our aim is to develop an Interactive Decision Support System (IDSS) approach, following the previous works presented in [5] to [7]. Its purposes are threefold:

1. solve efficiently the cases of unidirectional tolerancing (pre-study of complex mechanical systems or machine-design);
2. provide designers with an helpful framework (geometrical specifications by tolerance zones in three independent directions);
3. prepare automatic tolerancing methods for the two- and three-dimensional cases.

The proposed approach attempts to formalize and automate the manual method. In section 1, we briefly recall the graph model we associate with a mechanical assembly [9]. Then we determine all the configurations of the studied assembly. Recent papers have also explored the generation of the configurations by applying another strategy [8]. Section 3 is dedicated to tolerance chains construction. We first propose an extended syntax for the functional constraints so as to unambiguously interpret them. Next, the associated tolerance chains are built and give the essential individual dimensions for each part. In section 5, we discuss the dimensioning schemes. As practiced by some designers, we introduce the possibility of defining a new dimensioning scheme, including only manufactured dimensions, and to construct the corresponding tolerances chains. Finally, in section 6, we present all the relations the data and the unknowns must fulfill. At this stage, one can achieve the analysis or the synthesis (allocation) of the tolerances.

2. GRAPH MODEL

Let us consider a mechanical assembly, from which parts are assumed geometrically perfect and perfectly rigid. In unidirectional tolerancing, it may be represented by a view or a section where geometries are described by closed curves.

According to eventual clearances, parts can move and define new configurations of the assembly. In the proposed approach, the designer has to set the pairs of surfaces in contact or able to come into contact. The nature of each contact has also to be set. Three types of contacts are distinguished:

a) permanent contact (CP), for two fixed surfaces;
b) closed contact (CF), for two surfaces in contact but able to be separated;
c) opened contact (CO), for two separated surfaces able to come into contact.

Here we consider only mechanical assemblies with well-positioned parts: each part is in contact with at least another one, and two parts are in contact only through two surfaces. In addition, we assume that there is no redundancy: it is not possible to find a set of parts defining a loop of closed contacts. This would be physically unacceptable for rigid parts.

These data can easily be entered using a two-dimensional sketcher. They permit us to localize clearances, or to define the contact zones between the surfaces. Figure 1a represents the data relative to the treated example, as displayed by our prototype software (developed under Windows™ with C++): it contains three parts and five functional constraints. The opened padlocks (respectively closed) symbolize the opened contacts (respectively closed).

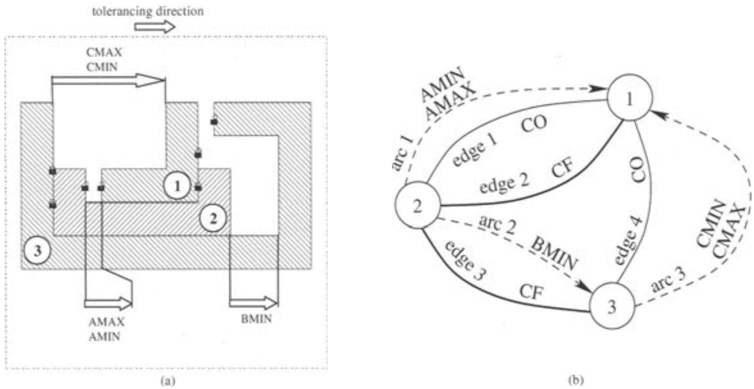


Figure –1. (a) Treated example. (b) Associated part graph: vertices represent parts, thin and bold edges represent contacts between parts while dotted arcs represent functional constraints.

The mechanical assembly is then modeled by a graph, denoted the part-graph. The vertices are the parts and two vertices are linked by an edge if the parts are in contact, or may come into contact.

For an assembly where all the parts are well-positioned, the part-graph is a connected non oriented 2-graph. Each edge is arbitrary numbered and has two values: the pair of the concerned surfaces and the type of contact $\xi \in \{CO, CF, CP\}$. In the part-graph of figure 1b, edges and vertices model the mechanical assembly of figure 1a. The edges express the contact between two parts, but also the ability to come into contact. As a consequence, we can obtain all the configurations by changing the values ξ associated with the edges (i.e. sliding the parts).

A configuration corresponds to a set of closed contacts between parts. It can be modeled by the partial graph obtained by suppressing all the edges of type of contact CO in the part-graph (CO-valued edges). For instance, the initial configuration of figure 1a is modeled by the partial graph extracted from the graph of figure 1b where only the CF-valued edges are considered (i.e. the bold edges).

There are as many partial graphs as configurations contained in the part-graph. For a configuration with well-positioned parts, the partial graph is a connected simple graph. In the absence of redundancy, there is no cycle. So the partial graph is a tree.

By sliding parts, we change contacts between parts, and new configurations are obtained. For determining all the configurations of the assembly, we use the following property: the number of CO-valued edges (resp. CF) is a constant along the configurations. The proof of this property is directly based on the fact that the partial graph is a tree (the reader should consult [9] for more details).

Signed distances (relative to the tolerancing direction) between two surfaces in a given configuration are computed using classical algorithms of graph theory on the partial graph.

In order to take into account the functional constraints we supplement the part-graph. For each functional constraint, we add an arc between the concerned parts. Each new arc is valued by the concerned parts and surfaces, its type and its attribute (see section 4). At this stage the graph contains edges and arcs as shown in figure 1b: it is oriented and non oriented at the same time. These arcs are often associated with some CO-valued edges of the part-graph. This is logical, because when building the assembly, the designer tries to make the clearances clearly apparent, in order to impose the functional constraints.

3. RESEARCHING ALL THE CONFIGURATIONS

The fact that the partial graph is a tree is the foundation of the algorithm for the determination of all the configurations. Let m_{CO} be the number of CO-valued edges and m_{CF} the number of CF-valued edges. We have:

$C^{m_{CO}}_{m_{CF}+m_{CO}}$ possible combinations changing a value CF into CO while keeping m_{CO} and m_{CF} constant.

Unfortunately, each combination does not necessarily correspond to a valid configuration, with well-positioned parts and no redundancy. A combination defines a valid configuration if the associated partial graph is a tree. Moreover, in this combinatorial process, an opened contact can be defined between two surfaces of two parts which have moved, such that physically the surfaces collide. It is necessary to check that the distance between surfaces connected by a CO-valued edge is strictly positive.

So, the proposed algorithm consists in first generating all the combinations, then checking if each combination is valid, i.e. the partial graph is a tree and the distance along the CO-valued edges is positive. It is worth noting that the initial configuration given by the designer has to be checked first.

Table -1: Possible Combinations.

Edge	I	II	III	IV	V	VI
1	CO	CO	CO	CF	CF	CF
2	CO	CF	CF	CO	CO	CF
3	CF	CO	CF	CO	CF	CO
4	CF	CF	CO	CF	CO	CO

The proposed example has $m_{CO} = 2$ and $m_{CF} = 2$, and the initial configuration is valid. The six obtained combinations are listed in table 1. The initial configuration corresponds to column III. By applying the checking rules, we obtain the four valid configurations of the mechanical assembly presented in figure 2 (combinations II, III, IV and V of table 1). Combination I leads to an inter-penetration between parts 1 and 2. In combination VI, the part 3 is “floating”.

4. THE TOLERANCE CHAINS DETERMINATION

The syntax of a functional constraint gives the following explicit information: the two concerned parts and their associated surfaces, the type(s) (i.e. minimum and/or maximum), and the value(s) of the distance(s) that should exist between the two surfaces. But a functional constraint also contains an implicit meaning: the extreme position of the surfaces to be taken into account. The difficulty in automating unidirectional tolerancing comes from the fact that it is not always possible to indicate the functional constraint in the relevant configuration, where parts and surfaces have the right position.

In order to unambiguously interpret the functional constraints, we propose to add an attribute in their definition which describes the right position to consider and has three possible values:

1. “actual”: the surfaces are in the position in the current configuration;
2. “far”: the surfaces are as far apart as possible;
3. “close”: the surfaces are as near as possible.

It is worth noting that the expression “as … as possible” depends on all the configurations of the mechanical assembly. Therefore, in order to improve the semantics for a functional constraint, we introduce the four following interpretations:

1. maximum “close”: to impose between two parts not allowed to come into contact;
2. maximum “far”: to limit the largest distance between parts;
3. minimum “close”: to limit the smallest distance between two parts not allowed to come into contact;
4. minimum “far”: to define the minimal distance when the two parts are far apart.

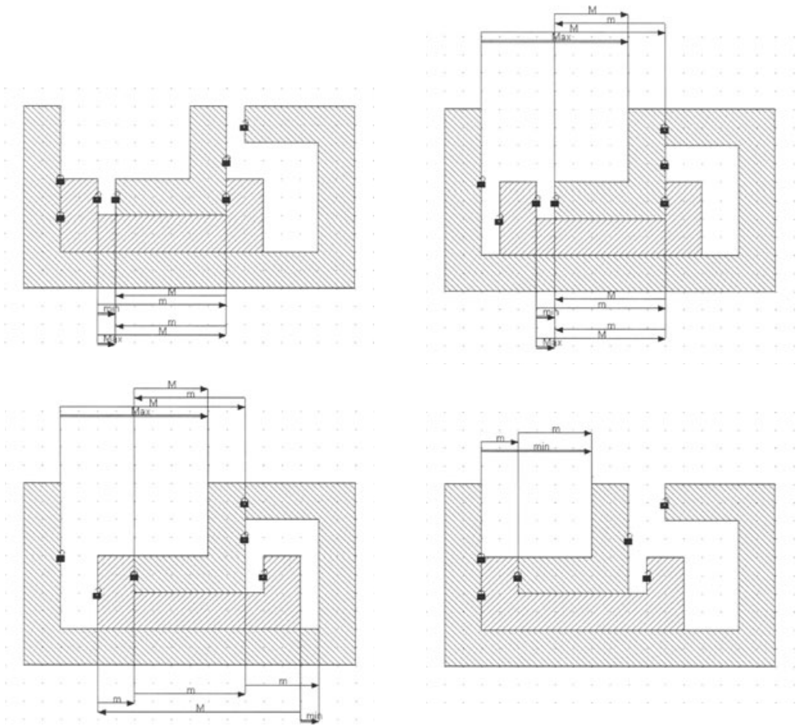


Figure –2. Valid configurations and associated tolerance chains.

These semantic interpretations of functional constraints correspond to actual situations. On the part-graph of figure 1b, AMIN corresponds to an uni-limit minimum “far” functional constraint while AMAX is an uni-limit maximum “far” one. BMIN corresponds to an uni-limit minimum “close” one. CMIN is an uni-limit minimum “close” functional constraint while CMAX is an uni-limit maximum “far” one.

Sometimes, the designer cannot impose the specifications using one of the former interpretations. In such a case, the “actual” attribute has to be used on the relevant configuration.

Following the tolerancing process, we now have to build the tolerance chains associated with each functional constraint. Among all the valid configurations of the mechanical assembly found in section 3, we consider those where the surfaces associated with the functional constraint are in "good" position (relative to the "far" or "close" attribute). This requires firstly computing the distance between the surfaces for each configuration, then selecting those for which the distance is minimum ("close") or maximum ("far").

For each functional constraint one has the initial surface, the final surface and the unique walk linking them in the partial graphs of the associated configurations. By grouping the surfaces belonging to each part of the walk it is now possible to define the individual dimensions of the parts that occur in the given functional constraint.

The nominal value of these individual dimensions is given by the absolute value of the difference of the abscises of the surfaces, measured in the tolerancing direction. Moreover, the order of appearance of the surfaces of a part in the previous walk gives the direction of the individual dimension between these two surfaces or, more usefully, the sign of the individual dimension in the algebraic tolerance chain.

In our example, we finally obtain the tolerance chains presented in figure 2, for the four valid configurations.

5. DIMENSIONING SCHEMES

The individual dimensions that come from the tolerance chains directly describe the functionality of the mechanical assembly. However, they present drawbacks concerning:

a) the dimensioning scheme;
b) the manufacturing process.

They can be incomplete because insufficient functional constraints have been imposed. At the same time, they also can be redundant. This can confuse the manufacturer even though it is not necessarily a problem for the designer. Moreover, some individual dimensions cannot be manufactured directly.

Extended companies prefer to keep the initial dimensioning scheme, because parts are often realized outside with unknown tool machines. However some integrated companies prefer to establish a dimensioning scheme corresponding more directly to the manufacturing process (with "industrial" or "manufactured" or "simulated" dimensions). This is convenient to ensure the feasibility of all the dimensions, and to impose associated tolerances in direct connection with the capabilities of the available tool machines.

In such a case, the designer has to propose a new dimensioning scheme:

a) complete, to define totally each part;
b) not redundant, to make the resolution and the manufacturing processes easier;
c) using the initial individual dimensions as much as possible;
d) established in collaboration with the manufacturer, to avoid auxiliary dimensions calculations;
e) bi-limiting in general, to facilitate the manufacturing and the resolution processes.

Roughly speaking, for each part, one has to position once and for all a surface relative to the other surfaces. Let n_S be the number of surfaces of the part. We have to define n_S-1 dimensions among $C^2_{n_s-1}$ possible combinations linking two surfaces, while avoiding redundancy and forgotten surfaces.

In order to help the designer to make a valid choice among all these combinations, we introduce for each part a new non-oriented graph, called a surface-graph. The vertices are the n_S surfaces and the edges represent the "industrial" dimensions. This graph must verify the following properties:

a) it is connected to ensure the completeness of the dimensioning scheme;
b) it has no cycle to avoid redundancy.

The proposed dimensioning scheme being validated by testing the properties of each surface-graph, one has to build new tolerance chains taking into account only the "industrial" dimensions. To achieve this, the previous method relative to the tolerance chains is used. But instead of considering the partial graph of the relevant configuration, a new partial graph is constructed, by replacing the parts with the surface-graphs. This graph is also a tree, so a unique walk exists between the two extreme surfaces of a functional constraint.

6. TOLERANCE ANALYSIS AND SYNTHESIS

At this stage, we have defined the nominal geometry and the contacts, imposed the functional constraints, and built the tolerance chains. If necessary, an "industrial" dimensioning scheme has been chosen leading to new tolerance chains. Solving the tolerancing problem consists in determining the minimum and/or the maximum values of the dimensions occurring in the problem ("worst" case).

The mathematical relations, automatically generated during the previous steps, come from:

a) tolerances chains;
b) non-penetration constraints along the CO-valued edges for all the configurations;
c) coherence constraints involved by the “industrial” dimensioning scheme.

The tolerances chains lead to a linear system of equations linking the functional constraints and the individual (or “industrial”) dimensions. But this usual presentation of the tolerancing problem has to be completed.

During the search of all the configurations, we check that there is no contact between surfaces connected by a CO-valued edge by using the nominal geometry. However, due to the variations of the dimensions during the resolution process, it is possible that this condition becomes violated. It is therefore necessary to impose additional functional constraints expressing the existence of the configurations in the vicinity of the nominal.

In addition, when dimensioning the parts and to avoid redundancy, some neighboring surfaces are not linked by a dimension. The distance between these two surfaces is deduced from the other dimensions of the part. When these dimensions change in the resolution process, the previous distance can become negative, violating the topology of the part. So the dimensions of a given part must fulfill inequalities expressing the integrity of its topology.

In the case of the tolerance analysis, the designer chooses minimal and maximal values for all the unknown dimensions. Then, the resulting values of the functional constraints and the resulting clearances between surfaces are computed while preserving the coherence of the dimensioning scheme. This method is implemented in our software, and permits, by a predictor/corrector process, the assignment of values to the dimensions, or if necessary, the modification of the values of the functional constraints.

In the case of tolerance synthesis, the designer imposes values for the functional constraints and has to solve the previous linear system of equations and inequations. In general there are more unknowns than relations. Equations and inequations may be linearly dependent. They do not have the same importance, due to the values of the functional constraints and/or the number of unknowns. Moreover, the values of the functional constraints do not have the same reliability. For instance, the tolerance of the gap in a joint is generally smaller and more rigorous than the tolerance of a distance to an obstacle. Finally, the designer expects results close to the nominal values he has introduced.

As a consequence, the resolution of this problem is complex and may require the simultaneous use of resolution algorithms for linear systems, artificial intelligence (expert system, fuzzy logic) and optimization. In our software, at the moment we only use a linear programming algorithm, minimizing the distance between the computed dimensions and the nominal dimensions for the tolerance synthesis.

7. CONCLUSION

The basis of an Interactive Decision Support System dedicated to the unidirectional tolerancing of mechanical assemblies has been presented. Prototype software has been developed in order to demonstrate the relevance and the efficiency of the proposed approach. The preliminary results obtained on some mechanical studies have to be confirmed on a wider benchmark. The IDDS proposed provides a framework for extending the present approach and testing new ideas. We are investigating the resolution problem with artificial intelligence and/or classical optimization approaches.

REFERENCES

1. A. Clément, A. Rivière, M. Temmerman, "Cotation tridimensionnelle des systèmes mécaniques", PYC Edition, ISBN: 2-85330-132-X, 1994.
2. P. Bourdet, L. Mathieu, "Tolérancement et métrologie industrielle", CETIM, ISBN: 2-85400-440-X, 1998.
3. U. Roy, C. R. Liu, T. C. Woo, "Review of dimensioning and tolerancing: representation and processing", Computer-Aided Design, Vol. 23, N°7, 1991, pp. 466-483.
4. K. M. Yu, S. T. Tan, M. F. Yuen, "A review of automatic dimensioning and tolerancing schemes", Engineering with Computers, Vol. 10, 1994, pp. 63-80.
5. R. Södeberg, "CATI: a computer aided tolerancing interface", Advances in Design Automation - ASME 1992, DE-Vol. 44-2, , Vol. 2, 1992, pp. 157-164.
6. M. N. Mliki, "Contribution à l'étude et à la réalisation d'un système de cotation fonctionnelle assistée par ordinateur", PHD Thesis, Marseille III, 1992.
7. B. Ramani, S. H. Cheraghi, J. M. Twomey, "CAD-based integrated tolerancing system", Int. J. Prod. Res., Vol. 36, N° 10, 1998, pp. 2891-2910.
8. E. Ballot, P. Bourdet, F. Thiébaut, "Determination of relative situations of parts for tolerance computation", 7^{th} CIRP International Seminar on Computed-Aided Tolerancing, Cachan, France, 2001.
9. Ch. Forster, J.-P. Boufflet, F. Lecouvreur, "Automatic determination of tolerance chains in unidirectional tolerancing", 26^{th} ASME Design Automation Conference, Baltimore, Maryland, USA, 2000

IMPROVEMENT OF THE SIMULATION TOOLS DEDICATED TO COMPOSITE STRUCTURES SUBJECTED TO CRASH LOADS

P. Rozycki and D. Coutellier
LAMIH, University of Valenciennes and Hainaut-Cambrésis, France

Abstract: In the framework of transport industries, vehicle design optimisation represents a key priority issue. Indeed, this is a result of the huge pressures imposed on industrialists particularly in terms of active safety. Moreover, among of all the materials used for design, some such as composite materials require special attention. It is consequently within this framework that we propose to describe in this paper the contributions which were made to the numerical simulation tools for structural analysis. The latter can involve crash type loading for example. In a first instance, certain behavioural phenomena and their integration into the computational code will be described. Secondly, some elementary validations will be presented. Finally, some applications relating to the dimensioning of panels in composites subjected to linear or point impacts will be revealed. The conclusions of these studies will enable us to show, on the one hand, a certain reliability of the developments which permit their use as design tools but also, on the other hand, to foresee the prospects of future research.

Key words: Composite materials, numerical simulations, design, crash

1. INTRODUCTION

Structural design optimisation in the transport industries highlights an increasingly high use of data-processing tools. Indeed, computation software allows industrialists to meet the requirements induced both by the economic and ecological contexts, such as cutting manufacturing costs and time. Nowadays, they are completely integrated within the various certification, standardisation or safety processes whose drastic criteria impose a constant increase in the structure's reliability and viability.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 105–116.

From this point of view, our research team has developed, in close relations with industrialists, a new finite element within an explicit code. Its main specificity is to come up to the dimensioning of multi-layered multi-material structures subjected to crash loads [1] [2] [3] [4]. These materials are a combination of metal and fibre/resin composite layers and they are devoted to the design of certain aircraft fuselages.

However, the purely composite laminated materials make significantly great strides, essentially due to their mechanical characteristic properties: by including them into some structural parts, it is possible to obtain responses for extreme requests. Consequently, it appears necessary to optimise the means suggested by the computer codes so that these new tools facilitate the design of these new structures for the industrialists.

The aim of this article is to present the improvements brought about to the previous finite element with regard to fibre/resin composite plies. Therefore, some general background information on the finite element is initially presented. An experimental quantisation of strain rate sensitivity for unidirectional composite materials is then tackled. This part continues by a theoretical approach adapted to the description of viscous phenomena. Finally, some simple and complex validation cases are presented which allow to conclude on the future prospects generated by this study.

2. THE MULTI-LAYERED MULTI-MATERIAL FINITE ELEMENT

2.1 Generality

The multi-layered multi-material element [5] is a three or four nodes finite shell element. It is based on the Mindlin-Reissner theory. The attached assumptions mainly impose a state of plane stresses and an invariance of the deformation through the thickness. In addition to both assumptions, it is assumed that the transverse and shear effects are taken into account. The interpolation functions relating to the finite element are bilinear.

Owing to the ply concept, the stacking and the nature of materials composing the structure to be designed (multi-layered multi-material, laminated composites, etc) are reproduced. The constituent materials can be only of the metallic, composite or foam type. Each ply has its own material characteristics, its own behaviour laws as well as a point of integration through thickness, located in the middle section of the ply.

2.2 Unidirectional composite elementary ply modelling

The finite element offers the choice between two possibilities for the modelling of the unidirectional composite elementary ply. The first model is known as “bi-phase” [6] [7] [8] and distinguishes the fibre behaviour from the matrix one. The second model is known as “global” and a homogenized behavioural law describes the behaviour of an elementary ply [9] [10]. The optimisation is based on the latter model: the essential reason lies in the fact that this model features an easy experimental methodology in order to identify material characteristics. It is thus an essential asset for the software users.

The global model allows the transcription of considerable phenomena such as elasticity, plasticity and certain types of degradation. The latter correspond to the micro-cracking of the matrix parallel to fibres, the debonding of fibre/resin and finally the rupture of fibres. The plastic flow is taken into account using a standard “von Mises” plasticity criterion coupled to an isotropic hardening law. Owing to the effective stress concept, the damage is integrated into the plastic flow. All the material parameters corresponding to this elastic plastic damaging law are identified thanks to some simple or cyclic tensile and compressive tests on five laminates with preset orientations [11].

2.3 Positioning of the problem

In order to justify the validity of the finite element introduced for the multi-layered multi-material structure dimensioning, some results are shown here. They are related to dynamic three point bending on plates composed of aluminium and composite glass-E/epoxy layers [1] [12] [13]. We can note a very good correlation between the experimental results and the values resulting from the numerical simulation with the new multi-material element (figure 1). For each case, the maximum error featured is lower than 12%. As one could expect, the superposition of elements with different behaviours does not agree at all with the treatment of the structures subjected to bending loads. Our development thus introduces new possibilities for the numerical simulation study of multi-material laminated structures. Other applications on tubes impacted in axial compression have also produced numerical results in agreement with the experimentation [2] [3] [4] [12] [13]. The tubes were formed of one steel layer reinforced by rolling up glass fibres and epoxy resin. They can be connected to energy absorbers.

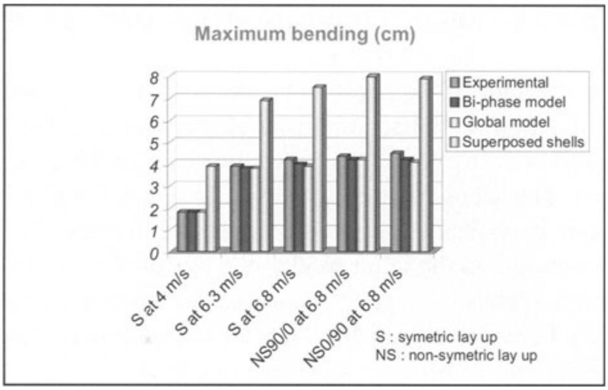


Figure -1. Maximum bending to dimension GLARE plates.

The results previously cited have confirmed the validity of the finite element within the integration of a multi-layered multi-material study framework. However, the applications often relate to accidental shocks (automotive crash, bird impact on a wing section, meteorite impact on a protection shield, etc) during which an influencing parameter can intervene: the strain rate.

3. STRAIN RATE SENSITIVITY OF UNIDIRECTIONAL COMPOSITE MATERIALS

For conventional materials such as metallic ones, behaviour laws suitably include the influence of the strain rate parameter. However, for laminated composites, the complexity of their intrinsic behaviour means that the static study of these materials must be supplemented by a dynamic one in order to check, and in the optimal case to quantify, the sensitivity of the material parameters to the load velocity.

3.1 Some experimental observations

The experimental campaign has concerned strain rate sensitivity of the material parameters of elastic plastic damaging law. Indeed, it is then possible to maintain certain reproducibility as well as feasibility for the computational code users with regard to characterisation tests. Moreover, it is also possible to preserve the global model underlying. The tests taken into account are consequently composed of tensile tests on laminates $[0]_4$, $[\pm 45]_S$, $[\pm 67.5]_S$, $[+45]_4$ and of a compressive test on laminate $[0]_6$.

The experimental device consists of a dynamic tension/compression machine (rapid hydraulic jack). For each type of test, four main load velocities were applied: 5 mm/min, 500 mm/min, 0.2 m/s and 4 m/s. Some intermediate velocities were introduced, thus increasing the results' accuracy. The strain rate covered ranges from quasi-static state to approximately 150 s^{-1}.

Two essential conclusions arise from the experimental tests: first, as figures 2 and 3 show, elastic moduli, initial yield stress and longitudinal fracture stress undergo a significant increase starting from a certain threshold strain rate. These increases are about 30% to 300% according to the constants. Then, we note that in the longitudinal case (figure 2), there is a behaviour law conservation since the aforementioned remains fragile and elastic. However, in the shearing case (figure 3), the type of behaviour tends to become "damaging elastic" under strain rate influence.

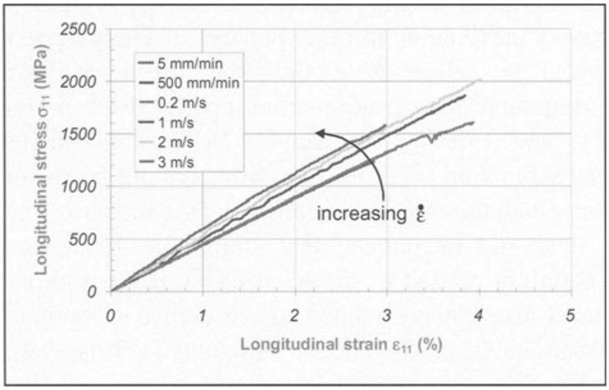


Figure -2. Dynamic tensile tests on laminate $[0]_4$.

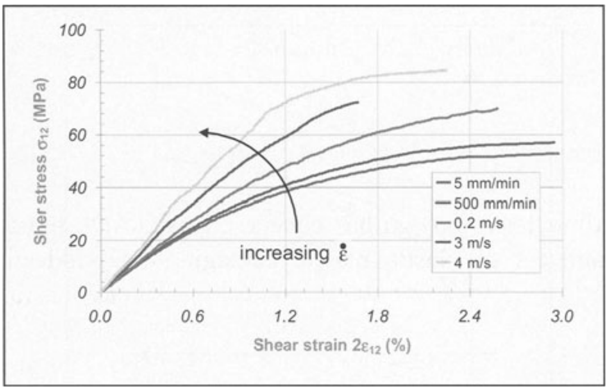


Figure -3. Dynamic tensile tests on laminate $[\pm 45]_S$.

Consequently, the assumption adopted is that behaviour law can be described overall as a "viscous" elastic plastic damaging law and, at a given strain rate, can be always identified, starting from a behaviour "reference" law. Lastly, the evolution of all previous constants are, for this glass-E/epoxy composite, in an analytical form power according to the strain rate [14].

3.2 Theoretical approach: "viscous" modelling

The type of behaviour observed in experiments is not easily identifiable: it represents neither visco-elasticity nor visco-plasticity. However, it is possible to approach the definition suggested by Lemaître and Chaboche [15] with regard to viscous fluids.

Indeed, epoxy resins, and more generally reticulate polymers, are sensitive to the strain rate during loads [16]: viscosity primarily comes from the chain segment movements through barriers of energy potentials. These phenomena are all the more active (since they are thermally activated) as the vitreous transition temperature from epoxy resin is relatively low (between 50°C and 150°C). We can be led to think that this rise in temperature, as a function of strain rate, brings about some micro-changes in resin phases, which thus tend to become "soft" solids.

However, it should be noticed that strain rate influence on material behaviour is certainly related to the resin dependence with respect to this strain rate, but is also closely related to the choice of components. Some studies on carbon/epoxy unidirectional laminates [17] have shown a strain rate insensitivity of the elastic moduli: carbon fibres have a Young modulus on average three times higher than that of glass-E fibres and thus inhibit the viscous effects of resin (in the elastic range).

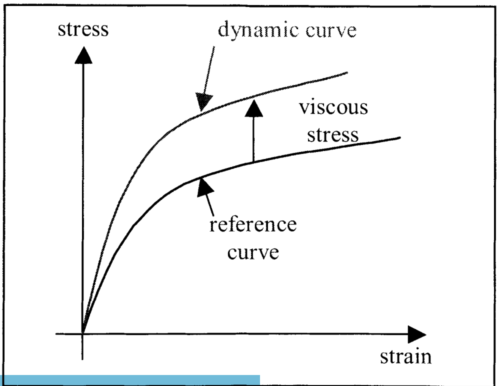


Figure -4. Definition of viscous stress.

Modelling results from the approach of viscous fluids [14]: the "pseudo-visco-elasticity" of resin (and in fact of laminated composite) are taken into account using a viscous stress, which is itself combined with the elastic stress (figure 4). Indeed, in experiments, it is always possible to determine the behaviour at a given strain rate according to behaviour at a reference strain rate.

In order to integrate the pseudo-visco-elastic nature of resin, the idea is to consider that dissipation's potential form is identical to the thermodynamic potential one, with the help of some characteristic functions of resin viscosity. Therefore, the stress tensor becomes as equation (1) according to the state law and the complementary law.

$$\underline{\underline{\sigma}} = C^0 : \underline{\underline{\varepsilon}}^e + \Theta : \underline{\underline{\dot{\varepsilon}}}^e \quad (\Theta : \textit{viscous tensor}) \tag{1}$$

The development of equation (1) for longitudinal, transverse and shear direction results in a description of elastic range. However, it is simplified owing to the approximation due to the dynamic rapid load type we are faced with. It can then be attributed to this general equation describing the elastic range:

$$\underline{\underline{\sigma}} = C^0 (I + F) : \underline{\underline{\varepsilon}}^e \quad (F\text{: \textit{viscous functions}}) \tag{2}$$

In a phenomenological way, the material state (resin or composite) can always be given starting from its state in the reference configuration. We can thus suppose that the state for a given strain rate corresponds to the quasi-static state of another glass E/epoxy composite material from which material characteristics are different. Since the elastic moduli can increasingly evolve with the strain rate, this assumption results in postulating a damage progress faster than that of the reference case. It is possible to show that the constant damage evolution intrinsic to modelling [14], is a function root of the viscosity shear function:

$$Y_{ij}(\dot{\varepsilon}) \approx \sqrt{(1 + F_{12}(\dot{\varepsilon}))}\, Y_{ij} \tag{3}$$

where Y_{ij} represent each damage parameter

The previous assumption also suggests that the plasticity field moves accordingly (with an evolution of initial yield stress). Indeed, the initial yield stress undergoes an evolution of the same type as the elasticity moduli (4). Resin plasticity is coupled with the model by supposing that the plastic strain evolution is similar to the reference case. However, they are delayed since the material becomes increasingly "rigid".

$$R_0(\dot{\varepsilon}) = \left(1 + F_R(\dot{\varepsilon})\right)R_0 \tag{4}$$

Finally, we put forward the additional assumption that the energy rate required to break the laminate (in the transverse and/or shear directions) is insensitive to the strain rate.

4. VALIDATIONS OF THE IMPLEMENTATION

4.1 Elementary numerical validations

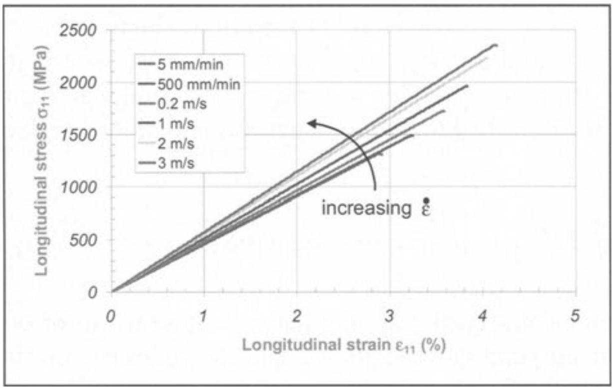


Figure -5. Numerical dynamic tensile tests on laminate $[0]_4$.

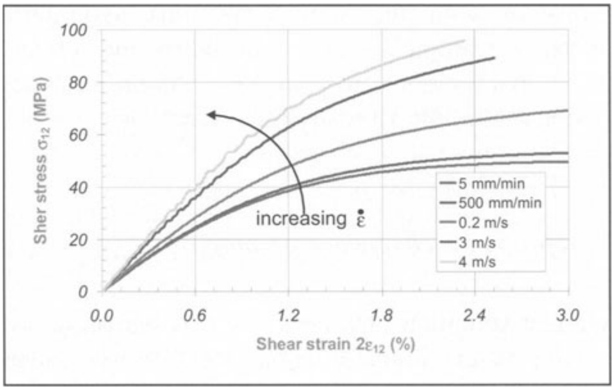


Figure -6. Numerical dynamic tensile tests on laminate $[\pm 45]_S$.

Initially, it is necessary to check that the "primary" theoretical behaviours are suitably represented: therefore, we have carried out validations known as elementary [14]. They correspond to the simulation of experimental tensile or compressive tests derived from the static or dynamic study framework. In each case, the test piece is represented – since its geometry makes it possible – by a single shell element. Figures 5 and 6 as compared to figures 2 and 3 visually show that the approach developed seems to correlate satisfactorily with the experiment.

4.2 Application to the case of plates subjected to impacts

Since the "viscous" behaviour law seems to be validated satisfactorily for the identification tests, we have started an experimental campaign in connection with more complex loads such as three point dynamic bending and impacts on plates $[0_2\ +45_2\ -45_2\ 90_2]_S$, $[+45_2\ -45_2]_{2S}$ in order to check the reliability of the developments. In this part, we have some preliminary results [18] derived from the examinations of an impacted plate $[0_2\ +45_2\ -45_2\ 90_2]_S$. A spherical ball, coupled to a carriage, proposes a mass of 26.5 kg and a velocity of 4 m/s at impact moment.

In figure 7, we compare the experimental and numerical evolutions of applied load versus time. The first observations tend to show that the strain rate sensitivity integration is a significant precaution since it allows to correct the structure rigidity adequately: the average strain rate observed at impact point is indeed about 100 s^{-1} (figure 8). The temporal description of load seems to use the behavioural law optimisation more satisfactorily than that proposed by "quasi-static" modelling. Indeed, the first maximum load observed supported by the plate is around 1.5 ms.

However, figure 7 highlights that the maximum load rate, applied to the plate always remains lower than the experimental one although a temporal description is satisfactory. Indeed, during tests we have observed the appearance of delamination between certain plies (hatched surface on figure 7). However the numerical modelling adopted does not take into account this phenomenon; we suppose that this difference in load rate results from a brutal "numerical" degradation of the structure rather than a gradual degradation observed in experiments. The state of damage is distributed in a prompt and integral way depending on the thickness of the laminate whereas in experiments the appearance of delamination authorizes the conservation of a certain structure integrity and thus of its rigidity. Studies are currently in progress in order to apply a methodology developed within the laboratory [19] to validate this assumption.

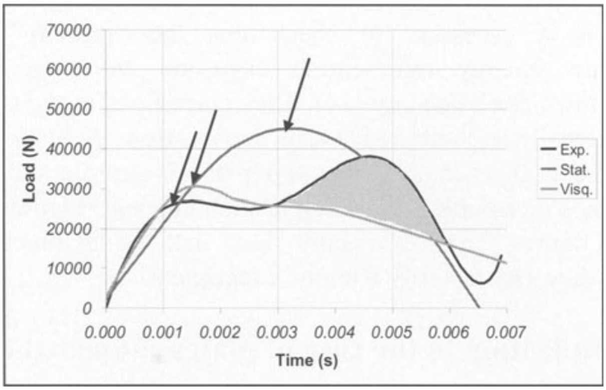


Figure -7. Experimental and numerical comparisons for an impact.

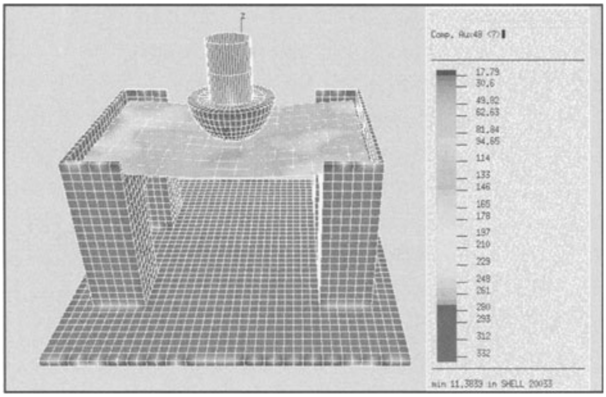


Figure -8. Numerical contour of the strain rate.

5. CONCLUSIONS AND FUTURE PROSPECTS

The few results presented enable us to highlight a certain relevance of the original approach adopted. Indeed, the elementary or complex validations show that the theoretical and data-processing developments lead to being taken into account the strain rate influence satisfactorily. The optimisation of the data-processing tools allows the computational code users to possess a powerful tool. They can simulate structure dimensioning including not only multi-layered multi-material but also laminated fibre/resin composites.

The whole approach was transposed with the same conclusions as previously, to fabric glass E/epoxy or carbon/epoxy laminates. These results consequently encourage us to continue and refine our work, in particular by proposing a more suitable description integrating the temperature rise phenomena. Other studies are also in progress in order to integrate the major delamination phenomena.

ACKNOWLEDGMENTS

We thank Mr. E Haug and Mr. X Ni from ESI Software Company as well as the ANRT, which have allowed the realisation of this thesis subject under a CIFRE agreement. We address also our warmest thanks to ONERA of Lille, and more particularly to Mr. E Deltombe and Mr. D. Dormegnie, for their support and for the realisation of the whole experimental tests. Finally, we thank the CNRS for their permanent support of the laboratory.

REFERENCES

[1] D. COUTELLIER, C. GAUTHIER, Y. RAVALARD, X. NI, E. HAUG, "Modélisation en flexion des multicouches et multi-matériaux en dynamique rapide", JNC 10, D. Baptiste & A. Vautrin, 1996, pp. 969-980.

[2] D. COUTELLIER, C. GAUTHIER, P. ROZYCKI, Y. RAVALARD, "Numerical simulation of the tensile behaviour of multi-layered multi-material within an explicit software code", European Journal of Mechanical Environmental Engineering, vol 43, 1998, pp.51-60.

[3] C. GAUTHIER, D. COUTELLIER, Y. RAVALARD, "Modélisation d'un élément multicouches et multi-matériaux dans le logiciel d'éléments finis PAMCRASH™", Giens95, 1995, pp. 373-378.

[4] C. GAUTHIER, "Contribution à la modélisation du comportement en crash des structures stratifiées métal/composite. Développement d'un élément de coque multi-couches multi-matériaux", thèse de l'Université de Valenciennes et du Hainaut-Cambrésis, 1996.

[5] Pam Crash™, "Solver Notes Manual & Solver Reference Manual", ESI Group Software product company, Rungis, 1998.

[6] E. HAUG, H. CHARLIER, J. CLINCKEMAILLE, "Recent trends and developments of crashworthiness simulation methodologies and their integration into industrial vehicle design cycle", Third European Cars/Trucks Simulation Symposium (ASIMUTH), 1991.

[7] E. HAUG, A. DE ROUVRAY, "Crash response of composite structure", Chapter 7 of Structural Crashworthiness and Failure, Ed. by N. Jones & T. Wierrzbicki, Elsevier, London, 1993.

[8] A.K. PICKETT, J. RUCKERT, E. HAUG, "A fracture damaging law suitable for anisotropic short fibre/matrix materials in an explicit finite element code", Journal of Composites, Vol. 21, n°4, 1990, pp 297-304.

[9] P. LADEVEZE, "Sur la mécanique de l'endommagement des composites", JNC 5, Ed. Pluralis, Paris, 1986, pp. 667-683

[10] O. ALLIX, P. LADEVEZE, E. LE DANTEC, "Modélisation de l'endommagement du pli élémentaire des composites stratifiés", JNC 7, Ed. Pluralis, Lyon, 1990, pp. 715-724.

[11] D. COUTELLIER, P. ROZYCKI, "New functionalities for PAMCRASH™ multi-layered multi-material element", Composites Part A – Applied Science and Manufacturing, 2000, Vol-31, pp. 841-851.
[12] P. ROZYCKI, "Flexion trois points dynamique sur des plaques de multicouches multi-matériaux", Mémoire de D.E.A., Partie n°1, Université de Valenciennes et du Hainaut-Cambrésis, 1996.
[13] P. ROZYCKI, "Compression dynamique de multicouches multi-matériaux", Mémoire de D.E.A., Partie n°2, Université de Valenciennes et du Hainaut-Cambrésis, 1996.
[14] P. ROZYCKI, "Contribution au développement de lois de comportement pour matériaux composites soumis à l'impact", thèse de l'Université de Valenciennes et du Hainaut-Cambrésis. Valenciennes, 2000.
[15] J. LEMAITRE, J.L. CHABOCHE, "Mécanique des matériaux solides", Ed Dunos, Paris, 1985, Chapitre 1 : pp. 26-31, chapitre 4 : pp. 145-148.
[16] J.L. CHABOCHE, "Formalisme général des lois de comportements: applications aux métaux et polymères. Introduction à la mécanique des polymères", Ed. G'sell C. & Haudin J.M., Nancy, 1995, pp. 119-140.
[17] O. ALLIX, J.F. DEU, P. LADEVEZE, "A delay-damage meso-model for the prediction of localisation and fracture of laminates subjected to high-rate loading", ECCM'99, European Conference on Computational Mechanics, München, Germany, 1999.
[18] P. ROZYCKI, "Validations numériques de la loi élasto-plastique endommageable visqueuse – Essais de flexion trois points dynamiques et impact de plaques", rapport interne de l'Université de Valenciennes, 2001, à paraître.
[19] J.C. WALRICK, "Contribution au développement d'une nouvelle méthodologie pour l'étude du délaminage dans les structures stratifiées composites : application à l'impact basse vitesse", thèse de l'Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes, 1999.

ON THE KINEMATIC CALIBRATION OF A STEWART PLATFORM

F. Dudita*, M. Neagoe* and G. Gogu**
Transilvania University of Brasov, Romania.
***Laboratoire de Recherches et Applications en Mécanique Avancée, Institut Français de Mécanique Avancée et Université Blaise Pascal, Clermont-Ferrand, France.*

Abstract: This paper presents several results regarding the kinematic calibration of a Stewart platform. Firstly, a general method for the precision modeling of parallel robots is proposed by the authors and applied in the case of the Stewart platform. Thus, an accuracy model of the Stewart platform is established, taking into account the errors of the geometrical parameters. Next, on the basis of the precision model and the experimental data for the accuracy of the Stewart platform, some algorithms used in kinematic calibration were generated and tested; overcoming some numerical difficulties, the actual deviations of the geometrical parameters were identified. Thus, the correct values of geometrical parameters and, implicitly, the correct kinematic model used in command/control were obtained. Finally, the importance of the calibration in the optimization of the kinematic models is emphasized, based on the numerical results. According to these aspects, the accuracy of the Stewart platform model was significantly increased.

Key words: Stewart platform, accuracy, modelling, testing, kinematic calibration.

1. INTRODUCTION

The *parallel robots* represent a relative recently category of robots, which have tended to become an interest-center for researchers and practitioners. Although they offer multiple advantages (suppleness, high speed and accuracy, large loading etc.), parallel robots entail many difficulties in kinematic and dynamic modeling and, implicitly, in precision modeling and calibration.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 117–128.

Precision models play a central role in the calibration process. The modeling of parallel robots accuracy is little broached in the literature in comparison with serial robots. Vischer presents a precision model for a Stewart platform based on the accuracy of 138 geometric parameters [21]. Wang used 132 parameters to model the accuracy of the Stewart platform [22, 23]. Daney proposed a model based on the accuracy of 42 geometric parameters for the same type of parallel robot [3]. For the calibration of the Stewart platform different methods are used: direct methods based on the direct kinematic model [18, 21], inverse methods based on the inverse kinematic model [20, 21, 22, 26, 27], optimization/identification algebraic methods [9, 10], semi-parametrical methods [21] and constrained calibration methods [3, 6, 9, 11, 15, 24, 25].

A robot is a complex system, on which many error factors act; therefore it has an actual behavior that deviates from the desired (commanded) one, established on the basis of the *nominal model*. Robot accuracy can be increased, without structural or constructive improvements, through *calibration*, which allows in the command process the use of an *actual* (*correct*) *model*, more accurately related to the real robot. Calibration attempts to identify the most important error factors, to model theirs influences on the robot accuracy and to obtain the actual (real, correct) values of the modeling parameters by analyzing the experimental data from accuracy testing. Thus, the nominal models can be optimized and the actual models (used in command/control process) ensure the increase of robot accuracy.

General considerations about these aspects are published by Mooring [14], Hollerbach [8], Khalil [12], Bernhardt [1], Schrörer [19] etc.

In this paper, the authors propose a general method, used in precision modeling of parallel robots, applied to a Stewart platform (section 3); based on the error model obtained and on the general considerations about calibration, a kinematic calibration algorithm for the Stewart platform is generated and then tested full stop (section 4). For a better understanding, a brief description of the Stewart platform and its parameterization are presented (section 2).

2. STEWART PLATFORM: BRIEF PRESENTATION AND PARAMETERIZATION

The Stewart platform (Fig. 1) is made up of an upper moving plate (tool base) connected to a fixed plate (robot base) by 6 telescopic arms, whose main part is a long prismatic joint (Fig 1,b). A lower universal joint connects the robot base to the prismatic joint and a spherical joint connects the other end of the prismatic joint to the tool base.

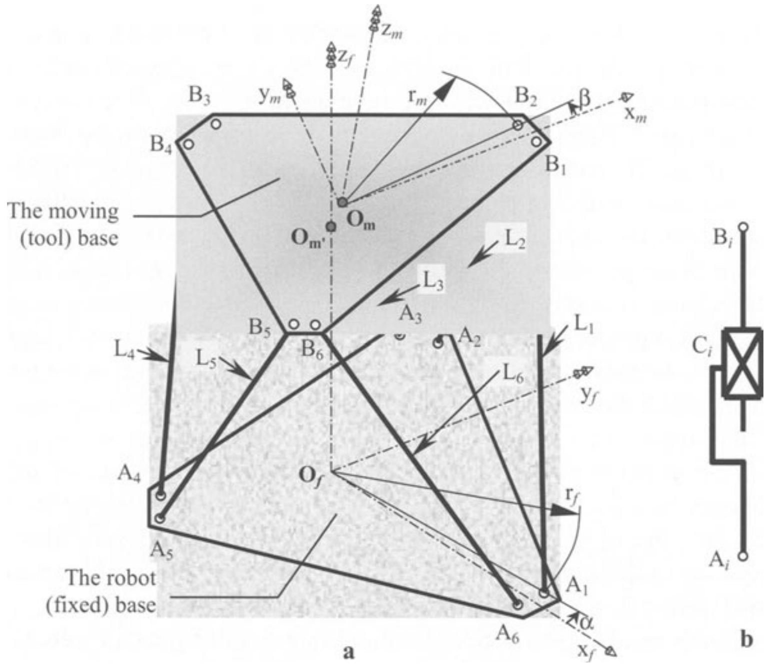


Figure -1. The Stewart platform

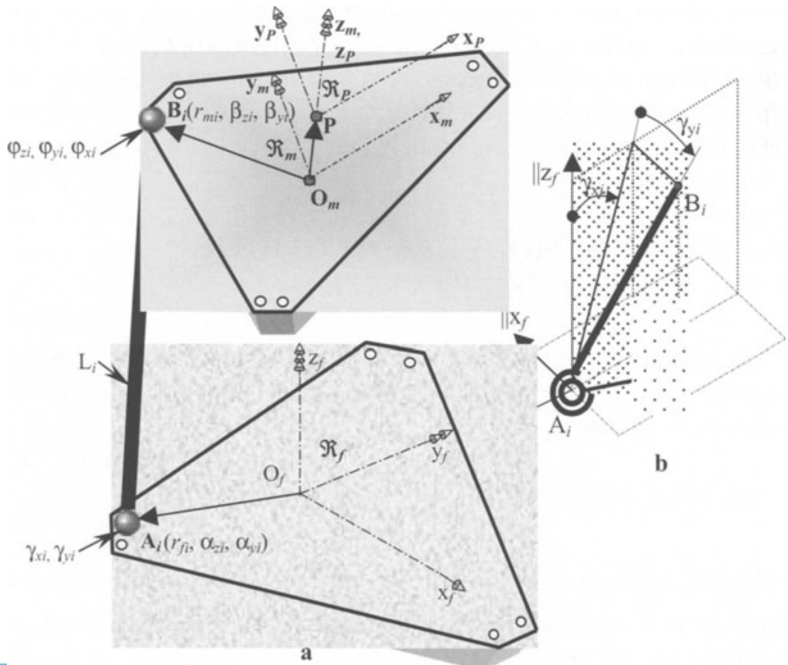


Figure -2. Decomposition of the parallel structure in open chains and its parameterization

The parallel structure geometry, defined by the relative position of the 12 joint center points (A_i and B_i, i=1..6, Fig. 1), is characterized, for the analysed platform, by the following parameters: r_f= 270 mm; r_m= 195 mm; α = 4.25°; β = 5.885°. We associate to the Stewart platform three main frames (Fig. 2): *robot base frame* $\mathfrak{R}_f(O_f\ x_f\ y_f\ z_f)$, *moving plate frame* $\mathfrak{R}_m(O_m x_m y_m z_m)$ and *tool base frame* $\mathfrak{R}_P(O_P x_P y_P z_P)$. In the initial configuration, the moving plate is located in $O_{m'}$ ($O_m \equiv O_{m'}$, $h = O_f O_m$) and the frame $\mathfrak{R}_m$ is parallel to the frame $\mathfrak{R}_f$. The pose of the tool base, relative to the robot base, is described by 6 parameters: *3 positioning parameters* – the vector $O_{m'}O_m$ components (Fig. 1), expressed in the $\mathfrak{R}_f$ frame and *3 orienting parameters*, describing the orientation of the $\mathfrak{R}_m$ frame relative to the reference frame $\mathfrak{R}_{m'}$, with the following significances: $\theta_1 - z$ axis angle; $\theta_2 - x$ axis angle; $\theta_3 - y$ axis angle.

For the position of points A_i in the $\mathfrak{R}_f$ frame we use a set of spherical coordinates (α_{zi}, α_{yi}, r_{fi}) and similarly for points B_i (defined in the $\mathfrak{R}_m$ frame, Fig. 2): (β_{zi}, β_{yi}, r_{mi}). The joint displacements in the universal joints A_i are modeled by angles γ_{xi} and γ_{yi} (Fig. 2,b); similarly, the displacements of spherical joints B_i are modeled by angles (φ_{zi}, φ_{yi}, φ_{xi}).

Precision modeling is a special topic in the field of parallel robots. These models are useful in calibration process and represent one of the most difficult problems we must solve. Therefore, the next section presents briefly a *new general method* used in *precision modeling* of *parallel structures* in which we consider the objective of the kinematic calibration proposed to be solved (to identify the actual values of geometrical parameters) and we present explicitly the precision modeling for geometrical parameter deviations. A complete error model of this Stewart platform is shown in [17].

3. ERROR MODEL OF GEOMETRICAL PARAMETER DEVIATIONS

The influence of geometrical parameter deviations on end effector errors is described by a linear model, where the error Jacobian $\mathbf{J_G}$ is the matrix of the following system:

$$[\mathbf{w}]_P = [\mathbf{J_G}]_P \cdot [d\mathbf{G}], \tag{1}$$

where $[\mathbf{w}]_P = [d_x\ d_y\ d_z\ \delta_x\ \delta_y\ \delta_z]_P^T$ is the hexadimensional vector of end effector errors, corresponding to the characteristic point P (O_mP=h_m, Fig. 2), reduced in the $\mathfrak{R}_P$ frame; $[d\mathbf{G}]$ is a 36 dimensional vector of the errors of geometrical parameters used in kinematic modeling:

$$[d\mathbf{G}] = \left[\delta\alpha_{z1}\ \delta\alpha_{y_1}\ dr_{f1}\ dr_{m1}\ \delta\beta_{y_1}\ \delta\beta_{z1}\ \ldots.\ \delta\alpha_{z6}\ \delta\alpha_{y_6}\ dr_{f6}\ dr_{m6}\ \delta\beta_{y_6}\ \delta\beta_{z6}\right]^T.$$

The following assumptions are used in the error modeling:

- The joint errors (in active joints C_i, Fig. 1,b) are neglected.
- The geometrical parameters are affected by constant deviations.

A preliminary step is applied in the *error Jacobian* $\mathbf{J}_\mathbf{G}$ modeling: *decomposition of parallel structure in 6 open chains*, one for each arm: $O_f A_i B_i O_m P$ (Fig. 2). Next, a three steps algorithm is used.

Step 1. The end effector errors description for all 6 open chains

The end effector errors being expressed in the $\Re_p$ frame, we have to use *D–F* (type *K*) homogenous operators [5, 7] in direct kinematic modeling of each arm. The kinematic modeling is applied to an *equivalent structure*: to each geometrical parameter we associate a *fictive joint* (prismatic or rotational joint). Thus, in the direct kinematic modeling of the arm *i* we use the following homogenous operators: $\mathbf{A_{01}} = \mathbf{R_z}(\alpha_{zi})$; $\mathbf{A_{12}} = \mathbf{R_y}(\alpha_{yi})$; $\mathbf{A_{23}} = \mathbf{T_x}(r_{fi})$; $\mathbf{A_{34}} = \mathbf{R_x}(\gamma_{xi})$; $\mathbf{A_{45}} = \mathbf{R_y}(\gamma_{yi})\cdot \mathbf{T_z}(L_i)$; $\mathbf{A_{56}} = \mathbf{R_z}(\varphi_{zi})$; $\mathbf{A_{67}} = \mathbf{R_y}(\varphi_{yi})$; $\mathbf{A_{78}} = \mathbf{R_x}(\varphi_{xi})$; $\mathbf{A_{89}} = \mathbf{T_x}(-r_{mi})$; $\mathbf{A_{9\_10}} = \mathbf{R_y}(-\beta_{yi})$; $\mathbf{A_{10\_11}} = \mathbf{R_z}(-\beta_{zi}) \cdot \mathbf{T_z}(h_{mp})$.

The end effector errors of each arm can be expressed by applying the general relations used in precision modeling of open chains [7]:

$$\begin{gathered}
{}^i[\mathbf{w}]_P = [\mathbf{J}_i]\left[\delta\alpha_{zi}\ \delta\alpha_{yi}\ dr_{fi}\ \delta\gamma_{xi}\ \delta\gamma_{yi}\ \delta\varphi_{zi}\ \delta\varphi_{y_i}\ \delta\varphi_{xi}\ dr_{mi}\ \delta\beta_{yi}\ \delta\beta_{zi}\right]^T = [\mathbf{J}_i][d\mathbf{\Phi}_{Gi}], \\
[\mathbf{J}_i] = \left[J_{\alpha_{zi}}\ J_{\alpha_{yi}}\ J_{r_{fi}}\ J_{\gamma_{xi}}\ J_{\gamma_{yi}}\ J_{\varphi_{zi}}\ J_{\varphi_{y_i}}\ J_{\varphi_{xi}}\ J_{r_{mi}}\ J_{\beta_{yi}}\ J_{\beta_{zi}}\right]^T,
\end{gathered} \tag{2}$$

where each column of the Jacobian $\mathbf{J}_i$ describes the influence of one error–source on the end effector errors.

Step 2. Dependent error identification

The joint displacements in the *passive joints* A_i and B_i are *dependent displacements;* as a result, the displacements errors in these joints are also *dependent errors*. Knowing that the *end effector errors are identical for all 6 arms* (due to the parallel connectivity), all dependent errors can be expressed starting from the following equations:

$$[\mathbf{w}]_P = [\mathbf{J}_1][d\mathbf{\Phi}_{G1}] = \ldots = [\mathbf{J}_6]\,[d\mathbf{\Phi}_{G6}]. \tag{3}$$

From eq. (3) 5 independent matrix equations result; separating the dependent term from the independent ones, we obtain:

$$\left[J_{\gamma_{x1}}\ J_{\gamma_{y1}}\ J_{\varphi_{z1}}\ J_{\varphi_{y1}}\ J_{\varphi_{x1}}\right]\left[\delta p^1_{dep}\right] - \left[J_{\gamma_{xk}}\ J_{\gamma_{yk}}\ J_{\varphi_{zk}}\ J_{\varphi_{yk}}\ J_{\varphi_{xk}}\right]\left[\delta p^k_{dep}\right] =$$

$$= \left[J_{\alpha_{zk}} \; J_{\alpha_{yk}} \; J_{r_{fk}} \; J_{r_{mk}} \; J_{\beta_{yk}} \; J_{\beta_{zk}} \right] \left[\delta p_{ind}^{k} \right] - \left[J_{\alpha_{z1}} \; J_{\alpha_{y1}} \; J_{r_{f1}} \; J_{r_{m1}} \; J_{\beta_{y1}} \; J_{\beta_{z1}} \right] \left[\delta p_{ind}^{1} \right],$$

$$k = 2..6, \tag{4}$$

where we use the following notations: $\left[dp_{dep}^{i} \right] = \left[\delta\gamma_{xi} \; \delta\gamma_{yi} \; \delta\varphi_{zi} \; \delta\varphi_{yi} \; \delta\varphi_{xi} \right]^{T}$, $\left[dp_{ind}^{i} \right] = \left[\delta\alpha_{zi} \; \delta\alpha_{yi} \; dr_{fi} \; dr_{mi} \; \delta\beta_{yi} \; \delta\beta_{zi} \right]^{T}$.

Next, the 5 systems (4) are assembled in one matrix equation:

$$\left[\mathbf{J}_{\Phi}^{*} \right] \cdot [\delta\Phi] = \left[\mathbf{J}_{G}^{*} \right] \cdot [d\mathbf{G}], \tag{5}$$

where $[\delta\Phi] = \left[\delta\gamma_{x1} \; \delta\gamma_{y1} \; \delta\varphi_{z1} \; \delta\varphi_{y1} \; \delta\varphi_{x1} \; \ldots \; \delta\gamma_{x6} \; \delta\gamma_{y6} \; \delta\varphi_{z6} \; \delta\varphi_{y6} \; \delta\varphi_{x6} \right]^{T}$ is the global vector of dependent errors, and

$$\left[\mathbf{J}_{\Phi}^{*} \right] = \begin{bmatrix} \mathbf{J}_{\Phi 1} & -\mathbf{J}_{\Phi 2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{J}_{\Phi 1} & \mathbf{0} & -\mathbf{J}_{\Phi 3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{J}_{\Phi 1} & \mathbf{0} & \mathbf{0} & -\mathbf{J}_{\Phi 4} & \mathbf{0} & \mathbf{0} \\ \mathbf{J}_{\Phi 1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{J}_{\Phi 5} & \mathbf{0} \\ \mathbf{J}_{\Phi 1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{J}_{\Phi 6} \end{bmatrix},$$

$$\left[\mathbf{J}_{G}^{*} \right] = \begin{bmatrix} -\mathbf{J}_{G1} & \mathbf{J}_{G2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{J}_{G1} & \mathbf{0} & \mathbf{J}_{G3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{J}_{G1} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{G4} & \mathbf{0} & \mathbf{0} \\ -\mathbf{J}_{G1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{G5} & \mathbf{0} \\ -\mathbf{J}_{G1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_{G6} \end{bmatrix}, \tag{6}$$

$\mathbf{J}_{\Phi k} = \left[J_{\gamma_{xk}} \; J_{\gamma_{yk}} \; J_{\varphi_{zk}} \; J_{\varphi_{yk}} \; J_{\varphi_{xk}} \right]$, $\mathbf{J}_{Gk} = \left[J_{\alpha_{zk}} \; J_{\alpha_{yk}} \; J_{r_{fk}} \; J_{r_{mk}} \; J_{\beta_{yk}} \; J_{\beta_{zk}} \right]$, $k = 1..6$.

Finally, the dependent errors can be calculates by the following equation:

$$[\delta\Phi] = \left[\mathbf{J}_{\Phi}^{*} \right]^{-1} \left[\mathbf{J}_{G}^{*} \right] \cdot [d\mathbf{G}] = \left[\mathbf{J}^{*} \right] \cdot [d\mathbf{G}]. \tag{7}$$

The matrix $\left[\mathbf{J}_{\Phi}^{*} \right]$ is a full rank square 30×30 matrix; consequently, it is an inverted matrix.

Step 3. End effector errors determination

The matrix $\mathbf{J}^*$ (eq. 7) is a 30×36 size matrix, which can be subdivided into 6 submatrices of 5 rows; each submatrix represent the error Jacobian of dependent errors from one arm related to the deviations $d\mathbf{G}$. Considering one of the open chains, e.g. arm 1, the end effector errors expressed in $\Re_p$ are:

$$[\mathbf{w}]_P = [\mathbf{J}_1]\left[\delta\alpha_{z1}\ \ \delta\alpha_{y1}\ \ dr_{f1}\ \ \delta\gamma_{x1}\ \ \delta\gamma_{y1}\ \ \delta\varphi_{z1}\ \ \delta\varphi_{y1}\ \ \delta\varphi_{x1}\ \ dr_{m1}\ \ \delta\beta_{y1}\ \ \delta\beta_{z1}\right]^T =$$
$$[\mathbf{J}_1]\left[\mathbf{J}^*_{\Phi 1}\right]\ [d\mathbf{G}], \tag{8}$$

where $\left[\mathbf{J}^*_{\Phi 1}\right]$ is a special matrix [17], which relates the vector $d\mathbf{\Phi}_{G1}$ to the vector $d\mathbf{G}$.

Finally, considering eq. (8) and (1), the error Jacobian $\mathbf{J}_\mathbf{G}$ can be obtained through the following relations:

$$[\mathbf{J}_\mathbf{G}] = [\mathbf{J}_1]\left[\mathbf{J}^*_{\Phi 1}\right] = [\mathbf{J}_2]\left[\mathbf{J}^*_{\Phi 2}\right] = \ldots = [\mathbf{J}_6]\left[\mathbf{J}^*_{\Phi 6}\right]. \tag{9}$$

In conclusion, the modeling method presented in the paper allows the generalization of error modeling for parallel robots through an algorithmic and systematic approach. Although the error model obtained is relatively complex (due to the inversion of a 30×30 matrix), the error Jacobian $\mathbf{J}_\mathbf{G}$ can be computed by using one of the equations (9).

4. KINEMATIC CALIBRATION OF STEWART PLATFORM

In the calibration process of the Stewart platform we have considered the following assumptions: a) the end efector errors are exclusively generated by geometrical parameter deviations; b) the relative displacements in active joints (joint C_i, i=1..6, Fig. 1,b) are not affected by errors. This problem is solved in two relevant cases:

1. Kinematic calibration using only position measurements;
2. Kinematic calibration using pose (position and orientation) measurements.

4.1 Kinematic calibration using only end effector position errors

In this case, for each robot configuration i experimentally tested, from eq. (1) we use only the first 3 scalar equations:

$$\begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix}_i = \left[\mathbf{J}_{Gi}^{p}\right] \cdot [d\mathbf{G}], \tag{10}$$

where $\mathbf{J}_{Gi}^{p}$ is the upper 3×36 submatrix from $\mathbf{J}_{Gi}$ (see eq. 1), corresponding to the positioning errors. For *m* configurations tested we obtain *m matrix equations* (10), which can be assembled into the following *overdetermined* linear system:

$$[d\mathbf{X}] = \left[\mathbf{J}_{G}^{p}\right] \cdot [d\mathbf{G}], \quad \left[\mathbf{J}_{G}^{p}\right] = \begin{bmatrix} \mathbf{J}_{G1}^{p} \\ \dots \\ \mathbf{J}_{Gm}^{p} \end{bmatrix}, \tag{11}$$

where $d\mathbf{X}$ is the global vector of end effector positioning errors ($d\mathbf{X} = [d_{x1}, d_{y1}, d_{z1}, \dots, d_{xm}, d_{ym}, d_{zm}]^T$).

In the attemption to solve eq. (11), using the *linear least squares method*, we must first eliminate some numerical difficulties related to the Jacobian matrix $\left[\mathbf{J}_{G}^{p}\right]$: rank deficiency, ill condition number [8], identification of the redundant and unidentifiable parameters; in the case of Stewart platform we eliminated the parameters $r_{m1}, r_{m2}, \dots, r_{m6}$.

After all these steps, the system (11) was solved applying the method of *zeroing small singular values* [8]. The parameter deviations having been identified, we established the actual values of the geometrical parameters. Then, the kinematic model was optimized and the correct kinematic model could be use in command/control process. The end effector errors are depicted in Fig. 3, for a representative set of configurations tested, in both cases: before and after calibration. As shown in Fig. 3, we can draw the following conclusions:

– The robot positioning accuracy is about 3 times better after calibration, compared with before calibration (Fig. 3,a).
– As the end effector orienting errors have been neglected, the orienting accuracy after calibration is only slightly damaged (Fig. 3,b).
– The implementation (error compensation) step of calibration consists in chaining the nominal values with the actual values of the coordinates of points A_i and B_i (see Fig. 2).
– A marked improvement in positioning error uniformity after calibration is obtained.

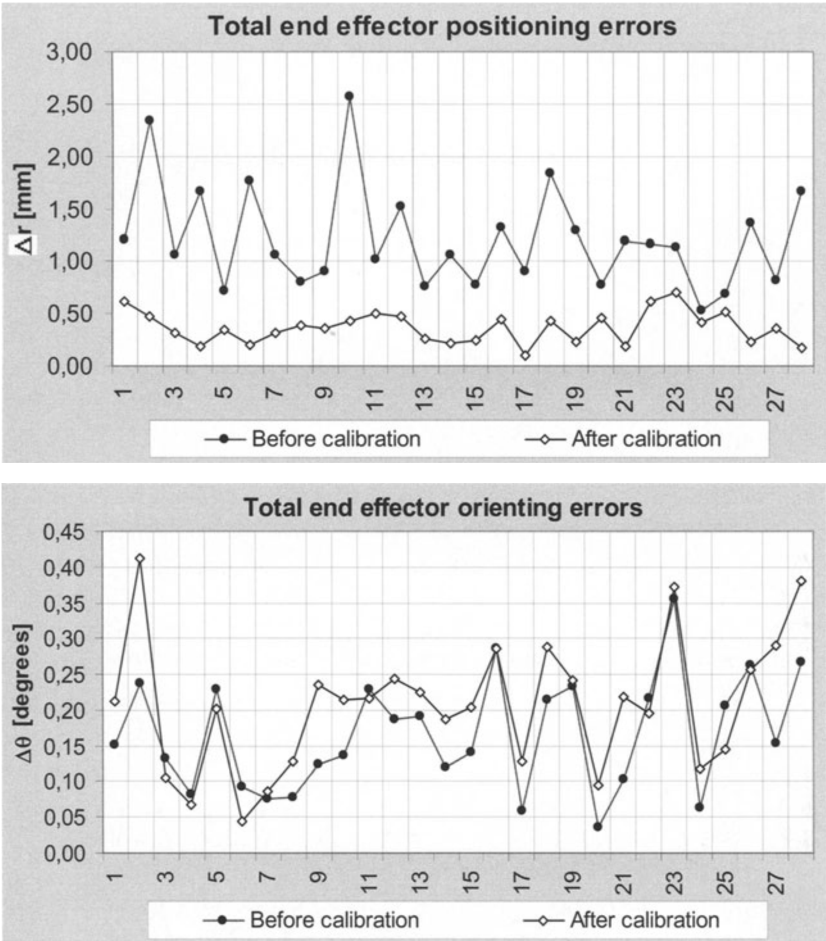


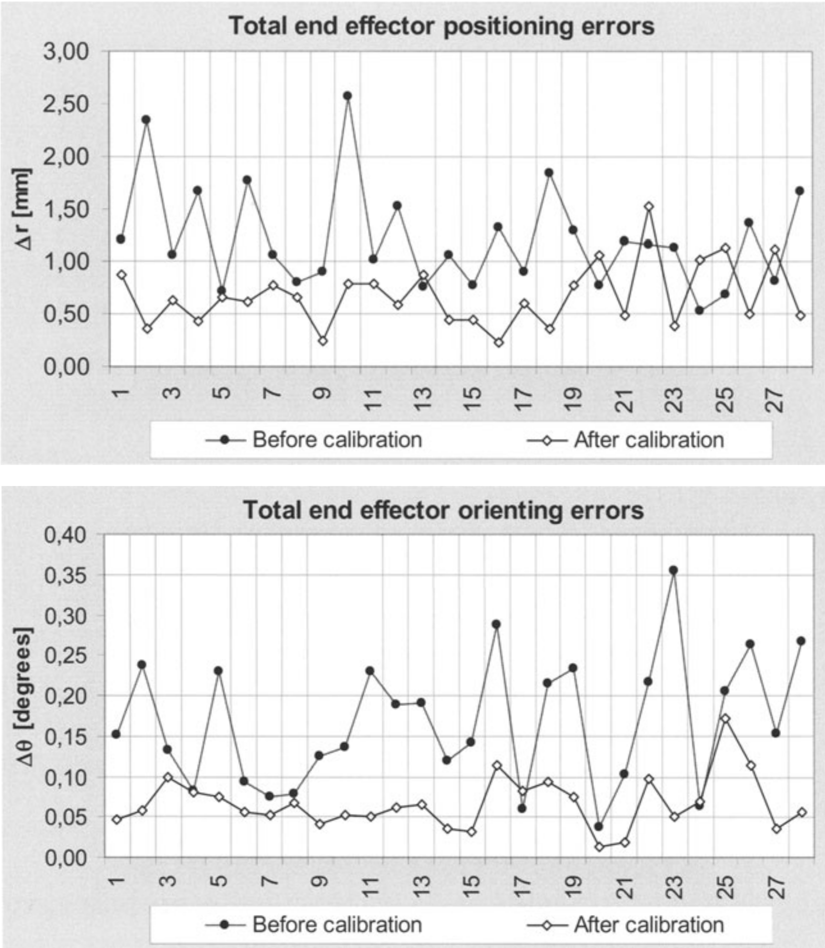
Figure -3. Kinematic calibration using only positioning errors

4.2 Kinematic calibration using end effector pose errors

Each configuration tested offers 6 scalar equations (see eq. 1); for *m* configurations we obtained a linear overdetermined system thus:

$$\begin{bmatrix} d\mathbf{X}_1 \\ d\boldsymbol{\theta}_1 \\ \dots \\ d\mathbf{X}_m \\ d\boldsymbol{\theta}_m \end{bmatrix} = \left[\mathbf{J}_G\right] \cdot \left[d\mathbf{G}\right], \ \left[\mathbf{J}_G\right] = \begin{bmatrix} \mathbf{J}_{G1} \\ \dots \\ \mathbf{J}_{Gm} \end{bmatrix}, \tag{12}$$

where $d\mathbf{X}_i = [d_{xi}, d_{yi}, d_{zi}]^T$ is the vector of end effector positioning errors, and $d\boldsymbol{\theta}_i = [\delta_{xi}, \delta_{yi}, \delta_{zi}]^T$ is the vector of orienting errors, for the i^{th} configuration tested.



a

b

Figure -4. Kinematic calibration using pose errors

In order to solve the system (12), in the *linear least squares* sense, and obtain the actual deviations of geometrical parameters, we had to apply the same steps as in the subsection 4.1. Finally, the end effector errors after calibration (Fig. 4) were established. In conclusion, the kinematic calibration in this case allowed the increase of both the positioning accuracy (~ 2 times, Fig. 4,a) and orienting accuracy (~ 2 times, Fig. 4,b). Also, we can see a marked improvement in error uniformity after calibration.

5. CONCLUSIONS

Based on the modeling and experimental testing done on a Stewart platform, regarding kinematic calibration, the following conclusions could be drawn:

a) As a result of the kinematic studies developped by the authors, a *general precision modeling* applied to *parallel robots* was proposed in the first step; this modeling was illustrated by simulation and tested on a Stewart platform.

b) On the basis of the general modeling proposed and of the general recommendations from the literature, regarding the calibration, an algorithm used in the kinematic calibration of a Stewart platform was generated and tested, with the aim of the optimization of the kinematic model and to use the actual model in the command/control process.

c) *The identification of modeling parameter deviations* implies solving a overdetermined linear system; the selection of the relevant tested–configurations and the numerical processing of the Jacobian matrix are the main difficulties of this step. Improved accuracy could be obtained only if these difficulties were solved.

d) In the case of the calibration using only the end effector positioning errors, the identification step solution requires the elimination of some difficulties related to the selection and exclusion of redundant and unidentifiable parameters. The actual kinematic model assures the increasing positioning accuracy, with little decrease of orienting accuracy; this disadvantage was eliminated by calibration using the end effector pose errors.

e) In the *error compensation* step we must only to change the nominal values of the coordinates of points A_i and B_i (see Fig. 2) with its actual values.

REFERENCES

1. R. BERNHARDT, S.L. ALBRIGHT. "Robot calibration", Chapman&Hall, 1993.
2. J.H. BORM, C.H. MENQ. "Determination of the optimal measurement configurations for robot calibration based on observability measure", Journal of Robotic Research, vol. 10(1), 1991, pp. 51-63.
3. D. DANEY. "Etalonnage géométrique des robots parallèls", PhD Thesis, Université de Nice – Sophia Antipolis, 2000.
4. M. DRIELS, U. PATHRE. "Significance of observation strategy on the design of robot calibration experiments", Journal of Robotic Systems, vol. 7(2), 1990, pp. 197-223.
5. F. DUDITA, D. DIACONESCU, G. GOGU "Linkages: inventics and kinematics in a phylogenetic approach (in Romanian), Ed. Tehnica, Bucuresti, 1989.

6. Z. GENG, L.S. HAYNES. “An effective kinematics calibration methods for Stewart platform”, In ISRAM, 1994, pp. 87-92.
7. G. GOGU, P. COIFFET, A. BARRACO. "Représentation des déplacements des robots", Ed. Hermes, Paris, 1997.
8. J. HOLLERBACH, CH. WAMPLER. "The calibration index and taxonomy for robot kinematic calibration methods", International Journal of Robotics Research, vol. 15, No. 6, 1996, pp. 573-591.
9. C. INNOCENTI. “Algorithms for kinematic calibration of fully-parallel manipulators”, In J-P. MERLET and B. RAVANI, editors, Computational Kinematics, Kluwer, 1995, pp.241-250.
10. C. INNOCENTI. “Polynomial solution of the spatial Burmester problem", ASME Journal of Mechanical Design, 117(1), 1995, pp. 64-68.
11. W. KHALIL, D. MURARECI. “Autonomous calibration of parallel robots", In Proc. 5th IFAC Symposium on Robot Control, vol. 2, 1989, pp.425-429.
12. W. KHALIL, M. GAUTIER, CH. ENGUEHARD. "Identifiable parameters and optimum configurations for robots calibration", Robotica, vol. 9, 1991, pp. 63-70.
13. J.P. MERLET. "Les robots parallèles", Ed. Hermes, Paris, 1990.
14. B. MOORING, Z. ROTH, M. DRIELS. "Fundamentals of Manipulator Calibration", John Wiley&Sons Inc., New York, 1991.
15. D. MURARECI. "Contribution à la modélisation géométrique et à l’étalonnage des robots séries et parallèles”, PhD Thesis, École Centrale de Nantes, 1997.
16. A. NAHVI, J. HOLLERBACH, V. HAYWARD. "Closed-loop kinematic of calibration a parallel-drive shoulder joint", Proc. IEEE Int. Conf. Robotics and Automation, 1994, pp. 407-412.
17. M. NEAGOE. "Contributions to the study of industrial robot precision" (in Romanian), PhD Thesis, Transilvania University of Braşov, 2001.
18. M.P. OLIVIERS, J.R.R. MAYER. "Global kinematic calibration of a Stewart platform”, ASME DSC, 57(1), 1995, pp. 129-136.
19. K. SCHRÖRER. "Theory of kinematic modeling and numerical procedures for robot calibration", In Robot calibration, Chapman & Hall, 1993, pp. 157-196.
20. S. TOYAMA, S. HATAE. "Error analysis of platform type of robot by means of screw algebra”, In 20th ISIR Int. Symp. on Industrial Robots, 1989, pp. 635-642.
21. P. VISCHER. "Improve the accuracy of parallel robot”, PhD Thesis, École Polytechnique Fédérale de Lausanne, 1996.
22. J. WANG. "Workspace evaluation and kinematic calibration of Stewart platform”, PhD Thesis, Florida Atlantic University, Boca Raton, 1992.
23. J. WANG, O. MASORY. "On the accuracy of a Stewart platform – part I: The effect of manufacturing tolerances", In IEEE Int. Conf. On Robotics and Automation, 1993, pp.114-120.
24. H. ZHUAN, Z.S. ROTH. "A method for kinematic calibration of Stewart platforms", In ASME Annual Winter Meeting, vol. 29, 1991, pp. 43-49.
25. H. ZHUAN, Z.S. ROTH. "Method for kinematic calibration of Stewart platforms", In Journal of Robotic Systems, 10(3), 1993, pp. 391-405.
26. H. ZHUANG, O. MASORY, Y. JIAHUA. "Kinematic calibration of a Stewart platform using pose measurements obtained by a single theodolite", In IEEE Int. Conf. on Intelligent Robots and Systems, 1995, pp.339-334.
27. H. ZHUANG, Y. JIAHUA, O. MASORY. "Calibration of Stewart platforms and other parallel manipulators by minimizing inverse kinematic residuals", Journal of Robotic Systems, 17(7), 1998, pp. 395-405.

THE ISOCONDITIONING LOCI OF PLANAR THREE-DOF PARALLEL MANIPULATORS

S. Caro*, D. Chablat*, P. Wenger* and J. Angeles**
*IRCCyN, Ecole Centrale de Nantes, France
**Department of Mechanical Engineering & Center for Intelligent Machines, McGill University, Montreal, Canada

Abstract: This paper deals with a special class of three-degree-of-freedom parallel manipulators. First, the singular configurations of the two Jacobian matrices are studied. The isotropic configurations are then found based on the characteristic length of this manipulator. The isoconditioning loci for the Jacobian matrices are plotted to define a global performance index allowing the comparison of the different working modes. The resulting index is compared with the Cartesian workspace surface and the average of the condition number.

Key words: parallel manipulator, optimum design, isoconditioning loci, characteristic length

1. INTRODUCTION

Various performance indices have been devised to assess the kinetostatic performances of serial and parallel manipulators. As the literature on performance indices is extremely rich to fit in the limits of this paper, the interested reader is invited to look at it in the rather recent references cited here. A dimensionless quality index was recently introduced by Lee, Duffy, and Hunt [1] based on the ratio of the Jacobian determinant to its maximum absolute value, as applicable to parallel manipulators.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 129–138.

This index does not take into account the location of the operation point of the end-effector, because the Jacobian determinant is independent of this location. The proof of the foregoing result is available in [2], as pertaining to serial manipulators, its extension to their parallel counterparts being straightforward. The *condition number* of a given matrix, on the other hand, is known to provide a measure of invertibility of the matrix [3]. It is thus natural that this concept found its way into this context. Indeed, the condition number of the Jacobian matrix was proposed by [4] as a figure of merit to minimize when designing manipulators for maximum accuracy. In fact, the condition number gives, for a square matrix, a measure of the relative roundoff-error amplification of the computed results [3] with respect to the data roundoff error. As is well known, however, the dimensional heterogeneity of the entries of the Jacobian matrix prevents the straightforward application of the condition number as a measure of Jacobian invertibility. The *characteristic length* was introduced in [5] to cope with the above-mentioned heterogeneity.

In this paper, we use the *characteristic length* to normalize the Jacobian matrix of a three-degree-of-freedom (dof) planar manipulator and to calculate the isoconditioning loci for all working modes.

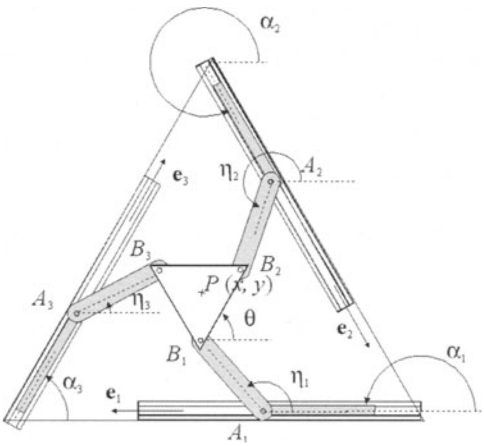


Figure -1. A three –DOF parallel manipulator

2. PRELIMINARIES

A planar three-dof manipulator comprising three parallel PRR chains is shown in Fig. 1. This manipulator has been frequently studied, in particular in [6,7]. The actuated joint variables are the displacements of the three prismatic joints, the Cartesian variables being position vector **p** of operation point P and orientation θ of the platform. The trajectories of points A_i define an equilateral triangle whose geometric center is point O, while points B_1, B_2 and B_3, whose geometric center is point P, lie at the corners of an equilateral triangle. We thus have $\alpha_i = \pi + (i\text{-}1)(2\pi/3)$, for $i = 1,2,3$. Moreover, $l = l_1 = l_2 = l_3$, with l_i denoting the length of A_iB_i and $r = r_1 = r_2 = r_3$, with r_i denoting the length of B_iP, in units of length that need not be specified in the paper. The layout of the trajectories of points A_i is defined by radius R of the circle inscribed in the associated triangle.

2.1 Kinematic Relations

Velocity ˙,**p** of point P can be obtained in three different forms, depending on which leg is traversed, namely,

$$\dot{\mathbf{p}} = \dot{\mathbf{a}}_i + \dot{\eta}_i\mathbf{E}(\mathbf{b}_i - \mathbf{a}_i) + \dot{\theta}\mathbf{E}(\mathbf{p} - \mathbf{b}_i),\ i \in [1,3] \tag{1}$$

with matrix $\mathbf{E}$ and velocity $\dot{\mathbf{a}}_i$ of points A_i defined as

$$\mathbf{E} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},\quad \dot{\mathbf{a}}_i = \dot{\boldsymbol{\rho}}_i \frac{\rho_i}{\|\rho_i\|} = \dot{\boldsymbol{\rho}}_i \begin{bmatrix} \cos(\alpha_i) \\ \sin(\alpha_i) \end{bmatrix} = \dot{\boldsymbol{\rho}}_i \mathbf{e}_i$$

where $\mathbf{e}_i$ is a unit vector in the direction of the ith prismatic joint. We would like to eliminate the three idle joint rates $\dot{\eta}_1$, $\dot{\eta}_2$ and $\dot{\eta}_3$ from eq.(1), which we do upon dot-multiplying the former by $(b_i - a_i)^T$, thus obtaining

$$(\mathbf{b}_i - \mathbf{a}_i)^T \dot{\mathbf{p}} = (\mathbf{b}_i - \mathbf{a}_i)^T \dot{\mathbf{a}}_i + (\mathbf{b}_i - \mathbf{a}_i)^T \dot{\theta}\mathbf{E}(\mathbf{p} - \mathbf{b}_i),\ i \in [1,3] \tag{2}$$

Equation (2) can now be cast in vector form, namely,

$$\mathbf{At} = \mathbf{B}\dot{\boldsymbol{\rho}} \text{ with } \mathbf{t} = \begin{bmatrix} \dot{\mathbf{p}} & \dot{\theta} \end{bmatrix}^T \text{ and } \dot{\boldsymbol{\rho}} = \begin{bmatrix} \dot{\rho}_1 & \dot{\rho}_2 & \dot{\rho}_3 \end{bmatrix}^T \tag{3}$$

with $\dot{\boldsymbol{\rho}}$ thus being the vector of actuated joint rates. Moreover, **A** and **B** are, respectively, the direct-kinematics and the inverse-kinematics matrices of the manipulator, defined as

$$\mathbf{A} = \begin{bmatrix} (\mathbf{b}_1 - \mathbf{a}_1)^T & -(\mathbf{b}_1 - \mathbf{a}_1)^T \mathbf{E}(\mathbf{p} - \mathbf{b}_1) \\ (\mathbf{b}_2 - \mathbf{a}_2)^T & -(\mathbf{b}_2 - \mathbf{a}_2)^T \mathbf{E}(\mathbf{p} - \mathbf{b}_2) \\ (\mathbf{b}_3 - \mathbf{a}_3)^T & -(\mathbf{b}_3 - \mathbf{a}_3)^T \mathbf{E}(\mathbf{p} - \mathbf{b}_3) \end{bmatrix} \tag{4a}$$

$$\mathbf{B} = \begin{bmatrix} (\mathbf{b}_1 - \mathbf{a}_1)^T \mathbf{e}_1 & 0 & 0 \\ 0 & (\mathbf{b}_2 - \mathbf{a}_2)^T \mathbf{e}_2 & 0 \\ 0 & 0 & (\mathbf{b}_2 - \mathbf{a}_2)^T \mathbf{e}_2 \end{bmatrix} \tag{4b}$$

When **A** and **B** are nonsingular, we obtain the relations

$$\mathbf{t} = \mathbf{J}\dot{\boldsymbol{\rho}} \text{ with } \mathbf{J} = \mathbf{A}^{-1}\mathbf{B} \text{ and } \dot{\boldsymbol{\rho}} = \mathbf{Kt}, \mathbf{K} = \mathbf{B}^{-1}\mathbf{A}$$

with **K** denoting the inverse of **J**.

2.2 Parallel Singularities

Parallel singularities occur when the determinant of matrix **A** vanishes [8,9]. At these configurations, it is possible to move locally operation point P with the actuators locked, the structure thus resulting cannot resist arbitrary forces, and control is lost. To avoid any performance loss, it is necessary to have a Cartesian workspace free of parallel singularities. For the planar manipulator studied, such configurations are reached whenever the axes A_1B_1, A_2B_2 and A_3B_3 intersect (possibly at infinity), as depicted in Fig.2.

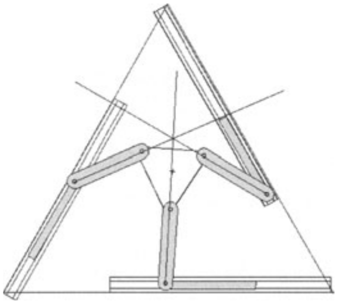


Figure -2. Parallel singularity

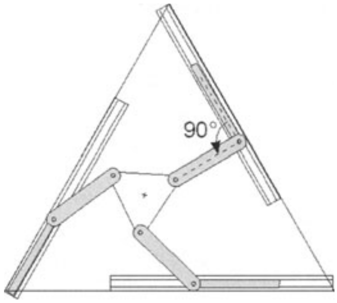


Figure -3. Serial singularity

In the presence of such configurations, moreover, the manipulator cannot resist a force applied at the intersection point. These configurations are located inside the Cartesian workspace and form the boundaries of the joint workspace [8].

2.3 Serial Singularities

Serial singularities occur when $\det(\mathbf{B}) = 0$. In the presence of theses singularities, there is a direction along which no Cartesian velocity can be produced. Serial singularities define the boundary of the Cartesian workspace. For the topology under study, the serial singularities occur whenever $(\mathbf{b}_i - \mathbf{a}_i)^T \mathbf{e}_i = 0$ for at least one value of i, as depicted in Fig.3 for $i = 2$.

3. ISOCONDITIONING LOCI

3.1 The Matrix Condition Number

We derive below the loci of equal condition number of the matrices **A**, **B** and **K**. To do this, we first recall the definition of condition number of an $m \times n$ matrix **M**, with $m<n$, $\kappa(\mathbf{M})$. Using Frobenius norm, $\kappa(\mathbf{M})$ is the ratio of the smallest, σ_s, to the largest, σ_l, singular values of **M**, namely,

$$\kappa(\mathbf{M}) = \sigma_s / \sigma_l \tag{5}$$

The singular values of matrix **M** are defined, in turn, as the square roots of the nonnegative eigenvalues of the positive-definite $m \times m$ matrix $\mathbf{M}\mathbf{M}^T$.

3.2 Non-Homogeneous Direct-Kinematics Matrix

To make matrix **A** homogeneous, as needed to define its condition number, each term in the third column of **A** is divided by the characteristic length L [10], thereby deriving its normalized counterpart $\overline{\mathbf{A}}$:

$$\overline{\mathbf{A}} = \begin{bmatrix} (\mathbf{b}_1 - \mathbf{a}_1)^T & -(\mathbf{b}_1 - \mathbf{a}_1)^T \mathbf{E}(\mathbf{p} - \mathbf{b}_1)/L \\ (\mathbf{b}_2 - \mathbf{a}_2)^T & -(\mathbf{b}_2 - \mathbf{a}_2)^T \mathbf{E}(\mathbf{p} - \mathbf{b}_2)/L \\ (\mathbf{b}_3 - \mathbf{a}_3)^T & -(\mathbf{b}_3 - \mathbf{a}_3)^T \mathbf{E}(\mathbf{p} - \mathbf{b}_3)/L \end{bmatrix} \tag{6}$$

which is calculated so as to minimize $\kappa(\overline{\mathbf{A}})$, along the posture variables ρ_1, ρ_2 and ρ_3. However, notice that **B** is dimensionally homogeneous, and does not need to be normalized.

3.3 Isotropic Configuration

In this section, we derive the isotropy condition on **J** to define the geometric parameters of the manipulator. We shall obtain also the value L of the characteristic length. To simplify $\overline{\mathbf{A}}$ and **B**, we use the notation

$$\mathbf{l}_i = (\mathbf{b}_i - \mathbf{a}_i) \tag{7a}$$

$$k_i = (\mathbf{b}_i - \mathbf{a}_i)^T \mathbf{E}(\mathbf{p} - \mathbf{b}_i) \tag{7b}$$

$$m_i = (\mathbf{b}_i - \mathbf{a}_i)^T \mathbf{e}_i \tag{7c}$$

$$\gamma_i = \angle A_i B_i P \tag{7d}$$

We can thus write matrices $\overline{\mathbf{A}}$ and **B** as

$$\overline{\mathbf{A}} = \begin{bmatrix} \mathbf{l}_1^T & -k_1/L \\ \mathbf{l}_2^T & -k_2/L \\ \mathbf{l}_3^T & -k_3/L \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} \tag{8}$$

Whenever matrix **B** is nonsingular, that is, when $m_i \neq 0$, for $i = 1,2,3$, we obtain

$$\overline{\mathbf{K}} = \begin{bmatrix} \mathbf{l}_1^T / m_1 & -k_1/(L\, m_1) \\ \mathbf{l}_2^T / m_2 & -k_2/(L\, m_2) \\ \mathbf{l}_3^T / m_3 & -k_3/(L\, m_3) \end{bmatrix}$$

Matrix $\overline{\mathbf{J}}$, the normalized $\mathbf{J}$, is isotropic if and only if $\overline{\mathbf{K}}\,\overline{\mathbf{K}}^T = \tau^2 \mathbf{1}_{3\times3}$ for $\tau > 0$ and $\overline{\mathbf{K}} = \overline{\mathbf{J}}^{-1}$,i.e.,

$$(\mathbf{l}_i^T \mathbf{l}_i + k_i^2 / L^2)/m_i^2 = \tau^2, \;\; i \in [1,3] \tag{9a}$$

$$(\mathbf{l}_i^T \mathbf{l}_j + k_i k_j / L^2)/(m_i m_j) = 0, \;\; i \neq j, \;\; i, j \in [1,3] \tag{9b}$$

From eqs.(9a-b), we can derive the conditions below:

$$\|\mathbf{l}_1\| = \|\mathbf{l}_2\| = \|\mathbf{l}_3\| \tag{10a}$$

$$\|\mathbf{p} - \mathbf{b}_1\| = \|\mathbf{p} - \mathbf{b}_2\| = \|\mathbf{p} - \mathbf{b}_3\| \tag{10b}$$

$$\mathbf{l}_1^T \mathbf{l}_2 = \mathbf{l}_2^T \mathbf{l}_3 = \mathbf{l}_3^T \mathbf{l}_1 \tag{10c}$$

$$m_1 m_2 = m_2 m_3 = m_3 m_1 \tag{10d}$$

In summary, the constraints defined in the eqs.(10a-d) are:

- Pivots B_i should be placed at the vertices of an equilateral triangle;
- Segments A_iB_i form an equilateral triangle;
- The trajectories of point A_i define an equilateral triangle. Hence, $l = l_1 = l_2 = l_3$.

Notice that the foregoing conditions, except the second one, were assumed in chapter 2.

3.4 The Characteristic Length

The characteristic length is defined at the isotropic configuration. From eq.(9b), we determine the value of the characteristic length as,

$$L = \sqrt{-k_1 k_2 / (\mathbf{l}_1^T \mathbf{l}_2)}$$

By applying the constraints defined in eq.(9a), we can write characteristic length L in terms of angle γ, i.e.

$$L = \sqrt{2} r \sin(\gamma)$$

where $\gamma = \gamma_1 = \gamma_2 = \gamma_3$, was defined in eq.(7d) and $\gamma \in [0, 2\pi]$.

This means that the manipulator under study admits several isotropic configurations, two of which are shown in Fig.4a and Fig.4b, whereas the characteristic length L of a manipulator is unique [2]. When γ is equal to $\pi/2$, Fig.4a, *i.e.,* when A_iB_i is perpendicular to B_iP, the manipulator finds itself at a configuration furthest away from parallel singularities. To have an isotropic configuration furthest away from serial singularities, we obtain two conditions: $\mathbf{e}_i^T \mathbf{E}(\mathbf{b}_i - \mathbf{a}_i) = 0$ and $r = R/2$.

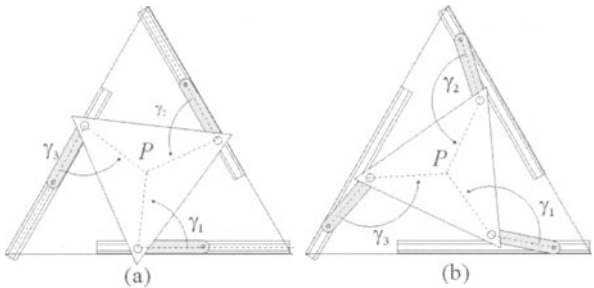


Figure -4. Two isotropic configurations with two values of ρ

3.5 Working Modes

The manipulator under study has a diagonal inverse-kinematics matrix **B**, as shown in eq.(4b), the disappearance of one of its diagonal entries thus indicating the occurrence of a serial singularity. The set of manipulator postures free of this kind of singularity is termed a working mode. The different working modes are thus separated by a serial singularity. The formal definition of the working mode is detailed in [8]. For the manipulator at hand, there are eight working modes, as depicted in Fig.5.

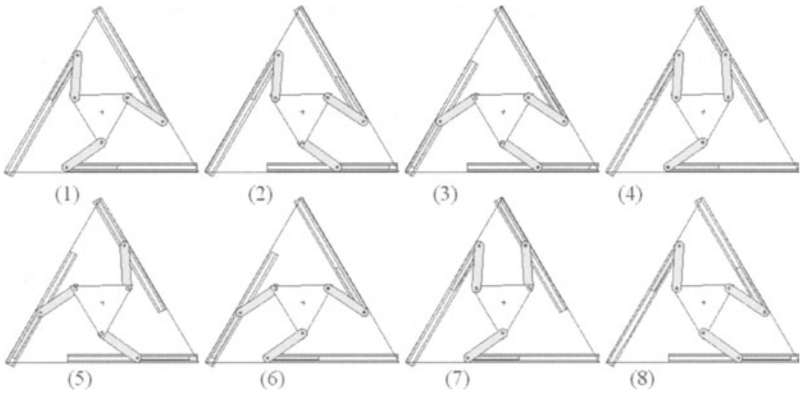


Figure -5. The eight working modes of the 3PRR manipulator

However, because of symmetries, we can restrict our study to only two working modes, if there are no joint limits. Indeed, working mode 1 is similar to working mode 5. For the first one, the signs of the diagonal entries of **B** are all negative. For the second one, they are all positive. A similarly reasoning is applicable to working modes 2-6, 3-7 and 4-8. Likewise, working modes 3-4 and 7-8 can be derived from working modes 2 and 6 by a rotation of 120° and 240°, respectively. Therefore, only working modes 1 and 2 are studied. We label the corresponding matrices as $\overline{\mathbf{A}}_i$, $\mathbf{B}$, $\overline{\mathbf{K}}_i$ for *i*th working mode.

3.6 Isoconditioning Loci

For each Jacobian matrix and for all the poses of the end-effector, we calculate the optimum conditioning according to the orientation of the end-effector. We can notice that there is a singular configuration for any orientation of the end-effector.

Fig.6 depicts the isoconditioning loci of matrix $\overline{\mathbf{A}}$. We depict in Fig.7 the isoconditioning loci of matrix $\mathbf{B}$. We notice that the loci of both working modes are identical. This is due to both the absence of joint limits on the actuated joints and the symmetry of the manipulator. For one configuration, only the signs of m_i change from a working mode to another, but the condition number κ is computed from the absolute values of m_i. The shapes of the isoconditioning loci of $\overline{\mathbf{K}}$ are similar to those of the isoconditioning loci of $\overline{\mathbf{A}}$; only the numerical values vary.

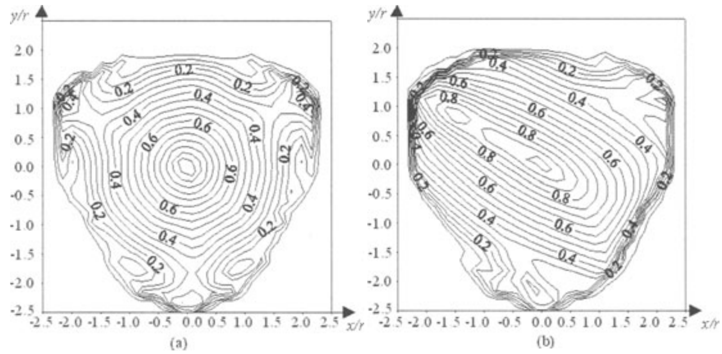


Figure -6. Isoconditioning loci of the matrix (a) $\overline{\mathbf{A}}_1$ and (b) $\overline{\mathbf{A}}_2$ with $R/r = 2$ and $l/r = 2$

For the first working mode, the condition number of matrix $\overline{\mathbf{A}}$ decreases regularly around the isotropic configuration. The isoconditioning loci resemble concentric circles. However, for the second working mode, the isoconditioning loci of matrix $\overline{\mathbf{A}}$ resemble ellipses. Characteristic length L depends on r. Two indices can be studied according to parameter R:

(i) the area of the Cartesian workspace, called S;

(ii) the average of the conditioning, called κ.

The first index is identical for the two working modes. Fig. 8 depicts the variation of S as a function of R/r, for $l/r = 2$. Its maximum value is reached when $R/r = 0.5$.

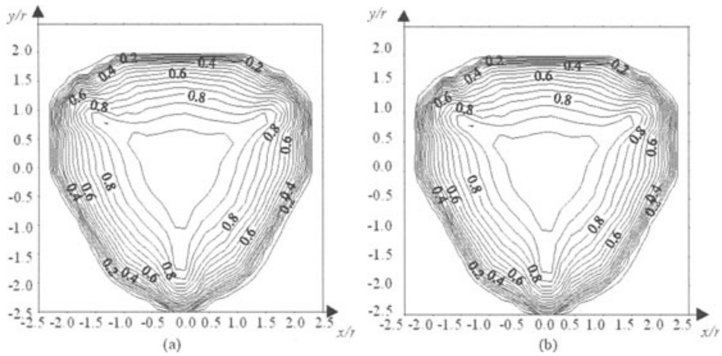


Figure -7. Isoconditioning loci of the matrix (a) $\mathbf{B}_1$ and (b) $\mathbf{B}_2$ with $R/r = 2$ and $l/r = 2$

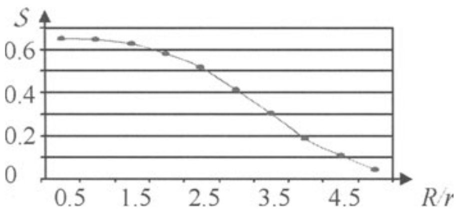


Figure -8. $S = f(\mathrm{R/r})$

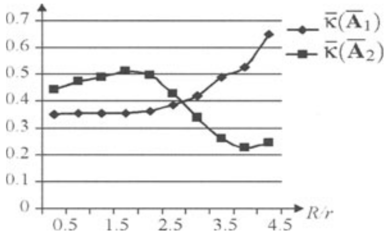


Figure -9. $\bar{\kappa}(\bar{\mathbf{A}}_1)$, $\bar{\kappa}(\bar{\mathbf{A}}_2) = f(\mathrm{R/r})$

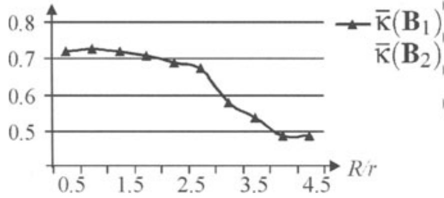


Figure -10. $\bar{\kappa}(\mathbf{B}_1)$, $\bar{\kappa}(\mathbf{B}_2) = f(\mathrm{R/r})$

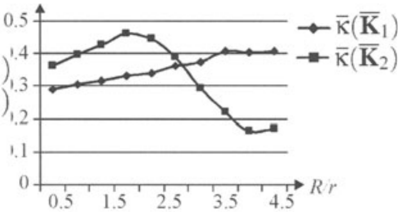


Figure -11. $\bar{\kappa}(\bar{\mathbf{K}}_1)$, $\bar{\kappa}(\bar{\mathbf{K}}_2) = f(\mathrm{R/r})$

For the three matrices studied, $\bar{\kappa}$ can be regarded as a global performance index. This index allows us to compare the working modes. Figures.9, 10 and 11 depict $\bar{\kappa}(\bar{\mathbf{A}})$, $\bar{\kappa}(\mathbf{B})$ and $\bar{\kappa}(\bar{\mathbf{K}})$, respectively, as a function of *R/r*, with $l/r = 2$.

The value of $\bar{\kappa}(\bar{\mathbf{A}}_1)$ increases with *R*. At the opposite, the maximum value of $\bar{\kappa}(\bar{\mathbf{A}}_2)$ is reached when $R/r = 2$. For both the working modes studied, $\bar{\kappa}(\mathbf{B}_1)$ and $\bar{\kappa}(\mathbf{B}_2)$ are identical for *R/r* fixed. For the first working mode, the minimum value of $\bar{\kappa}(\bar{\mathbf{K}}_1)$ and the maximum area of Cartesian workspace *S* occur at different values of *R/r*. This means that we must reach a compromise under these two indices. For the second working mode, there is an optimum of $\bar{\kappa}(\bar{\mathbf{K}}_2)$ close to the optimum of *S*, for $R/r = 2$.

4. CONCLUSIONS

We produced the isoconditioning loci of the Jacobian matrices of a three-PRR parallel manipulator. This concept being general, it can be applied to any three-dof planar parallel manipulator. To solve the problem of heterogeneity of the Jacobian matrix, we used the notion of characteristic length. This length was defined for the isotropic configuration of the manipulator. The isoconditioning curves thus obtained characterize, for every posture of the manipulator, the optimum conditioning for all possible orientations of the end-effector. This index is compared with the area of the Cartesian workspace and the conditioning average. The two optima being different, it is necessary to find another index to determine the optimum values. The results of this paper can be used to choose the working mode which is best suited to the task at hand or as a global performance index when we study the optimum design of these kinds of manipulators.

REFERENCES

[1] Lee J., Duffy J. and Hunt K., *A Pratical Quality Index Based on the Octahedral Manipulator*, The International Journal of Robotic Research, Vol. 17, No. 10, October, 1998, 1081-1090.

[2] Angeles J., *Fundamentals of Robotic Mechanical Systems*, Springer-Verlag, Second Edition, New York, 2002.

[3] Golub G. H. and Van Loan C. F., *Matrix Computations*, The John Hopkins University Press, Baltimore, 1989.

[4] Salisbury J. K. and Craig J. J., *Articulated Hands: Force Control and Kinematic Issues*, The International Journal of Robotics Research, 1982, Vol. 1, No. 1, pp. 4-17.

[5] Angeles J. and Lopez-Cajun C. S., *Kinematic Isotropy and the Conditioning Index of Serial Manipulators*, The Int. J. of Robotics Research, 1992, Vol. 11, No. 6, pp. 560-571.

[6] Merlet J.P., *Parallel robots*, Kluwer Academic Publ., Dordrecht, The Netherland., 2000.

[7] Gosselin C., Sefrioui J., Richard M. J., *Solutions polynomiales au problème de la cinématique des manipulateurs parallèles plans à trois degré de liberté*, Mechanism and Machine Theory, Vol. 27, pp. 107-119, 1992.

[8] Chablat D. and Wenger P., *Working Modes and Aspects in Fully-Parallel Manipulator*, Proceedings IEEE Inter. Conference on Robotics and Automation, 1998, pp. 1964-1969.

[9] Gosselin C. and Angeles J., *Singularity Analysis of Closed-Loop Kinematic Chains*, IEEE, Transaction on Robotics and Automation, Vol. 6, pp. 281-290, June 1990.

[10] Ranjbaran F., Angeles J., Gonzalez-Palacios M.A. and Patel R.V., *The Mechanical Design of a Seven-Axes Manipulator with Kinematic Isotropy*, Journal of Intelligent and Robotic Systems, 1995, Vol. 14, pp. 21-41.

DEVELOPMENT OF AN OPTIMISED AUTOMOTIVE AXLE

L. Heuzé*, P. Ray*, G. Gogu*, L. Serra** and F. André**
**Laboratoire de Recherche et Applications en Mécanique Avancé, Institut Français de Mécanique Avancée et Université Blaise Pascal, Clermont Ferrand, France*
*** Centre de Technologies Michelin, Clermont Ferrand, France*

Abstract: The elastokinematic performances of automotive axles guarantee driveability and security of passengers, also giving vehicles improved road-holding properties as well as increasing or reducing briskness. The search for improved performances leads to optimisation of mechanism kinematics. It also concerns integration of suitable bushings, which are elements in standard cars that are vital for passenger comfort in that they filter out vibrations. Such optimisation is needed since the design does not incorporate active suspension technologies. This paper presents a fast and efficient method which has been applied to the multi-criteria problem characterised by a large number of variables. It takes into account intended judicious specifications as well as inherent constraints to incorporation in the vehicle.

Key words: optimisation, automotive, axle, modelling, elastokinematic

1. INTRODUCTION

The aim of automotive axle design is to provide comfort for car passengers by means of optimum filtering of vibrations and running noises. This is done by integrating suitable rubber bushings in the structure. The passengers' comfort also involves limitation of body car movement, mainly rolling. Lastly, the design also includes determining the best road-holding properties, i.e. the ability of the car to follow the intended path. In actual fact, the suspension characteristics must match the intended ones in such a way that they lead to the desired behaviour, such as over-steering or under-

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 139–148.

steering. With all these properties in mind, research has mainly been directed at controlling the steering and camber angles, half-track variation and roll centre position under all restrictive conditions (i.e. lateral and longitudinal loads, vertical clearance etc.) [1], [2]. In fact, vehicle behaviour, vertical load transfer and tyre working conditions depend on these conditions [3].

Various optimisation techniques and models have been developed taking into account of the applied technologies on suspension mechanisms and the numerous parameters which are to be taken into consideration [4], [5], [6]. The principal objective of the present study is to apply a fast and efficient method to obtain the best elastokinematic axle characteristics. From this viewpoint, the task is based on the intention to model the studied vehicle on a kinematics and compliance measuring bench. If judicious specifications are to be respected and inherent constraints in the set up process to be taken into account, the modelling of the car and the optimisation method must provide the best configuration for the axle parameters.

2. STUDIED DESIGN PRESENTATION

The purpose of the study is to provide an optimised configuration of the Michelin *OCP* system (*Optimised Contact Patch*) (figure 1).

The design features an innovative rear axle to create an optimised contact patch between the tyre and the ground [7]. In fact, to date, conventional axles have demonstrated the same problem. On corners, the car tilts as a result of the roll effect. Since tyres follow the body car movement, they are thus no longer perpendicular to the road.

Consequently, the contact surface area is reduced, which means that perfect road-holding properties can no longer be guaranteed (figure 2) [8].

The principle comprises a means to uncouple vehicle roll movements from wheel leaning. In fact, it guarantees an optimum contact patch on corners and on straight lines using the following method : the larger the lateral force, the more the wheel leans towards the inside of the bend.

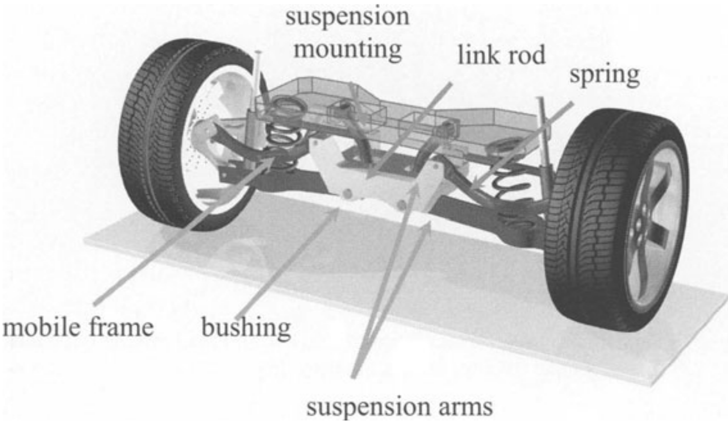


Figure -1. Michelin OCP system

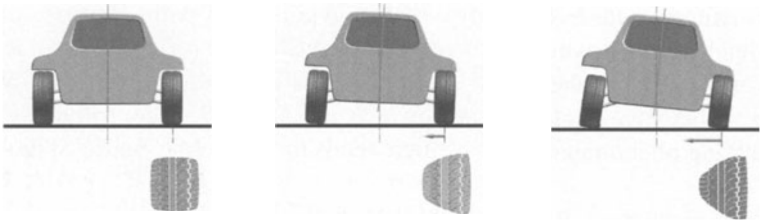


Figure -2. Effects of change in direction on contact surface

3. AXLE MODELLING

3.1 Working hypotheses

The main way to perform an accurate study of the elastokinematic features of an axle is to prepare a model of the axle mounted on a vehicle on a kinematics and compliance measuring bench (figure 3). Determination of the elastokinematic features is very important, as it affects the dynamic behaviour of the vehicle.

The testing device provides a means of determining the different forces, couplings, shifting and rotation of the contact patch between the tyres and the ground, as well as rotation and displacement of the wheel centre and the wheel rim plane. These data are obtained by constraining the vehicle axle by means of forces and shifting through the contact patch while the car body is fastened to the ground.

Figure -3. Testing device used for measuring forces and shifting between tyres and ground through the contact patch [9]

The measured values allows the engineer to anticipate typical features of vehicle driveability. For example, let us consider lateral force testing applied in the same direction through the contact patch on the both wheels of a rear axle. This test reproduces tyre loads in a bend, with ground lateral forces being oriented towards the centre of the hypothetical path. Because of the lateral load, the two wheels are vertically constrained (effect of the jacking forces), in particular the outer wheels. In this situation, if the outer-wheel steering is negative (i.e. steering towards the outside), it contributes to the over-steering phenomenon : the vehicle tends to veer to the centre of the path (figure 4).

In such a situation, it is apparent that small variations in parameters such as steering should create a significant change in driveability conditions. It is therefore necessary to try to obtain judicious elastokinematic specifications. Moreover, all types of vehicle loads (roll, lateral and longitudinal loads etc.) have to be taken into account. Among such specifications, we can mention searching for the minimum track variation under lateral forces or with bumps and rebounds in order to gain better control over the vehicle's behaviour.

3.2 Geometric constraints

The principal constraint in applying a suspension mechanism to a vehicle consists in considering the available volume under the car body. Suspension movements must be compatible with the environment [10]. In particular, we have to take into account elements such as the exhaust system, the petrol tank for rear suspension, or the transmission and engine block for front suspension. Beyond the static point of view, analysis of the constraints leads to a study of the volume swept by the suspension elements during their movements, to avoid interference between them, in particular during vertical wheel motion. Lastly, it is necessary to maintain adequate ground clearance (figure 5). For example, it must remain possible to park the vehicle straddling the pavement.

To sum up, the initial constraints define the range of optimisation geometric parameters.

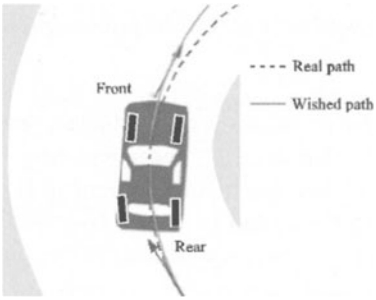


Figure -4. Over-steering

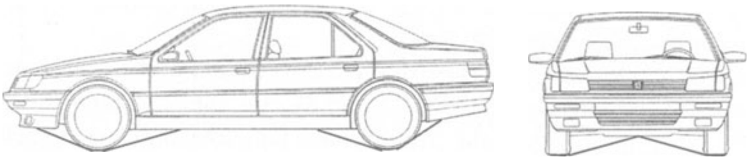


Figure -5. Packaging constraints

4. OPTIMISED AUTOMOTIVE AXLE

The optimisation operation is organised in several stages. Firstly, the Michelin *OCP* system can be considered from a two-dimensional aspect (figure 6). In this way it is possible to constitute the initial elements for construction of a more complex three-dimensional model. The modelling work is performed using *Adams* software.

After defining the general geometry, the first task is to consider the initial optimisation parameters, which define the axle geometry. All of these parameters define the vertical and horizontal positions of the mechanism links and take into account the geometric constraints explained above.

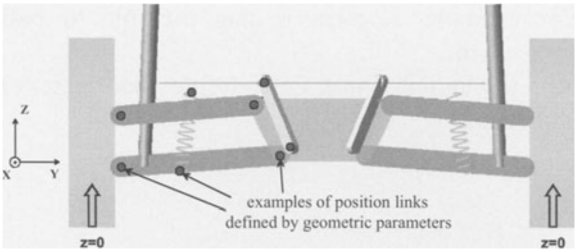


Figure -6. : Two-dimensional Adams model with vertical constraints

Once the parameter default values are ascertained, we must constrain the model in a similar way to that used on the measuring bench. The vertical position of the tyre bases is then established (tyre grip conditions) (figure 6). The suspension springs are preloaded in such a way that the vertical load of the tyres is equal to the initial experimental load for the vehicle on the bench. Gravity is considered as being equivalent to zero for the model, as the element weight has already been taken into account. In fact, the vertical tyre load already obtained is the result of the weight of the sprung and unsprung masses.

The last stage before the optimisation campaign is designed to obtain the correct wheel centre stiffness, which is an important parameter on which the passengers' comfort depends. This is obtained by determining the correct rigidity value for the suspension springs.

The intended elastokinematic specifications are defined while searching for better tyre working condition (figure 2) and the desired driveability (figure 4). We then have to take into account judicious objective functions. Among these studied functions, camber and half-track variation analyses under lateral forces or vertical motion can be mentioned (figure 7).

Optimisation begins with a design study for each of the m loads and each of the s objective functions $(g_j)_{j=1,\ldots,s}$ to determine the most influential input parameters. At this stage, one parameter is variable and the value of the other parameters is fixed. We use a static simulation for each stage of the parameter variation and we compare the results obtained with the ones obtained in the previous stage. The design study thus provides sufficient elements to determine the n influential input parameters, called $(x_i)_{i=1,\ldots,n}$, considered as optimisation variables.

The second stage of the optimisation is a combination of one or several designs of experiments (DOE), also performed for each loading case k (k=1,…, m) [11]. With respect to intended specifications, we must maintain each objective function g_j within the intended range $[a_{jk}, b_{jk}]$. Each load study is associated with a specific objective function, such as camber under lateral forces or half-track variation under bump-rebound motions. Other functions will define the optimisation constraints. Let us consider the following generic form for the studied loading case:

$$g_k = f(x_1, x_2, \ldots, x_n) \tag{1}$$

In DOE use, we must control the constraint values, such as steering under lateral force. The DOE comprises the following:

- a principal function to minimise like $h_k = \left| g_k - \dfrac{a_{jk} + b_{jk}}{2} \right|$
- constraints written as $g_{j \neq k} - b_{jk} \leq 0$ and $a_{jk} - g_{j \neq k} \leq 0$.

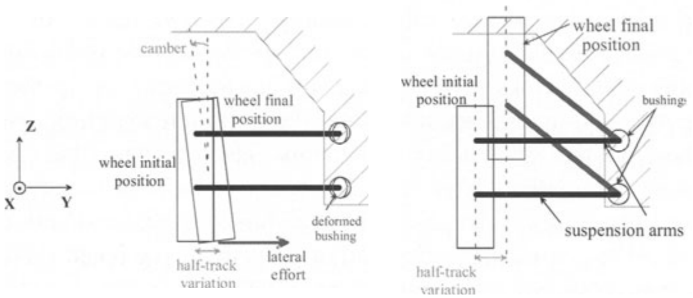


Figure -7. : Main features analysed under lateral loads and vertical motions

Because of the mechanical model complexity, we cannot write each objective function g_j in an analytical form. The purpose is not to obtain an analytical form of g_k, from which we should obtain a global optimum, but rather a DOE must help us explore the whole space described by parameter values and obtain the best configuration (x_1, x_2, …, x_n). If the selected trial performed in an initial DOE produces the best results, but does not respect the intended specifications, the optimum search operation can be repeated using another DOE with variation ranges closer to the previous optimum, until satisfactory results are obtained. For the purposes of our study, we preferred to use the full factorial design technique. Indeed, using a 550 MHz Pentium III processor, it is possible to perform 7776 trials over a period of 3 hours, i.e. trials using 5 parameters and 6 levels.

For a DOE, we could try to write the objective function in a form similar to the following, since the number of coefficients to be determined is less than or equal to the number of experiments:

$$g_k(x_1, \ldots, x_n) = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i + \sum_{i=1}^{n} \sum_{j>i}^{n} \alpha_{ij} x_i x_j \tag{2}$$

Determination of these coefficients should be carried out using the least squares method. Its setting is particularly long, as it must be performed for each load. Another solution would be to use a less complex form of the objective function, to avoid the need to take into account so many coefficients. However, a mechanism such as a suspension system is characterised by the number of interactions between geometric parameters. We also prefer to use only DOE, whereas manual trial selection would also be a swift method. In this way very good results are obtained quickly. Other optimisation tools, such as genetic algorithms or classic methods, would not be more suitable because of the need for a longer development time or the risk of obtaining optimum results only locally [4].

The 3D kinematic and the 3D elastokinematic studies are the continuation of the optimisation process, completing the model. They can be performed either by adding other elements, which have not yet been considered, such as longitudinal suspension arms, steering rods (figure 8) or by replacing perfect links of the model by elastic bushings. In these cases, the spring settings, the design study and the optimum search operations are all similar. Moreover, adding elements also means that additional optimisation parameters have to be taken into consideration. These parameters define new link positions and bushing stiffness values. After performing all stages of the optimisation procedure, we reach the optimum solution for the suspension mechanism parameters.

5. RESULTS

For each graph obtained by the elastokinematic study, specific analysis is performed on the range of output parameters and also on the slope of each curve. In fact this feature does have a considerable effect on vehicle behaviour close to the state of equilibrium.

The following graphs represent the principal characteristics of the suspension, at three stages of the optimisation process. The first graph is the most important one because it concerns the features of the *OCP* system. Lastly, a vehicle elastokinematic study is designed to examine a large number of parameter variations in detail. Because of the wide range of results, we have only presented the results for camber and half-track variations for the rear right-hand side wheel under lateral loads and its half-track variation under bump-rebound motions (figure 7).

For standard suspension systems, as a result of the bushing compliance under lateral loads, we obtain a camber as shown in figure 7. The application of the *OCP* system allows the camber trend to be radically reversed under lateral loads (i.e. negative camber) (figure 9). The optimisation task performed on the elastokinematic 3D model also improves this tendency. The increase in the camber curve slope thus obtained is about 180%. We can see that we improve the negative camber trend while controlling the increase in the right half-track variation. The resulting increase in that variation is about 40%, this remaining under the imposed limits (figure 10).

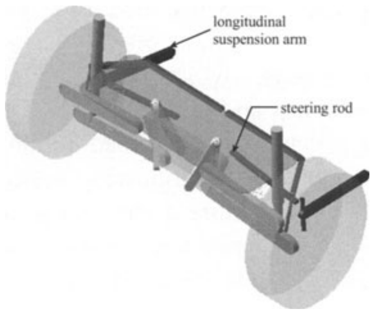


Figure -8. 3D model

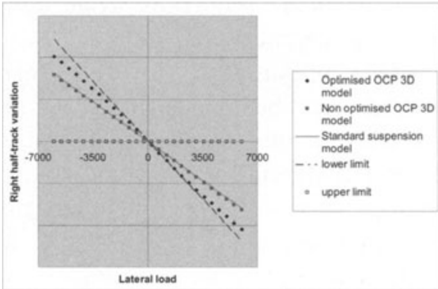


Figure -9. Right camber variation under lateral loads

This control is performed using a prudent range of trials obtained from the experiment designs. This selection takes into account all parameters for each vehicle load described by the intended specifications. It is then possible to maximise or minimise the parameters, while maintaining the other parameters in a correct range. In particular, optimisation leads to a decrease of 35% in the half-track variation under bump-rebound motion (figure 11), while the best results we could achieve were a variation of virtually zero.

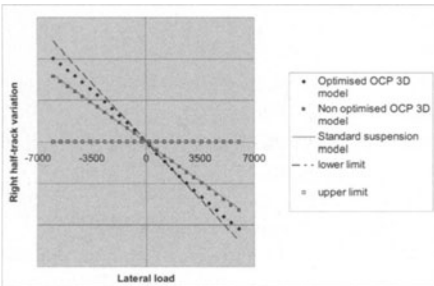


Figure -10. Right half-track variation under lateral loads

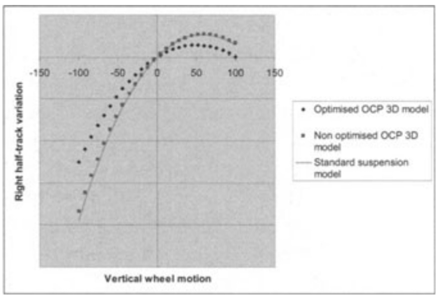


Figure -11. Right half-track variation under bump-rebound motion

6. CONCLUSIONS

The elastokinematic features of a vehicle have a considerable impact on its dynamic behaviour, such as road-holding. The principal method of studying these features is to represent the vehicle on a measuring bench using a numerical model. The objective is then to improve accurately the elastokinematic properties while optimising the model and obtaining the intended specifications.

Axle optimisation involves many parameters, which define both geometry and bushing stiffness. We then need a fast and efficient method to achieve optimum intended specifications, on which the vehicle road-holding properties depend. The method comprises several stages. Following design studies, a series of experiment designs allow us to explore a wide range of configurations and to determine a satisfactory optimum, while also respecting the imposed constraints. A further assignment should deal with the comparison between numerical and experimental results, in order to improve the numerical model's quality.

REFERENCES

1. T.C. CRAHAN : "Modeling steady state suspension kinematics and vehicle dynamics of road racing cars. Part1 : Theory and methodology", SAE Transactions, 1996, volume 105, pp 1-34
2. R.S. SHARP, M. BETTELLA : "Automotive Suspension and Chassis Interactions", SIA – Congrès de Dynamique du Véhicule, Ecole Centrale Lyon, 6-7 Juin 2001
3. J.M. D'SOUZA, J.M. STARKEY : "A force-based suspension modelling approach for vehicle steady-state handling", International Journal of Vehicle Design, 1998, volume 19, pp 167-186
4. I. ESAT : "Genetic algorithm-based optimization of a vehicle suspension system", International Journal of Vehicle Design, 1999, volume 21, pp 148-160
5. M.V. BLUNDELL : "The influence of rubber bush compliance on vehicle suspension movement", Materials and Design, 1998, volume 19, pp. 29-37
6. Z. LOZIA : "Model of steered tyred wheel, three-dimensional problem", Journal of Theoretical and Applied Mechanics, n°35, 1997, volume 4, pp. 849-870
7. SOCIÉTÉ MICHELIN : "Véhicule automobile équipé d'un système de contrôle de l'angle de carrossage des roues en virage", brevet européen n° de publication : EP1070609
8. J. FOULQUIER. "Une invention dans le domaine de la liaison au sol permettant l'optimisation en temps réel de l'aire de contact pneu / sol", SIA – Congrès de Dynamique du Véhicule, Ecole Centrale Lyon, 6-7 Juin 2001
9. MTS SYSTEMS CORPORATION : "About MTS Vehicle Dynamics", http://www.mts.com/vehicles/index.asp
10. T. HALCONRUY. "Les liaisons au sol", Editions ETAI, ISBN 2-7268-8250-1, 1995
11. D. BENOIST, Y. TOURBIER, S. GERMAIN-TOURBIER. "Plans d'expériences: constructions et analyse", Editions Tec et Doc, ISBN 2-85206-988-1, 1994

SELF-LOOSENING MODEL FOR BOLTED ASSEMBLIES UNDER TRANSVERSE STRESSES

H. Aziz, B. Combes and R. Lavabre
LGMT-INSA, Département de Génie Mécanique, Toulouse, France

Abstract : The present article investigates self-loosening of bolted assemblies under the effects of transverse loads. Firstly, we discuss the relaxation phenomenon that occurs at the end of tightening. We then perform modeling and calculus for self-loosening conditions for a short screw, neglecting the tilt of the screw in its tapping. In particular, we determine a maximum value for the friction coefficient such that this self-loosening occurs. Lastly, we indicate the procedure to confirm these results and to extend the study to the case of long screws.

Key words : bolted assembly, screw, unscrewing, screw-loosening, self-loosening.

1. INTRODUCTION

Conventional wisdom as contained in mechanical construction text books has it that "a well chosen and well tightened bolt does not loosen spontaneously". However, a host of devices intended precisely to prevent untimely unscrewing are manufactured and successfully marketed, none of which prevents operating incidents caused by the unexpected loosening of a bolted assembly being a relatively frequent occurrence. Although such events do not always lead to the total release of the parts, loosening leads to increased mobility of the parts and can cause fatigue fractures.

Factors inductive to loosening are well-known: apart from sizing or assembly errors, we can mention impacts, vibrations, temperature variations, creep of materials, etc. A wide range of techniques have for long been used by manufacturers to prevent unscrewing. Among those most frequently used, we can mention locking the rotation of threaded parts (using pins,

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 149–158.

deformable washers, etc.), braking by reinforcing friction on a fixed part or in the thread (using lock washers, close-fit threads, etc.), binding the thread, etc. The multiplication of these techniques testifies to the complexity of the problem, and the difficulty of solving it in a sure and economical way. But sometimes the very device designed to avoid unscrewing leads to rupture of the assembly by exceeding the tightening limits, damage to the parts through surface caulking during assembly, or through damage incurred during disassembly.

Since Goodier and Sweeney in 1945 [1], a number of authors have investigated self-loosening, either to analyse the physical causes or to determine the technological conditions that may prevent it. In particular, Hess [2] and Junker [3] can be mentioned in this context. Their works distinguish different kinds of phenomenon, according to whether the assemblies are subjected to an axial moment, an axial force or a transverse force on the bolt axis.

Self-loosening under the effect of axial moment is a relatively frequent phenomenon that is fairly easy to understand [4]. A part, fastened by a screw, turns about the axis of this screw and drives it in its rotation by friction. Axial tightening of the screw, acting on the helicoidal form thread, causes a torque reaction that encourages movement in the unscrewing direction. This explains why, in the event of an alternating load due to irregularity of the torque or to a vibration in torsion, we shall observe a general unscrewing movement.

Self-loosening under the effect of an axial force is a much rarer phenomenon. Indeed, its very existence is open to debate. Some of the experiments described in the literature involve impacts and we may surmise that that these impacts are capable of inducing transverse loads that would explain the observed screw-loosening. Other experimenters obtain screw-loosening (and even screwing!) of bolts subjected to very weak tightening through axial vibrations. Through experiments and by calculation, we conducted tests [5] to determine the conditions that would be conducive to spontaneous screw-loosening under the effect of a purely axial force of a simple but realistic bolted assembly. We only attained that point with tightening levels very much lower than those used in mechanical assemblies.

Finally, spontaneous screw-loosening under the effect of a transverse force is the type most frequently encountered. While we can readily reproduce this condition in experiments, modelling and calculating it are delicate matters because we need to take into account displacements, friction and deformations in the complex forms of the thread. Different authors have insisted on a number of aspects, especially flexion of the thread [6], tilting of the screw about an axis perpendicular to the exerted force [7] and swivelling of the screw around a parallel and laterally shifted axis [8]. In all events, under an alternating load, the screw comes to rest sometimes against one side and sometimes against the other side of the thread, and slides in the tapped hole with low rotation, always oriented in the screw-loosening direction.

We shall first present relaxation of the screw torsion that occurs at the end of screwing. We shall then study the behaviour of a screwed assembly under the effect of a transverse load: we shall see, in the case of a short screw, the conditions for sliding under the head, sliding in the thread, and the process that leads to self-loosening. Finally, we shall state the limits to this study and suggest how further studies may be conducted to take matters further.

2. RELAXATION AT THE END OF SCREW-TIGHTENING

The assembly considered here comprises a right-threaded screw, engaged in a tapped hole in a massive base, with a part being clamped in place between the screw-head and the tapped base. For numerical applications, the ISO metric thread is retained, with a half angle of thread of $2\alpha = 60°$ and the average thread lead angle $\beta \approx 3°$ for ordinary thread pitches and $\beta \approx 2°$ for fine pitches.

Let us refer to figure 1 and, as stated in [9], consider the screw subjected to an external torque, C, exerted on the head (using a wrench or some other screw-tightening tool) and to friction and reaction in relation to its supports: torque, C_s, due to friction under the head, torque, C_t, due to contact in the threading and tension, Q, of the stem. The transverse force F is null and the screw equilibrium results in the relation: $C + C_s + C_t = 0$.

The expression for the friction torque under the head is $C_s = \mu_s r_s Q$ during screwing and $C_s = - \mu_s r_s Q$ during unscrewing, where μ_s is the coefficient of friction and r_s the equivalent part's action radius under the head.

The torque in the thread is expressed by $C_t = \mu_t' r_t Q$ during screw-tightening and by $C_t = \mu_t'' r_t Q$ during unscrewing, where μ_t' and μ_t'' are the equivalent coefficients of friction and r_t is the equivalent action radius. An elementary calculation shows that :

$$r_s = \frac{2}{3}\frac{r^3_{s\ out} - r^3_{s\ in}}{r^2_{s\ out} - r^2_{s\ in}} \qquad r_f = \frac{2}{3}\frac{r^3_{t\ out} - r^3_{t\ in}}{r^2_{t\ out} - r^2_{t\ in}} \tag{1}$$

$$\mu_t' = \frac{\sin\beta + \mu_t \cos\beta}{\cos\alpha - \mu_t \sin\beta} \qquad \mu_t'' = \frac{-\sin\beta + \mu_t \cos\beta}{\cos\alpha + \mu_t \sin\beta} \tag{2}$$

$r_{s\ out}$ and $r_{s\ in}$ are the outer and inner radii of the annular bearing surface of the head, while the same applies for r_t. μ_t is the coefficient of friction on contact of the threads.

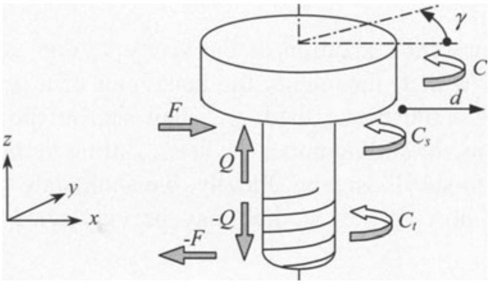


Figure -1. Forces and torques exerted on the screw

During screw-tightening, the stem of the screw is subjected to torsion that tends to make the threaded portion rotate in the screwing direction. When the external torque, C, is no longer applied to the head, torsion remaining in the stem still tends to make the threaded portion rotate in the screwing direction and the torque in the thread retains the same sign (i.e. C_t remains positive). Conversely, the same torsion tends to bring the head back in the unscrewing direction and the frictional torque under the head then changes its sign (i.e. C_s becomes negative).

The torsional moment that remains in the stem at the end of the tightening is limited by the least value (in absolute value terms) of torques C_s and C_t. For example, with a standard size connecting screw:

- $\mu_t' > \mu_s$ due to the wedging effect resulting from the half angle of the thread, α, and the lead angle, β;
- $r_t < r_s$ as the mean radius of the thread is less than the mean radius of the seat surface of the head.

The wedging effect of the thread is generally predominant: $\mu_t' r_t > \mu_s r_s$ and, after screwing, we can see very slight sliding of the screw head in the screw-loosening direction, bringing the torsional moment in the stem back to the value of the sliding torque of the screw head :

$$C_s = -\mu_s r_s Q = -C_t \tag{3}$$

If, during later use of the assembly, some transverse loads lead to the thread or the bearing surface of the head sliding, then rotation of these surfaces will be eased and the screw will end up losing its torsion. If, finally, this sliding occurs both under the head and on the thread, spontaneous screw-loosening can be obtained. We shall now proceed to analyse this phenomenon.

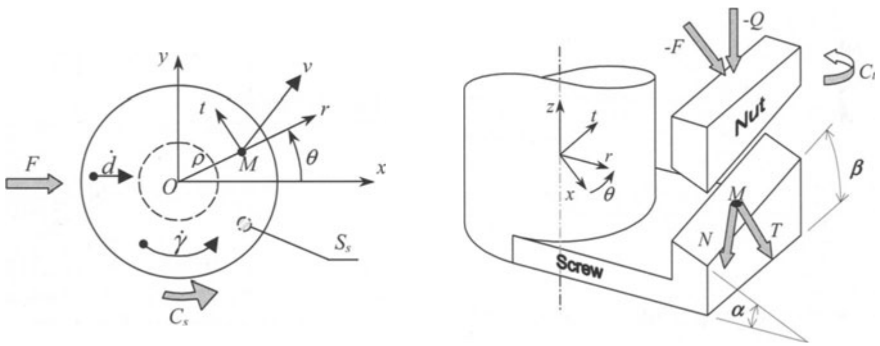
3. SELF-LOOSENING OF A SHORT SCREW SUBMITTED TO TRANSVERSE SLIDING

Here, we shall study short screws (Figure - 1) for which a transverse load F generated by a transverse movement d of the tightened part along the $\vec{x}$ axis generates only slight rocking torque about the $\vec{y}$ axis. We shall neglect the latter torque and consider the stem of the screw to work without noticeable bending in traction, torsion and shearing. We shall also assume that the tension Q of the stem is uniformly distributed along the contact surfaces, as much at the support of the screw head as in the thread.

3.1 Sliding under the head

Given γ, the rotation of the screw around its $\vec{z}$ axis, according to figure 2 (a), at one point of the contact surface of co-ordinates (ρ, θ), the sliding velocity of the tightened part in comparison with the screw and its module will be given by :

$$\vec{v} = \dot{d}\vec{x} - \rho\dot{\gamma}\vec{t} \qquad\qquad v = |\vec{v}| \tag{4}$$



(*a*) – under the screw head (*b*) – on the thread

Figure -2. Velocity and load

Considering the distribution of the sliding velocity on the bearing surface, S_s , the contact loads result in a force F along $\vec{x}$ and a torque C_s along $\vec{z}$:

$$F = \mu_s \frac{Q}{S_s} \int_{S_s} \vec{x} . \frac{\vec{v}}{v} ds \qquad C_s = \mu_s \frac{Q}{S_s} \int_{S_s} \rho\, \vec{t} . \frac{\vec{v}}{v} ds \tag{5}$$

During screw-loosening, the torsional moment in the stem of the screw is low. As a result, the tangential velocity $r_s\dot{\gamma}$ due to the rotation of the head is small compared to the translation velocity $\dot{d}$:

$$|C_s| << |r_s F| \quad \Rightarrow \quad \varepsilon_s = \left|\frac{r_s\dot{\gamma}}{\dot{d}}\right| << 1 \tag{6}$$

The expressions for F and C_s are then simplified in :

$$F \approx \pm\mu_s Q \qquad\qquad C_s \approx -\varepsilon_s\mu_s r_s Q \tag{7}$$

3.2 Slide in the thread

At one point of the contact surface, of co-ordinates (ρ, θ), the sliding velocity of the nut in comparison with the screw is (Figure – 2 (b)) :

$$\vec{v} = -\dot{x}\vec{x} - \rho\dot{\gamma}\vec{y} + (\dot{x}\cos\theta\tan\alpha + \dot{x}\sin\theta\tan\beta - \rho\dot{\gamma}\tan\beta)\vec{z} \tag{8}$$

Considering the distribution of the sliding velocity on the contact surface of the thread S_t, the contact loads result in a force $-F$ along $\vec{x}$ and a torque C_t along $\vec{z}$:

$$-F = \frac{Q}{S_t}\int_{S_t}\frac{\mu_t(\rho\dot{\gamma}\sin\theta - \dot{x}) - v(\sin\alpha\cos\theta + \sin\beta\sin\theta)}{v\cos\alpha - \mu_t\vec{v}.\vec{z}}ds \tag{9}$$

$$C_t = r_t\frac{Q}{S_t}\int_{S_t}\frac{v\sin\beta - \mu_t(\rho\dot{\gamma} - \dot{x}\sin\theta)}{v\cos\alpha - \mu_t\vec{v}.\vec{z}}ds \tag{10}$$

Given ε_t , the ratio of tangential and transverse velocities in the thread : $\varepsilon_t = |r_t\dot{\gamma}/\dot{x}|$ and with $A(\theta) = \sqrt{1 + \varepsilon_t{}^2 - 2\varepsilon_t\sin\theta + \cos^2\theta\tan^2\alpha}$, we can write :

$$\frac{F}{Q} = \frac{1}{2\pi}\int_0^{2\pi}\frac{\mu_t(1 - \varepsilon_t\sin\theta) + (\sin\alpha\cos\theta + \sin\beta\sin\theta).A(\theta)}{\cos\alpha.A(\theta) - \mu_t(\cos\theta\tan\alpha + \sin\theta\tan\beta - \varepsilon_t\tan\beta)}d\theta \tag{11}$$

$$\frac{C}{r_t Q} = \frac{1}{2\pi}\int_0^{2\pi}\frac{\sin\beta.A(\theta) - \mu_f(\varepsilon_t - \sin\theta)}{\cos\alpha.A(\theta) - \mu_t(\cos\theta\tan\alpha + \sin\theta\tan\beta - \varepsilon_t\tan\beta)}d\theta \tag{12}$$

These expressions were evaluated numerically for the case of a coarse ISO metric thread pitch for various values of the friction coefficient μ_f and ratio ε_f.

3.3 Screw-loosening conditions

A transverse movement of the screw is only possible when the force due to the sliding of the tightened part under the screw head is higher than that needed for sliding of the threads in the tapped hole. According to equations (7) and (9), this condition can be expressed as follows : $F/Q > \mu_s$

Figure 3 shows that, if the coefficients of friction are identical under the screw head and in the threaded region, this condition obtains for $\varepsilon_t > 0{,}75$, whatever the value of μ..

According to equation (7), the frictional torque under the screw head tends towards 0 when ε_s is small, i.e. when the slide velocity of the tightened part is high compared to the unscrewing velocity. If this condition is observed, and if a transverse movement of the screw occurs, unscrewing will occur only if torque C_t due to the contact in the thread becomes positive. The numerical evaluation of equation (10) leads to figure 4.

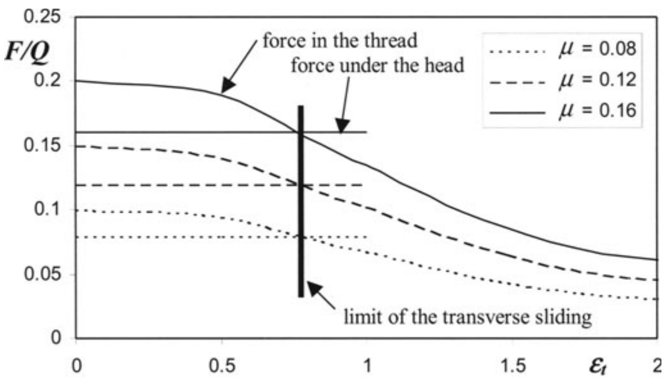


Figure -3. Transverse forces due to sliding

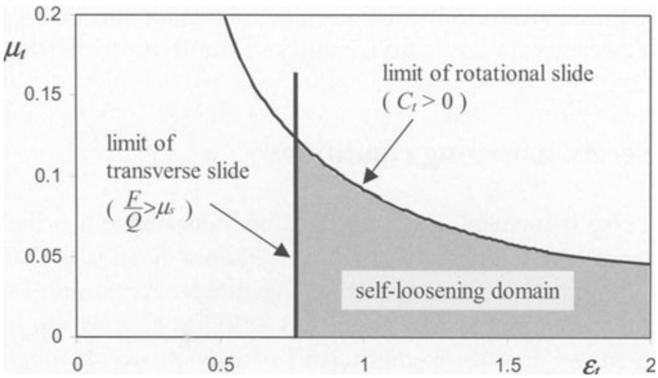


Figure -4. Self loosening domain

3.4 Self-loosening scenario

Let us consider an assembly comprising a short screw tightening a part subjected to an alternative transverse movement.

Immediately after screwing, the relaxation described above takes place. This relaxation brings the torque in the stem back to the value of the sliding torque of the screw head on the part.

The first transverse movements of the part leads to the screw head being able to rotate again in the unscrewing direction until the torque in the stem reaches a value close to 0. In this phase, the threaded portion of the screw that is acted on in the screwing direction, does not rotate because the torque due to the thread lead angle is opposite to the torsional moment. When the torsion of the stem is cancelled out, the combined action of the torque due to thread lead angle and the transverse force due to the slide under the head can cause the screw to move in its tapped hole. Figure 4 shows that this occurs only for a coefficient of friction at the thread contact of less than 0.12. For current values of μ_t (between 0.1 and 0.12), the ratio ε_t of tangential and transverse velocities at the thread is about 0.75 to 1. Conversely, the ratio ε_s of tangential and transversal velocities under the screw head is close to 0: rotation and radial displacement of the screw in its tapped hole are very slow as compared with the motion of the tightened part.

What results paradoxically from this is that an increase in the coefficient of friction μ_s under the head makes unscrewing easier by allowing translatory drive of the screw by the tightened part while the resistive torque C_s under the head remains almost null. This can explain the sometimes negative role played by certain anti-unscrewing washers.

4. LIMITS TO RESOLUTION AND PROSPECTS

We should bear in mind, however, that the above results are based on a number of assumptions. The most serious problems here appear to us to relate to variability in the coefficients of friction and to the tilting of the screw in its tapped hole.

We are all well aware that Coulomb's laws of friction only provide a rough approximation and the coefficients of friction which we have assumed to be constant depend on the sliding velocity. This can make starting up very slow motion of the screw in its tapped hole more difficult.

Further, we have neglected tilting torque around the $\vec{y}$ axis and, as a result, we have assumed uniform distribution of the tension of the stem Q over the contact surfaces, both on the support of the screw head and on the thread. This assumption appears to be justified only in the case of very short screws. For an assembly with greater thickness, transverse resistance to sliding of the head, acting along the portion between the tapped hole and the screw head, generates a tilting moment. The deformation of the thread and of the head, combined with bending of the stem, produces a rotation of the screw in its tapped hole around the $\vec{y}$ axis perpendicular to both the movement and the axis of the tightened part. These deformations can absorb slight movement of the part and prevent or delay the slide needed for unscrewing.

Moreover, the slide of the screw in its tapped hole is much more readily obtained under the effect of tilting torque than through transverse force alone : it begins on the first thread and gradually propagates through the depth of the tapped hole. We believe this effect to be preponderant where the screw is no longer very short, explaining most cases of self-loosening observed under transverse stress. Detailed study of this phenomenon would require us to use more refined models to consider deformations of the thread, the core of the screw and the tapped support itself, throughout the length of the thread in contact. Pursuing the model further would appear to be difficult without resorting to numerical methods.

Finally, we carried out a number of series of tests to check our forecasts, imposing alternative movement on a part tightened by a screw (Figure -5). We were able to measure the change in rotation and tension of the screw during self-loosening, together with movement of the screw in its tapped hole.

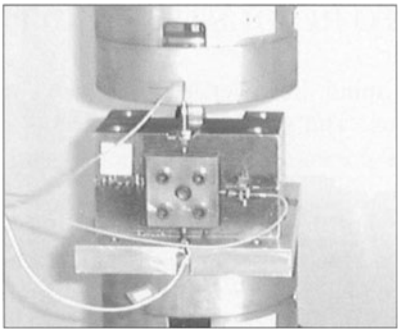


Figure -5. Testing apparatus – transverse stresses

5. CONCLUSION

This study introduced modelling of the self-loosening process for a bolted assembly, in the case of a short screw, neglecting tilting of the screw in its tapped hole. The self-loosening conditions were established and, in particular, we determined a maximum value for the coefficient of friction. Finally, we indicated how these results could be confirmed through further investigation and how the study could be extended to study the case of long screws.

REFERENCES

1. J.N. GOODIER, R.J. SWEENEY, "*Loosening by Vibration of Threaded Fasteners*", Mechanical Engineering, Vol. 67, 1945, pp. 798-802.
2. D.P. HESS, S.V. SUDHIRKASHIAP, "*Dynamic Loosening and Tightening of a Single-Bolt Assembly*", Journal of Vibration and Acoustics, Vol. 119, 1997, pp. 311-316.
3. J.G. JUNKER, "*New Criteria for Self-Loosening of Fasteners Under Vibration*", SAE Transactions, Vol. 78, 1969, pp. 314-335.
4. T. SAKAI, "Investigations of Bolt Loosening Mechanisms, 2nd Report, On the Center Bolts of Twisted Joints", Bulletin of the JSME, Vol. 21, No. 159, 1978, pp. 1391 – 1394.
5. H. AZIZ, B. COMBES, R. LAVABRE, "*Etude du dévissage spontané des assemblages boulonnés sous une force axiale*", Actes du 7ème Colloque AIP-PRIMECA sur la Conception Mécanique Intégrée, La Plagne, France, 2-4 avril 2001, pp. 74 - 80.
6. R.I. ZADOKS, X. YU, *"A Preliminary Study of Self-Loosening in Bolted Connections"*, Nonlinear Vibration ASME, DE-Vol. 54, 1993, pp. 79 – 89.
7. A. YAMAMOTO, S. KASEI, *"A solution for self loosening mechanism of threaded fasteners under transverse vibration"*, International Journal of the Japan Society for Precision Engineering, Vol. 18, 1984, pp. 261 - 266.
8. T. SAKAI, "Investigations of Bolt Loosening Mechanisms, 1st Report, On the Bolts of Transversely Loaded Joints", Bulletin of JSME, Vol. 21, No. 159, 1978, pp. 1385 – 1390.
9. J. GUILLOT, "*Assemblages par éléments filetés – Calcul*", Techniques de l'Ingénieur, Editions des Techniques de l'Ingénieur, Vol. B 5, 1987, pp. 560.1-25, 561.1-17, 562.1-14.

CRASHWORTHINESS OPTIMIZATION USING A SURROGATE APPROACH BY STOCHASTIC RESPONSE SURFACE

B. Soulier*, L. Richard*, B. Hazet* and V. Braibant***
**LURPA, ENS de Cachan, France*
***PSA Peugeot Citroën, Velizy, France*
****MECALOG SA, Antony, France*

Abstract: In the automotive passive safety field, numerical simulations gradually replace experimental crash-tests, and allow through parametric studies an improved definition of the architecture and the sizing of vehicles. In this context, this paper is focused on a methodology for crashworthiness optimization. After a review of difficulties inherent to the numerical modeling, we propose a global optimization strategy based on a surrogate approach : the resolution of the real optimization problem is replaced by a sequence of resolutions of approximate problems. An interpolation model is adopted in order to smoothen the objective function and constraints and to enable the analytical calculations of their gradients. The response surface model is build by a stochastic process. Unlike traditional techniques of construction of polynomial response surfaces by least squares regression, the approach developed, based on SPH (smooth particle hydrodynamics) methods, makes it possible to reproduce strong non-linearities of the objective functions and limiting constraints. Moreover, the flexibility of these models allows the updating of the approximation during the optimization process, which makes it possible to improve locally the quality of the approximations. We compare the quality of the approximations for various types of optimal design of experiments.

Key words: Optimization, Response Surface Methodology, Design of Experiment, Crashworthiness

1. INTRODUCTION

Car manufacturers are today brought to reduce their development times to answer expectations of their customers and to remain competitive. In this

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 159–168.

context, the designers are increasingly using numerical simulation techniques, which gradually replace physical testing on real prototypes.

At the same time, aided design methods make their appearance in engineering and design departments. They allow the parameterization of numerical models of vehicles, and guide the engineers in their choice of new technological solutions. Meanwhile, the field of passive security is an area of on-going progress currently favoured by car manufacturers. The implementation of optimization techniques in fast dynamics thus facilitates the design of vehicles with respect to this service. Numerical optimization of crashworthiness requires the meshing of the vehicle (fig1).

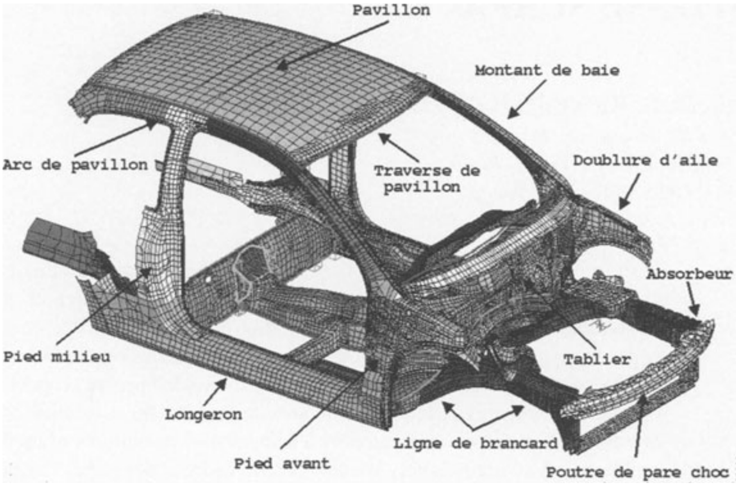


Figure -1. Mesh of a vehicle

In the automotive passive safety field, an objective function can be represented by the PMG acceleration.

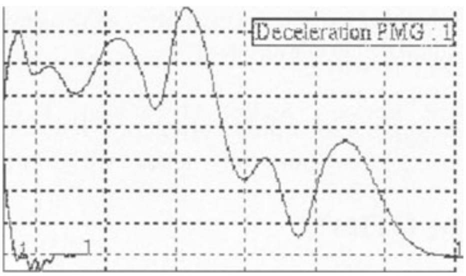


Figure -2. Acceleration of PMG

1.1 Optimization methods and concepts of approximations

Generally, the objective and constraint functions of a structural optimization problem are implicit with respect to the design parameters: they can be evaluated using finite elements analysis. This step is the most expensive operation of the minimization algorithm. Therefore, the most effective optimization methods [1] use the concepts of approximation. The initial design problem is replaced by a sequence of approximate and explicit problems. These approximation techniques are local when they are based on the linearization of the functions; they are considered to be global when identifying polynomial models or neural networks.

1.2 Formulation of crashworthiness optimization problems

The optimization problems encountered during the dimensioning of vehicles in crash are classically described by the following system:

$$\begin{cases} \min\limits_{x} F(\mathbf{x}, q, \dot{q}, \ddot{q}) \\ g(x, q, \dot{q}, \ddot{q}) \leq 0 \\ M\ddot{q} + g^{\text{int}}(q, \dot{q}) = g^{ext}(q, \dot{q}) \end{cases}$$

where F is the objective function of the optimization problem, g are the constraint functions of optimization, and $\mathbf{x}$ is the vector of the design variables. Displacements q, speeds $\dot{q}$ and accelerations $\ddot{q}$ of the structure are controlled by the discrete dynamics equation, integrated by an explicit time scheme.

Numerical simulation of automotive crash-tests is delicate because of the nature of the involved physical phenomena such as buckling of the structures. The calculated time response is therefore strongly non-linear, numerically very unstable, and their gradients cannot be computed analytically. We propose in this article the formulation of an approximation model, which aims to regularize optimization problems in fast dynamics. After a description of the interpolation model, we examine some of its properties, before illustrating its construction in an analytical example.

2. APPROXIMATION MODEL FORMULATION

2.1 Basics

The approximation model presented in this paragraph, based on bayesian statistics, is inspired by SPH techniques (Smooth Particle Hydrodynamics) [2][3], initiated in the Eighties in the field of astrophysics. These methods consist in approaching a function y by a balanced sum $\tilde{y}$ of exponential [4] or spline shape functions ϕ_i.

The model suggested generates continuous and differentiable response surface on the field of study, following the equation:

$$\tilde{y}(x) = \frac{\sum_{i=1}^{n} \phi_i(x) a_i}{\sum_{i=1}^{n} \phi_i(x)}$$

The interpolation coefficients a_i are obtained by a collocation method, and are solutions of the following linear system:

$$\begin{bmatrix} \dfrac{\phi_1(x_1)}{\sum \phi_i(x_1)} & \cdots & \dfrac{\phi_n(x_1)}{\sum \phi_i(x_1)} \\ \vdots & \ddots & \vdots \\ \dfrac{\phi_1(x_n)}{\sum \phi_i(x_n)} & \cdots & \dfrac{\phi_n(x_n)}{\sum \phi_i(x_n)} \end{bmatrix} \begin{Bmatrix} a_1 \\ \vdots \\ a_n \end{Bmatrix} = \begin{Bmatrix} y(x_1) \\ \vdots \\ y(x_n) \end{Bmatrix}$$

where the vector $(.x_1, \ldots, x_n)$ represents the list of simulation points where the function y is evaluated exactly, and $\phi_i(x)$ is the value of the correlation function at the x point.

2.2 Correlation function

The shape functions [5] are of a gaussian type (fig 3):

$$\phi_i(x) = e^{-k\|x - x_i\|^2}$$

where the distance $\|x - x_i\|$ is the length between the points x and x_i, and k a strictly positive normalization parameter.

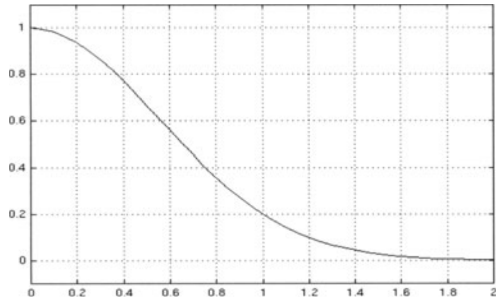


Figure -3. Gaussian-type shape function

The k parameter controls the shape of the surrogate surface by modifying the size of the support of the shape functions (figure 4). Low values of k produce very smooth surfaces, while high values of k produce staircase surfaces.

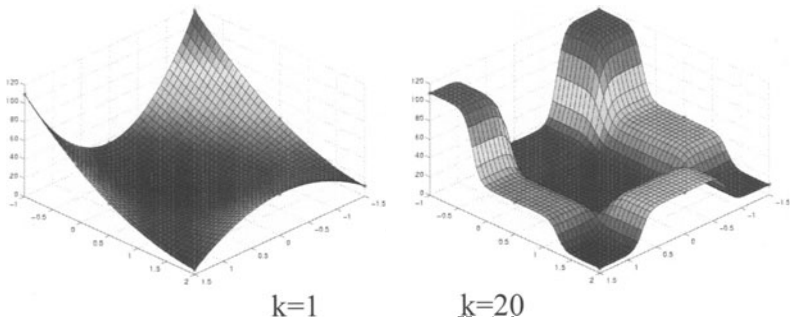


Figure -4. Influence of the k parameter on the shape of the interpolations

2.3 Modeling properties

The suggested modeling presents some interesting properties in the context of the resolution of optimization problems in fast dynamics:

- Ability to Reproduce non-linearities: contrary to linearizations and models of polynomial regression, the suggested interpolations are defined by linear combinations of non-linear functions, and thus restore the possibly strongly non-linear behavior of the objective function and limiting constraints,
- Flexibility: The approximations contain one shape function for each point in the sample of numerical experiments (x_1 , …, x_n). In an optimization process, this property makes it possible to enrich the approximations with information provided at each iteration by the mathematical optimization algorithm. This principle ensures convergence towards a real minimum

of the initial optimization problem, because approximate surfaces locally tend to the real modeled surfaces,

- Differentiation: The sensitivity analysis, essential to the calculation of the directions of descent used by optimization programs, is performed by analytical differentiation of the interpolation models.

3. DEFINITION OF THE SAMPLE OF SIMULATION POINTS

The determination of the sample of simulation points is obtained by a design of experiment strategy. Designs of Experiments are classically used to identify analytical models with a minimum of physical or numerical testing [6].

3.1 Factorial and Central-Composite design of experiment approach

The full-factorial design is based on the evaluation of all the combinations of factors to explore the field of variation. Three levels at least are necessary to estimate non-linearities in responses. An economic alternative to this type of design is the central-composite type. It is obtained by the union of a two-level full factorial design, a center point, and two additional points for each factor.

3.2 D-optimal design

Considering the costs of calculations raised in crash automotive, a D-Optimal Design of Experiment [7] has been developed to solve optimization problems with more than 5 design parameters.

4. COUPLING BETWEEN THE APPROXIMATION MODEL AND THE OPTIMIZATION PROGRAM

As previously introduced, the responses obtained in fast dynamics are unstable and contain numerical noise. Consequently, the related optimization problems are badly conditioned and present many local minima. We present an optimization algorithm adapted to the problems of the automotive crash. It is based on the coupling of a method of mathematical programming with the construction of interpolation surfaces, and consists in two main phases.

- an exploratory phase, which identifies a number of potentially optimal zones in the initial design space, by means of an examination technique.
- the actual optimization phase, in which several optimization sub-problems are solved by a mathematical programming method.

During the optimization by the mathematical programming algorithm, the response surfaces are updated. The solution of each optimization problem approached is included in the approximations of the following problem. This process is repeated until convergence of the algorithm is obtained, and makes it possible to tend locally to a solution of the real problem.

In addition, the exploratory phase leads to the definition of various sub-problems of optimization, which correspond to various local minima of the real problem.

5. ANALYTICAL EXAMPLE: THE ROSENBROCK FUNCTION

The Rosenbrock function is a classic unconstrained optimization test-case. It is very delicate to solve because the surface to be minimized presents a very lengthened, arc of circle-shaped valley. Its analytical expression is as follows:

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

The minimum of this function is the point (1,1), for which f=0. The exploratory phase of the algorithm results in the generation of an initial response surface built on a 4-level full factorial DOE.

Table 1. 4-level full factorial design

#	x_1	x_2	f	#	x_1	x_2	f
1	-1.75	-1.75	2323.6	9	-1.75	1.25	336.1
2	-0.25	-1.75	330.1	10	-0.25	1.25	142.6
3	1.25	-1.75	1097.3	11	1.25	1.25	9.8
4	2.75	-1.75	8675.3	12	2.75	1.25	3987.8
5	-1.75	-0.25	1104.8	13	-1.75	2.75	17.3
6	-0.25	-0.25	11.3	14	-0.25	2.75	723.8
7	1.25	-0.25	328.6	15	1.25	2.75	141.1
8	2.75	-0.25	6106.6	16	2.75	2.75	2319.1

This initial approximation leads to the identification of two valleys R1 and R2 defined by:

$$R_1 : \begin{cases} -0.56 < x_1 < 1.64 \\ -0.92 < x_2 < 1.28 \end{cases} \quad R_2 : \begin{cases} -2.00 < x_1 < -0.90 \\ 1.90 < x_2 < 3.00 \end{cases}$$

The optimization algorithm, based on approximations of f updated throughout convergence, is then used in each of these areas, and results in the following solutions:

$$P_1^* = (0.89;0.79) \text{ et } f^*(P_1) = 0.012$$
$$P_2^* = (-1.43;2.05) \text{ et } f^*(P_2) = 5.900$$

The resolution of the R1 problem is illustrated in figure 5. The layout of the initial interpolation in the R1 area is represented in figure 10a. Figure 10b shows the path of the optimization process on the final approximation, obtained after 5 successive updates.

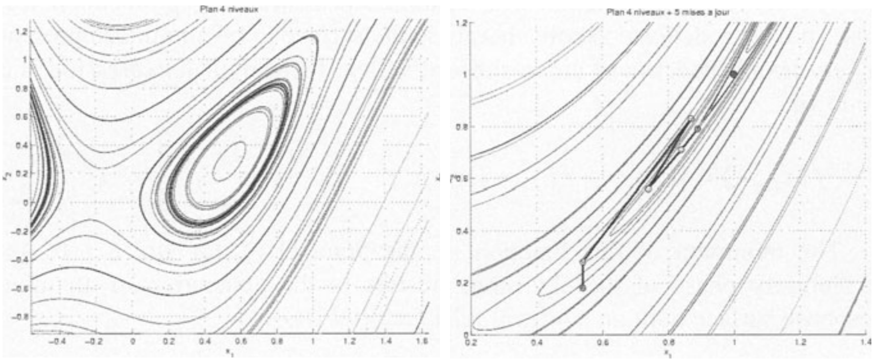


Figure -5. Resolution in the R1 area

6. INDUSTRIAL APPLICATION

We illustrate the application of the surrogate approach on the optimization of an automotive component. The objective is to minimize the deceleration of the PMG component. The design variables (figure 6) are the stiffness of 3 beams which make up the front part of the vehicle.

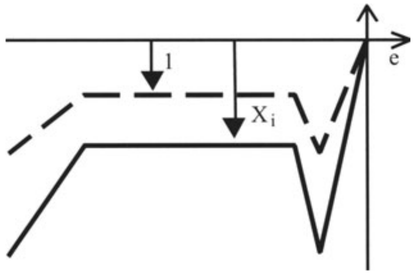


Figure -6. Design variables of the automotive crashworthiness optimization

The FE model contains 2000 elements and the runtime of a simulation on a Cray J916 runs into 1000s. We compare (figure 7) the deceleration response of the PMG component obtained by the approximation method to the crash-test response.

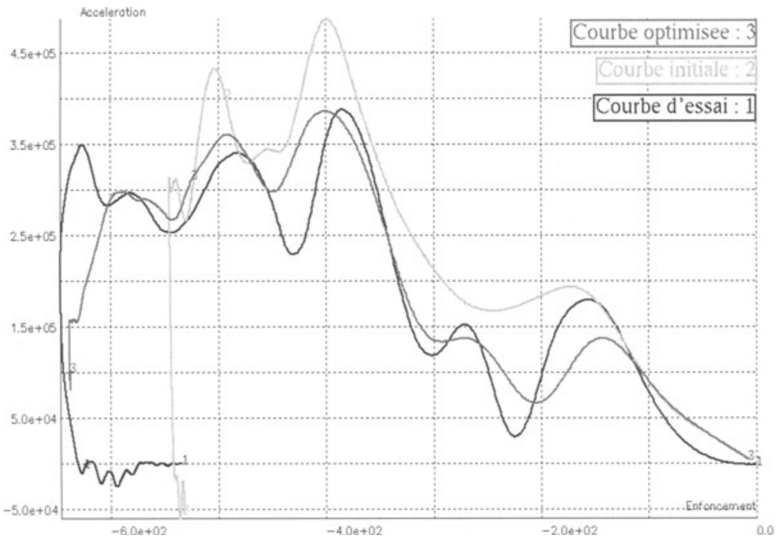


Figure -7. PMG deceleration curve

These results are obtained with a D-optimal Design of Experiment combining 49 numerical tests.

7. CONCLUSION

The proposed optimization method of non-linear problems, based on the construction of a sequence of interpolations, answers the problems of optimization in fast dynamics. Moreover, this method allows to identify several local minima and the iterative update of the approximations makes it possible to converge at a lower cost to the solutions of the real problems. This method is applied successfully to industrial problems of explicit dynamics in the optimization of pre-sizing models of motor vehicles in frontal collision.

REFERENCES

1. Fleury C., Braibant V., Optimization methods, Computer Aided Analysis and Optimization of Mechanical System Dynamics, E.J. Haug Editor, NATO ASI Series, Computer System Sciences, vol 9, pp 637-677, 1984
2. Monaghan J.J, Gingold R.A., Shock simulation by the Particle Method SPH, Journal of the Computational Physics, vol 52, pp 374-389, 1983
3. Nayrolles B., Touzot G., Villon P., Generalizing the Finite Element Method: Diffuse approximation and diffuse elements, Computational Mechanics, vol 10, pp 307-318, 1992
4. Rasmussen J., Non linear programing by cumulative approximation refinement, Structural Optimization, vol 15, pp 1-7, 1998
5. Currin C., Mitchell T., Morris M., Ylvisaker D., Bayesian prediction of deterministic function with applications to the design and analysis of computer experiments , J. of the American Statistical Association, vol 86, n° 416, dec 1991
6. Soulier B., Sur la modélisation expérimentale en mécanique: précision, optimization et applications industrielles, thèse de l'ENS Cachan, nov 1994
7. Mitchell T.J., An algorithm for the construction of D-optimal experimental designs, Technometrics, vol 16, n° 2, pp 203-210, 1974

MULTIDISCIPLINARY OPTIMIZATION OF COMPOSITE LAMINATES WITH RESIN TRANSFER MOULDING

C. H. Park*, W. S. Han*, A. Vautrin* and W. I. Lee**

** Mechanical and Materials Engineering Department, Ecole Nationale Supérieure des Mines de ST-Etienne, France*

*** School of Mechanical and Aerospace Engineering, Seoul National University, Korea*

Abstract: In this study, a multidiscplinary optimisation methodology of composite structures with the Resin Transfer Moulding (RTM) process is suggested in order to satisfy both the structural and the process requirements. Among the composite manufacturing techniques, the RTM process is distinguished for its own many advantages such as low manufacturing costs, complex shapes, high productivity, and good mechanical performance. In designing composite structures with the RTM process, tailoring the fibre preform architecture plays an important role as well as deciding the injection strategy. With some appropriate assumptions, a simplified optimisation method is suggested to find a near optimal design configuration. At first, the number of fibre mats, the stacking sequence of layer angles and the injection gate locations are found in order to satisfy the structural criteria, such as the stiffness requirement, and the process criteria, such as the mould filling time. Then, the thickness of the composite laminate and the fibre volume fraction are determined so as to reduce the weight keeping all the design criteria satisfied. This problem is a multi-objective optimisation problem and hence an objective function is appropriately formulated. As an optimisation technique, a genetic algorithm is used. The reliability of the present design methodology is assessed with some examples.

Key words: design methodology, genetic algorithm, multi-objective optimisation, resin transfer moulding (RTM)

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 169–178.

1. INTRODUCTION

During the last several decades, much attention has been paid to the optimisation of composite structures to improve their mechanical performance or reduce their weight. Also, some research has been carried out on the optimisation of manufacturing processes, especially in liquid composite mouldings such as Resin Transfer Moulding (RTM) and injection moulding, in order to reduce the manufacturing cost. However, these structural optimisations and process optimisations have been performed separately. Sometimes, the optimal design solution obtained only by structural optimisation yields a high manufacturing cost due to manufacturing complexities and long cycle times. As a consequence, it is necessary to consider both structural and process requirements for the design of composite structures.

In order to take into account the structural requirements in RTM, it is important to tailor fibre preform architecture such as fibre volume fraction and fibre orientation. From the processing point of view, it is very important to reduce the mould filling time so as to decrease the cycle time. It is also important to complete the mould filling before the resin cures. Thus, the injection strategy such as location of injection gate plays an important role.

In this study, the genetic algorithm is used as an optimisation algorithm. The genetic algorithm is proven to be a powerful and robust search method to deal with highly non-linear and large spaces having many possible local optima. It also performs well for integer programming problems such as stacking sequence optimisation problems.

As design variables, the stacking sequence of fibre mat and a location of injection gates are taken to satisfy all the structural and process requirements with a presumed fibre volume fraction, a number of fibre mats and a corresponding composite laminate thickness. And then, the thickness of the composite structure is minimized to reduce the weight of the part so that the design solution still satisfies the design criteria.

2. PROBLEM STATEMENT

2.1 Structural analysis

The composite laminate is assumed as the lay-up of the composite layers which are composed of unidrectional fibre mat (glass fibre) and thermosetting polymer resin (epoxy resin). Each layer is assumed to have a same thickness. The elastic moduli of the composites can be derived from the moduli of the fibre and of the matrix by the Halpin-Tsai equations [1].

$$E_1 = E_f V_f + E_m V_m \tag{1}$$

$$\nu_{12} = \nu_f V_f + \nu_m V_m \tag{2}$$

$$\frac{M}{M_m} = \frac{1 + \xi \eta V_f}{1 - \eta V_f} \tag{3}$$

$$\eta = \frac{(M_f / M_m) - 1}{(M_f / M_m) + \xi} \tag{4}$$

M = modulus E_2, G_{12}, or ν_{23}
V_f = fibre volume fraction
ξ = 2 for calculation for E_2, ξ = 1 for calculation for G_{12}
The displacements under the loads are calculated with classical lamination theory and finite element method [2].

2.2 Process analysis

The resin flow phenomenon during RTM process can be regarded as the flow through porous media, which has been described by Darcy's law. Darcy's law relates the fluid flow to the pressure gradient using the fluid viscosity and the permeability of the porous medium [3].

$$\bar{\mathbf{u}} = -\frac{\mathbf{K}}{\mu} \nabla P \tag{5}$$

The components of the permeability tensor transformed along axis x', at an angle θ to the principal x axis can be obtained by the next relations [4].

$$\mathbf{K} = \begin{pmatrix} K_{xx} & 0 \\ 0 & K_{yy} \end{pmatrix} \tag{6}$$

$$\mathbf{K}^{\boldsymbol{\theta}} = \begin{pmatrix} K_{xx}\cos^2\theta + K_{yy}\sin^2\theta & \left(-K_{xx}+K_{yy}\right)\sin\theta\cos\theta \\ \left(-K_{xx}+K_{yy}\right)\sin\theta\cos\theta & K_{xx}\sin^2\theta + K_{yy}\cos^2\theta \end{pmatrix} \tag{7}$$

The gapwised averaged permeability components $\overline{K}_{ij}$ for a lay-up of n layers each of thickness h^l with permeability components in the local mould co-ordinate frame K_{ij}^l will be given by [5]

$$\overline{K}_{ij} = \frac{1}{H}\sum_{l=0}^{n} h^l K_{ij}^l \tag{8}$$

where H is the total thickness of the layup. Once the permeabilities are known, the pressure distribution is computed with equation (9).

$$\nabla \cdot \left(\frac{\mathbf{K}}{\mu} \nabla P \right) = 0 \tag{9}$$

With the obtained pressure distribution, the velocity distribution can be computed as in equation (5).

In order to solve the governing equation, the Control Volume Finite Element Method (CVFEM) combined with Volume of Fluid (VOF) concept was applied [6,7].

3. OPTIMISATION PROCEDURE

3.1 Classification of the problem

In this study, two classes of the problems are treated.

According to the number of the injection gates, one-gate injection and two-gate injection are treated.

In the present study, two numbers of preassigned orientation angles are used : 0° and 90°, namely the cross-ply laminate.

3.2 Optimisation procedure

The final goal is to reduce the weight of the laminate plate while the design solutions satisfy all the design requirements. In this problem, there exist multiple variables and hence it is not so easy to consider all the variables at the same time. Thus, it is necessary to simplify the problem with appropriate assumptions. In the beginning, the fibre volume fraction and the number of layers are assumed appropriately. Only the stacking sequence of the layers and injection gate locations are considered as design variables and these two design variables are optimized to satisfy the structural and process requirements. Then, the minimum number of the layers N is found which satisfies all the design requirements. Up to this step, the fibre volume fraction has been maintained as a constant. Thus, the thickness of the laminate plate is proportional to the number of layers. With a given number of fibre mats, the thickness is inversely proportional to the fibre volume fraction in this stage. As the thickness is increased, the stiffness of the structure is enhanced. Now, the number of the layers, stacking sequence of layers and injection gate locations are found out to satisfy the structural and the process requirements. Then, the weight minimisation is performed. Two starting points are considered. The one is $N-1$ layers with an according stacking sequence and injection gate locations. The other is N layers with according design configurations. The former with $N-1$ layers does not satisfy both structural and process requirements. However, the selected design configuration with $N-1$ layers is near an optimal one among the design configurations with $N-1$ layers. The latter with N layers, of course, satisfies all the design requirements. Yet, some margin exists to reduce the thickness while satisfying the design requirements. There are two ways to determine an optimal thickness and a corresponding fibre volume fraction. With a given number of layers, namely a given amount of fibre mat, the stiffness of the structure increases and the mould filling time decreases as the thickness increases. The first way is to increase the thickness with $N-1$ layers until all the design requirements are satisfied. The other way is to decrease the thickness with N layers as thin as possible while all the design requirements are satisfied. In each case, the fibre volume fraction changes with the thickness variation. As a consequence, two sets of design configuration are found. Between these two solutions, the one that more reduces the weight of the structure is finally selected.

3.3 Objective function

In the present study, two injection cases are considered, i.e. one gate resin injection and two-gate resin injection. In case of one gate injection, an optimal gate location to minimize the mould filling time can be decided regardless of the layer angle, since it is assumed that the optimal gate location depends only on the mould geometry with a uniform permeability. Then, an optimal stacking sequence of the layers with different fibre orientations is found to minimize the maximum displacement of the composite structure under the given loading condition and to minimize the mould filling time. After that, a presumed gate location is checked with an obtained stacking sequence of layers.

To deal with the multi-criteria problem, a penalty parameter is introduced. Objective functions are formulated as

- Mould filling time

$$f(x_i) = \frac{1}{t} \tag{10}$$

- Mould filling time & displacement (stiffness)

$$f(x_i) = \frac{1}{d}\left(1 + \Gamma \frac{t_{,c}}{t}\right) \tag{11}$$

t : the mould filling time ($t_{,c}$: critical mould filling time)
d : the maximum displacement
Γ: penalty parameter

In order to balance between the design variables, the penalty parameter is introduced in the objective function. The value of the penalty parameter would be chosen by a compromise by the designer or the manufacturer.

In case of two-gate injection, the procedure is somewhat different. The optimal gate location is affected by the stacking sequence of the layers as well as by the mould geometry. The dependency of optimal gate locations upon layer angles are illustrated in Figure-1.

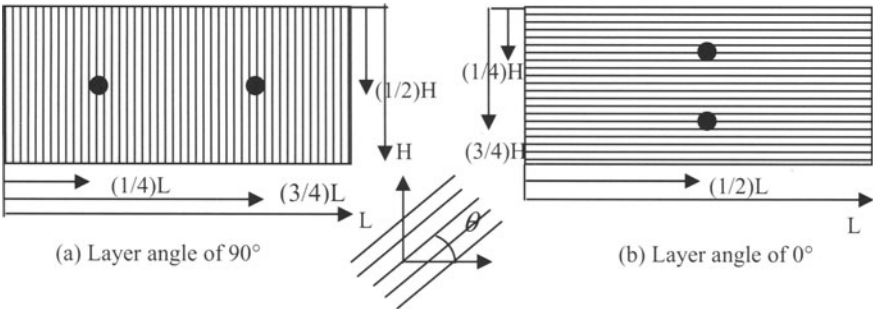


Figure -1. Optimal gate locations with different layer angles (L:H=2:1)

For this reason, the optimal gate location and the stacking sequence should be investigated at the same time so as to satisfy the stiffness requirement and to minimize the mould filling time. The objective function is formulated as equation (11).

4. GENETIC ALGORITHM

The genetic algorithm is inspired by the Darwinian model of biological evolution [8]. Since it is suitable for integer programming, it has been frequently adopted for stacking sequence optimisation problems.

In the present study, two design variables are considered : the stacking sequence of the layers with different fibre angle and the location of the resin injection gate. The stacking sequence of the preassigned angle is coded as a binary string. For two-angle set, 0° corresponds to 0 and 90° to 1. Thus a stacking sequence of 6 layers [0 90 0 0 90 0] is coded as [0 1 0 0 1 0]. Concerning the gate location, the nodes in the finite element calculation are taken as candidates. Their node number is coded as a binary string. For example, the node number 13 is decoded as [0 1 1 0 1] for 5-digit string.

If the objective function has several variables, as in the present case, the string can be represented by a concatenation of the different or same coding of the design variables. For example, a stacking sequence of layer = [0 0 0 90 90 0], and an injection gate node = 30, the binary coded concatenated string will be A=[0 0 0 1 1 0 , 0 1 1 1 1 0].

5. SAMPLE PROBLEM AND RESULTS

An example was a 40cm×20cm rectangular composite laminate plate for which thickness should be decided. The loading and boundary conditions are illustrated in Figure-2.

Figure -2. Loading and boundary conditions

Population for the genetic search is 30. Probability of crossover and mutation are given as 0.9 and 0.033, respectively.

The injection pressure is maintained as a constant of 0.1 MPa and the permeability ratio K_{11}/K_{22} is 53.91.

For the one injection gate problem, an optimal injection gate is found to be located at the center of the mould regardless of layer angles. Then an optimal stacking sequence is investigated. The results with stiffness maximisation with 7 layers are shown in Table-1 and with 8 layers in Table-2.

Table-1. The result of stiffness optimisation with 7 layers (1 gate injection)

Γ	stacking sequence [°]	displacement [mm]	mold filling time [sec]
0	90 90 90 90 90 90 90	15.70	3549.38
1	90 90 0 0 0 90 90	16.47	354.72

Table-2. The result of stiffness optimisation with 8 layers (1 gate injection)

Γ	stacking sequence [°]	displacement [mm]	mold filling time [sec]
0	90 90 90 90 90 90 90 90	10.54	3549.38
1	90 90 0 0 0 0 90 90	11.09	354.72

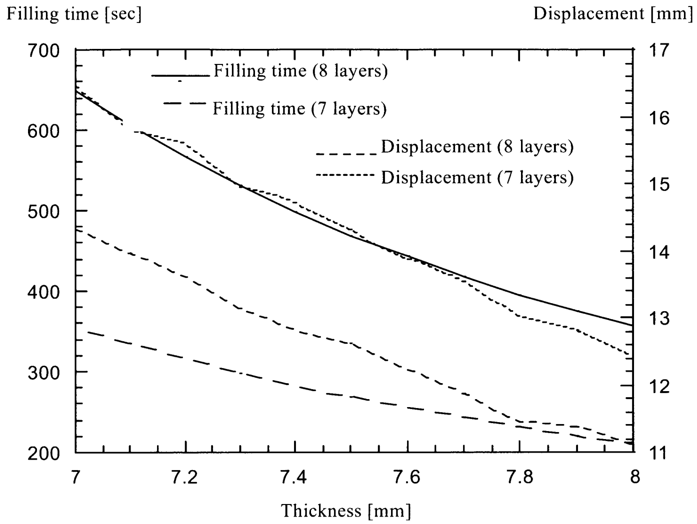


Figure -3. Mould filling time and displacement with thickness variation

In both cases, the stacking sequences only by stiffness optimisation, namely Γ = 0, yield a high mould filling time. However, the obtained stacking sequences with Γ = 1 show better performances in the process criteria while the stiffness performance hardly degenerates. Then, the thickness of the laminate plate is minimized in order to reduce the weight. In this study, the maximum allowable displacement is assumed to be 13mm and the maximum allowable mould filling time to be 500 sec.

Two stacking sequences are considered. The one is [90 90 0 0 0 90 90] with 7 layers and the other is [90 90 0 0 0 0 90 90] with 8 layers. In both cases, the fibre volume fraction is 0.45. The obtained stacking sequences satisfy the process criterion (mould filling time) in both cases. However, the former does not satisfy the structural requirement (stiffness). In the case of 7 layers, the thickness should be increased to enhance the stiffness of the composite laminate plate. On the contrary, the stacking sequence of 8 layers satisfies both the process and structural requirements. And, there is still a margin to reduce the thickness with the design requirements satisfied. The mould filling time and displacement according to the variation of the thickness are shown in Figure-3. With these results, two values of the thickness, i.e. 7.8mm with 7 layers and 7.4 mm with 8layers are selected while all the design requirements are satisfied. Between these two configurations, the better one resulting in a lighter structure is to be finally selected. The fibre volume fraction and total weight of the laminate plate according to the variation of the thickness are illustrated in Figure-4. As a result, the 7.8mm with 7 layers is selected.

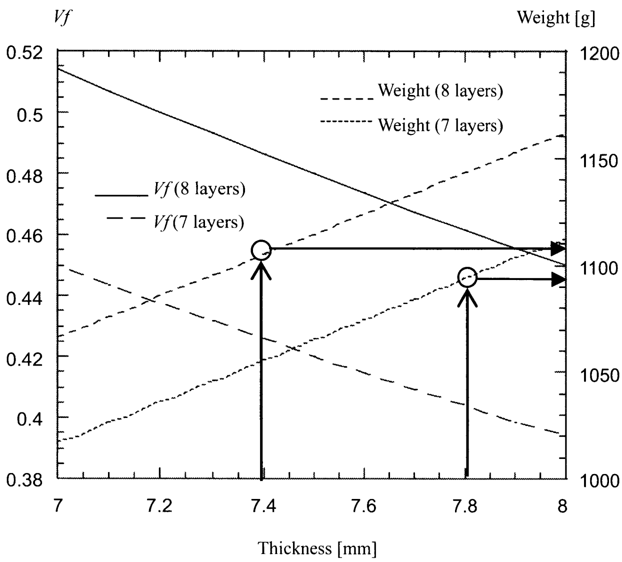


Figure -4. Fibre volume fraction and weight with thickness variation

Table-3. The result of stiffness optimisation with 7 layers (2 gate injection)

Γ	stacking sequence [°]	displacement [mm]	mold filling time [sec]	gate
0	90 90 90 90 90 90 90	15.70	1757.97	(L/4, H/2)
1	90 0 0 0 0 90 90	16.90	163.34	(L/4, H/2)

Table-4. The result of stiffness optimisation with 8 layers (2 gate injection)

Γ	stacking sequence [°]	displacement [mm]	mold filling time [sec]	gate
0	90 90 90 90 90 90 90	10.54	1757.97	(L/4, H/2)
1	90 90 0 0 0 90 90	11.09	159.21	(3L/8, H/2)

In a similar way, optimal design configurations are found in the case of 2 injection gates. In this case, the locations of injection gates and the stacking sequence should be found at the same time, since they are dependent on each other. The results are illustrated in Tables-3,4. Concerning gate locations, only the one gate location is shown since the other is located symmetrically.

6. CONCLUSIONS

In the present study, a design methodology is suggested so as to simultaneously optimise the structural and process parameters for a composite laminate plate manufactured with RTM. The stacking sequence of layers, locations of injection gates, number of the fibre mats and thickness of the laminate plate are optimized to reduce the weight of the structure while satisfying both the structural (stiffness criterion) and the process requirements (mould filling time criterion).

Because of its own characteristics, a genetic algorithm performs well for a stacking sequence optimisation. However, it requires more computational effort as the number of design variables increases. Thus, a simplified methodology with some appropriate assumptions is suggested to effectively find a near optimal design configuration.

In the case of a multi-objective optimisation problem, the designer or manufacturer should decide how much emphasis he will put on each objective considering the constraints and the situations he faces by modulating the penalty parameter in the objective function, as well as how the objective function should be formulated properly according to the given problem.

REFERENCES

1. J. C. Halpin, S. W. Tsai, “Effects of Environmental Factors on Composite Materials”, *AFML-TR* 67-423, June, 1969
2. G. Eyraud., W. S. Han, G. Verchery, “Finite Element Aided Design for Laminated and Sandwich Plates using Reanalysis Methods”, *Proceeding of the 9th International Conference on Composite Materials(ICCM9)*, **5**, 617-624, Madrid, 1993
3. F. A. L.,Dullien, “Porous Media Fluid Transport and Pore Structure”, *Academic Press*, 157-159, 1979
4. S. G. Advani, “Flow and Rheology in Polymer Composites Manufacturing”, *Elsevier Science*, 1994
5. A. E. Scheidegger, “Physics of Flow through Porous Media”, 3rd Ed.*University of Toronto Press, Toronto*, 1974
6. B. R. Baliga, S. V. Patankar, “Handbook of Numerical Heat Transfer”, *John Wiley & Sons, INC.*, 421-461, 1988
7. C. W. Hirt, B. D. Nichols, “Volume Of Fluid (VOF) Method for the Dynamics of Free Boundaries”, *J.Computational Physics*, **39**,201-225, 1981
8. D. E. Goldberg, “Genetic Algorithms in Search, Optimisation and Machine Learning”, *Addison Wesley*, 1989

Chapter 2

OPTIMIZATION OF MANUFACTURING SYSTEMS

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 179–180.

PROACTIVE SUPPORT FOR CONCEPTUAL DESIGN SYNTHESIS OF SHEET METAL COMPONENTS

F. Rehman and X-T. Yan
CAD Centre, Department of Design, Manufacture and Engineering Management, University of Strathclyde, Glasgow, UK

Abstract: This paper describes an approach of generating feasible design solutions in terms of form features according to functional requirements of sheet metal component design. The objective of this research is to develop a methodology of generating a solution space at the conceptual design stage and subsequently selecting an optimal solution for sheet metal component design. A decision made at the conceptual design stage causes consequences for all subsequent phases of the product life cycle. Normally a concept or a principle solution is selected on the basis of desired functional requirements, neglecting the consequence of selection on the performance of other life cycle phases like manufacturing and assembly etc. This approach causes problems at later product realization stages in terms of cost & time incurred due to redesign of product. This paper presents a novel approach of how consequences caused by design decisions on later life cycle phases specially manufacturing can be brought to the attention of designers at the conceptual design stage of sheet metal components

Key words: Functional Design, Conceptual Design, Sheet Metal, Manufacturing Consequences

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 181–190.

1. INTRODUCTION

Conceptual design is an early phase of the design process, which involves the generation of solution concepts to satisfy the functional requirements of a design problem. There can be more than one solution to a problem; this means that there is scope for producing improved designs if one could explore a full solution rather than only investigating a subset of possible solutions. The importance of conceptual design to the overall success of the product is crucial as once the conceptual design process has been finished, the majority of product cost and quality has been fixed by selecting particular concepts/solutions as the subsequent product life cycle activities (manufacturing, assembly, use, recycle/dispose) depends on these conceptual solutions. Moreover detailed design and manufacture cannot make up for a poor or inadequate conceptual design. Decisions taken during conceptual design affect all the downstream phases of product life cycle. The solution space for design could be explored effectively and an optimal solution at the conceptual design stage could be generated if the consequences caused by design decisions on later product life cycle stages are considered fully at the conceptual design stage. The subsequent sections in this paper discuss the current on-going research in developing a methodology to proactively synthesize conceptual design of sheet metal components using a knowledge of life cycle consequences approach and particularly manufacturing/assembly consequences.

2. ROLE OF CONCEPTUAL DESIGN IN DESIGN PROCESS

Conceptual design is the initial and the most abstract stage of the design process, starting with required functions (what the product is to do without saying how the design is to do it) and resulting in concepts (preliminary system configurations). Conceptual design can be defined as the transition between four different information states: 1) a set of required functions; 2) a initial set of possible concepts solutions to fulfill the function; 3) a set of behaviors that fulfill the functions; 4) final selected concept(s)/solution(s) that generate the behaviors. Functions, behavior and form have been identified as the major elements of information, which are manipulated in these states. The representation of these elements within the conceptual design process is described in the following sub-sections.

2.1 Functional representation in conceptual design

The word 'function' in design is regarded as a description of the intended action or effect produced by an object [1]. For example the function of shaft can be described as: *a shaft transmits torque.* There are two existing functional representations in mechanical artifact design. One is a natural language-like, non-mathematical representation, where verbs are used to describe what an object does, or is supposed to do. An example would be *what does a shaft do?* A *shaft* (object) *transmits torque* (function). An advantage of this type of representation is that it is close to the way designers express their ideas. But it is difficult to formalize this representation in a generalized way.

The other representation is a mathematical representation of function, where it is expressed as a transformation between input(s) and output(s). It is formalisable and more suitable for a computational environment. However if a man-machine environment is to be provided, using this representation, the commonly used functions expressed in the first representation would have to be mapped into the latter representation before any general functional reasoning support environment could be developed [2]. Similarly a conversion from mathematical representation to a more understandable form by human beings is required after any reasoning. This approach obviously adds additional overheads and can cause problems during reasoning of functions and interpretation of reasoning results. This research used natural language representation of functions in the form of verb (operator)_noun (operand) like *Provide (operator)_ Assembly (operand)*. This representation allows the recall of the precise meaning of function definition at any level of the function mapping and reasoning process.

2.2 Behavioral representation in conceptual design

Function reveals the intentions of the artifact; *form* specifies what the artifact is composed of and how the components are interconnected, while *behavior* spells out how the structure of the artifact achieves its functions. Most existing design systems explicitly represent only form, making little allusion to behavior as a reasoning step between function and structure [3]. However the functional model and form model of artifact/component alone are not sufficient to synthesize the artifact behavior. This is because of the fact that functional models do not adequately capture the interaction of forces and kinematic motions between the part geometry. For instance, the fit condition between a shaft and a bore cannot be expressed by spatial relationship since it does not provide functional design details such as contact pressure, contact force, rotational torque, rotational speed etc. at the shaft-bore interface. Behavior of a function is defined to be the set of values of parameters (which are related casually) of the function either at the specified time or a series over a period of time. Like function, behavior can also be represented by different representation schemes like bond graphs, behavior graphs and natural language type representations. Behavior of a function is context sensitive and as such, behavior comes into play only in the context of a design solution. This research proposes natural language type representation of behavior by using context based reasoning as described in section 4.2 in detail.

2.3 Form representation in conceptual design

The modeling of a form of a mechanical artifact is expressed in terms of its constituent components and sub-components, and the interactions between them. The form of each artifact representation consists of information about

1. Component/sub-component structure of the artifact.
2. Material properties.
3. Typical shape which characterize the artifact as a unique artifact.
4. List of additional features, which would be required to make the artifact work in a real-life environment.
5. The possible modes/situations in which the artifact might fail.

This information has been represented through a number of mediums by researchers in the systems developed to support the conceptual design process. It includes natural language syntax based representations, graph based approaches and graphical modeling of components. Depending on the nature and complexity of information about the form of a mechanical artifact any one of these or all of these methods can be used simultaneously to model the form of the component in each case of a concept/solution generated by the conceptual design process.

3. REVIEW OF RELATED WORK IN FUNCTION TO FORM MAPPING

In the past, the research efforts involving part functions were mainly focused in three major areas: (i) development of standard vocabularies for part functions; (ii) conceptual design with abstract part functions; and (iii) design with spatial relationships. In order to fill up the gap between the functional requirements of the product and rough realizations of concepts/solutions a number of research projects have been carried out to develop methodologies/mechanisms for function form mapping support at conceptual design stage. Some of these approaches are detailed below.

Pahl and Beitz [4] suggested a procedure for deriving the overall functions of a design from the given design problem statements; and then decomposing and recomposing individual sub functions to a hierarchical functional structure that could be mapped to appropriate physical elements. Though the method provides useful suggestions regarding the function decomposition process, it however does not relate the functions to the form of the product. Simon et al. [5] proposed a standardized representation of function consisting of schemes for function and associated flows along with taxonomies of generic functions and flows. Their formal representation provides the means for representing functions that have multiple input and output flow, properties and parameters associated with flows and the decomposition of functions into sub functions each potentially having its own distinct flows. Deng et al [6] developed a dual stage functional modeling framework, which supports functional design knowledge in an object oriented function representation. The identified functional design knowledge is organized into three levels: metal level, physical level, and geometric level.

The review shows that most of the work is done from the perspective of functional requirements/product design specifications only. Little attention has been focused in bringing later product life cycle issues/knowledge at the conceptual design stage, so that the concept/solution generated not only satisfies the functional requirements and context of design, but also takes into account the consequences that are likely to occur in later life cycle stages due to the proposed selected concept. This research proposes a methodology of incorporating life cycle consequences especially manufacturing consequences at the conceptual design stage.

4. FUNCTION TO FORM MAPPING IN THE CONTEXT OF SHEET METAL DESIGN

The term sheet metal refers to metal strip of thickness ranging from 0.3 mm to 5mm. However in some cases this range is not followed strictly. Sheet metal products are made up of different types of materials like ferrous, non-ferrous and alloys. The common functions of sheet metal products are of conveyance nature or of assembly nature. For example in case of conveyance nature functions, the mostly used functions are *convey, channel, direct, divide, guide*, etc. Lots of sheet metal residential & commercial products perform these types of functions such as *Air Intakes, Dormer Vents, Static Louvers, Roof Vents, and Ducts* etc.

In the case of assembly type functions, the mostly used functions are *assemble, constrain, enclose, fasten, fix, guide, join, link, locate, orient, position, support etc.* Industrial sheet metal products perform these types of functions like *automotive body panels, computer casings, electrical control enclosures etc.* These functions are achieved through different *manufacturing features*, which are inscribed on sheet metal during the manufacturing process. In other words, Product Design Elements (PDEs) (see figure 1) at component building level can most likely be a possible means of achieving a desired function requirement, though PDEs at component and sub-assembly levels are also possible. These PDEs include *slot, hole, bend, notch, lance, hem, curl, emboss, bead, rib* etc. Based on this argument, this research proposes the following PDE based function mapping.

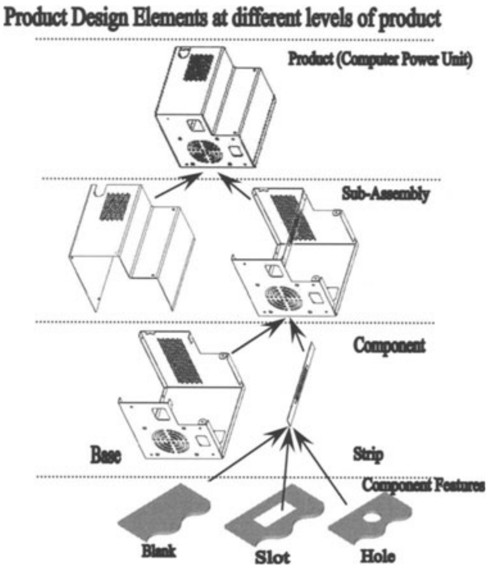


Figure -1. Product Design Elements (PDEs) at different levels of a product

4.1 PDE based design and function to PDE mapping

A PDE is a reusable design information unit (element) representing a potential solution means for a function requirement. Of relevance to this definition and looking from the viewpoint of component construction, a more commonly used term *feature* is considered to be an information element defining a region of interest within a product. A feature is described by properties from several different classes of properties, thus relating these (classes of) properties to one another [7]. It is argued in this approach that a feature is one type of PDEs at component building level. PDEs are hence used as the key to function-oriented design.

Designing by functions or “functional design”, therefore, refers to the process of generating a design solution from a product function point of view, using available well-understood function-PDEs relationships to identify suitable means in the form of PDEs as shown in figure 2. PDEs are the information carriers that allow the mapping between function requirements and physical solutions of a product. They are the vehicles to bring basic design information to the downstream product realization phases for embodiment, detailed part design and later life cycle processes. Figure 3 shows the process of PDE based function design, where a [*Provide-Assembly*] function can be realized by four possible means or PDEs at assembly level of a sheet metal component. This is derived from the mapping search algorithm defined for the mapping process. Similarly when a small function can be realized by a component building PDE, it appears as a potential candidate. [Assembly_slot] PDE is such an example in which a *slot* solution is nominated by the mapping algorithm for designer’s consideration.

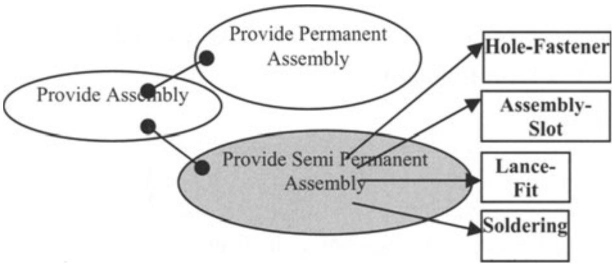


Figure -2. Function-Feature association

4.2 Context based function reasoning

With an extensive function requirement definition, function-PDE mapping can usually produce a long list of alternative PDEs for a designer to consider. This can be a demanding task if each of these PDEs is fully evaluated manually. In addition, the deadline for a design solution can be quite tight. To effectively support designers in these scenarios, this research developed a reasoning mechanism using design context knowledge. Reasoning using design context information spells out the behaviour of a function as discussed in section 2.2.

A design context is defined as the related background information of a design problem under consideration. A good understanding of design context information is essential to successful design and any design support system should investigate how design context information can be used to provide effective support. More specifically, design context information can include:

- Available information of functions defined for the product in its current working knowledge, which is partial solution information generated up till the current stage of the design process for a given problem;
- Information of users of the intended solution products, e.g. age, gender, product preference etc.;
- Materials selected for the solution product, more specifically the properties of the chosen material;

With the above context information available, the context information reasoning mechanism aims to detect any 'unfit' PDE from the initial mapped PDEs. The initial function requirement to **Provide-Assembly** has been matched with four possible means to implement this requirement. Searching the context information of solution material selection reveals that aluminum material has been selected for the solution, and the joining part has plastic material type. This activates a piece of knowledge that *Soldering* means cannot be used for the function, as aluminum components cannot be soldered to plastic components. Timely prompting of this context information about material assists designers to eliminate this infeasible option.

4.3 Manufacturing/Assembly consequence based function reasoning

Life cycle context or more specifically in this paper Manufacture/assembly context information is the definition of life-phase (Manufacture/assembly) systems constraints to product definitions. This group of context information is classified in this research in the library of manufacture/assembly consequences (MACs) as shown in figure 3. Designers are often unaware of these limitations and as design decisions become more related to other factors, it is very difficult, even if possible, for designers to foresee these potential decision consequences. This research isn't intend to exhaust all MACs for all scenarios, however, it does intends at this stage to generate sufficient useful and important MACs for function reasoning. Through the use of these MACs, it is possible proactively to remind designers of the potential consequences of their decisions. For example, committing *Hole-Fastener* as the selected PDE to realize the function to **Provide-Assembly,** triggers of a piece of MACs that violates the design for assembly principle as this decision results in more parts for the design, which in turn increases the time of the assembly process and cost of product.

Through the context information reasoning and MAC based reasoning, suitable PDEs for a function requirement can be reduced. In this example, designers are left to explore the remaining two alternatives (*Assembly-Slot,* *Lance-Fit*). This function mapping, reasoning and decision commitment process can usually continue for other function specifications until suitable means or PDEs are found for all implementable functions.

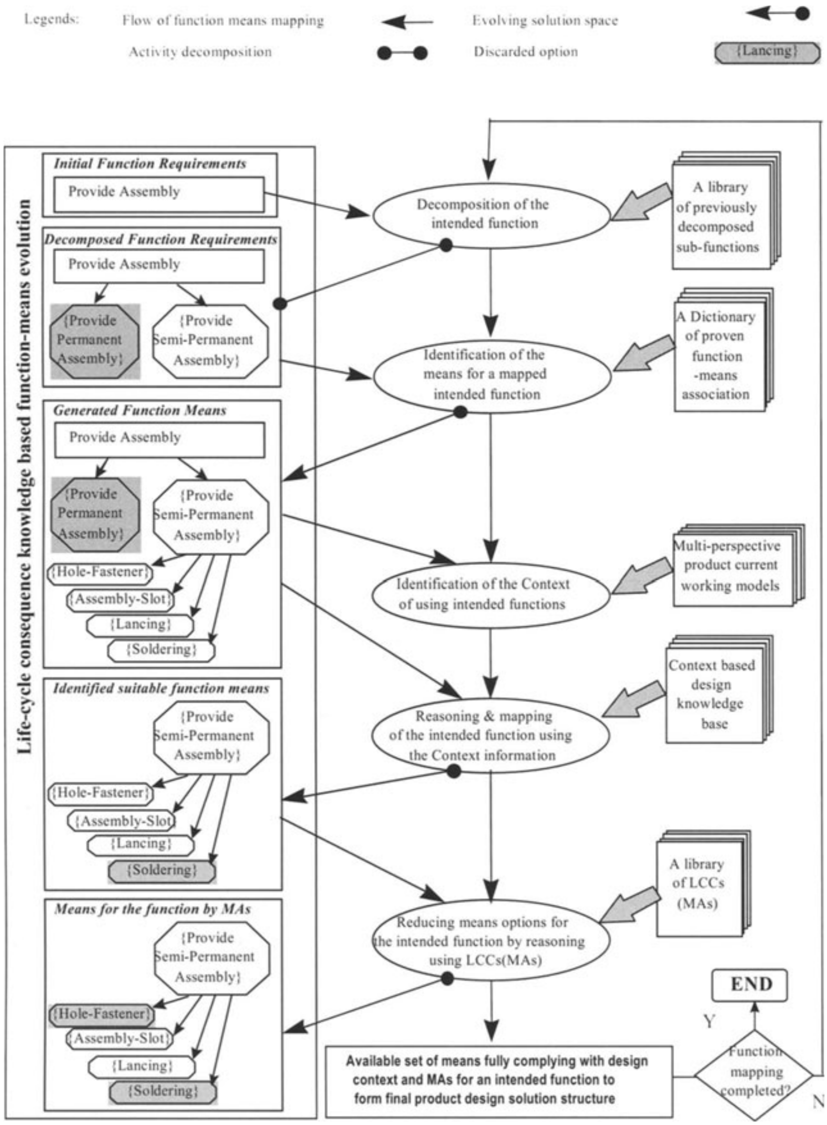


Figure -3. Function to PDE mapping and reasoning process

5. CONCLUSION & FUTURE WORK

This research proposes a PDE based function-means evolution methodology for generating conceptual solutions of sheet metal components through design context based and Manufacturing/Assembly consequences based reasoning mechanisms. Reasoning using design context information and MACs can further assist designers to concentrate on exploring design alternatives and generate more innovative design solutions thus reducing/eliminating the chances of redesign by considering manufacturing implications and increased costs due to the selection of a particular solution earlier at conceptual design stage. This research also highlights certain areas of function-means evolution methodology for future work such as:

- Determine the relationships of different parameters, which can influence the context based function reasoning e.g. the relationship of currently considered function with other functions of components and how this relationship can affect the mapping of considered function to some PDEs
- Extend the scope of this methodology, to incorporate at the conceptual design stage, consequences from other life phases apart from manufacturing/assembly such as the use of sheet metal components and their recycling.

REFERENCES

1. Welch, R.V. and Dixon, J.R., *"Representing function, behavior and structure during conceptual design"*, Proceedings of ASME Design Theory and Methodology Conference DE-Vol. 42, 1992, pp 11-18.
2. Chakrabarti, A., Thomas, P.B., *"An approach to functional synthesis of solutions in mechanical conceptual design, Part I: Introduction and knowledge representation"*, Research in Engineering Design, 1994, Vol.6, pp 127-141.
3. Gero, J.S., Tham, K.W., Lee, H.S., *"Behavior: A link between function and structure in design"*. Intelligent Computer Aided Design, D.C. Brown & H.Yoshikawa (Editiors)
4. Pahl G, Beitz W, *"Engineering design: A system approach"*, 2nd edition, Spring Verlag, London, 1996.
5. Feng, C.X, Huang, C., Kuisak, A., Li, P.G., *"Representation of functions and features in detail design"*, Computer Aided Design, Vol. 28, No. 12, 1996, pp 961-971.
6. Deng, Y. M., Tor, S.B., Britton, G.A., *"A dual-stage functional modeling framework with multi-level design knowledge for conceptual mechanical design"*, Journal of Engineering Design, 2000, Vol. 11, No. 4, 347-375.
7. Brunetti, G., Golob, B., *"A feature based approach towards an integrated product model including conceptual design information"*, Computer Aided Design, Vol. 32, 2000, pp. 877-887.

FUZZY VOLUMES

Principles and Application in the NC machining Fuzzy

K. Benfriha and A. Aoussat

Laboratoire CPI, Ecole Nationale Supérieur des Arts et Métiers, Paris, France

Abstract: The 3D simulation of objects in an evolutionist environment is hard to implement when we consider the amount of parameters and the high cost of computer resources. These facts bring us to introduce the notion of associated fuzzy volumes. This concept is based on the fuzzy geometry principles, which can be used in production with robots assistance. The technique is the associating to complex objects and their environment some fuzzy volumes, constitute of peaks network (points positioned in 3D). In which these points are situated around an object or on its surface in order to approach a target object. With the algorithm developed in this paper, which is based on a genetic algorithm, it is possible to detect interference during the simulation time. The NC machining is used application of this technology

Key words: Fuzzy Geometry, Fuzzy Volumes, Interferences, peaks Network, genetic algorithm

1. INTRODUCTION

The notion of volume is bound to the surfaces constituting its border [1]. For example, one of the necessary conditions for a surface to constitute a volume's boundary is that surface must be closed. The difference between a volumic modelization and a surfasic modezation is the notion of inside and outside, which the user will be able to have access to, which possesses a parameter quantity greater than the geometric elements constituting its border [2].

During the 1960's, several works brought about the implementation of geometric models of surface description in order to make up for the following difficulties [3]: To make available a geometric representation

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 191–200.

permitting representation of surfaces such as planes, cylinders, spheres, cones, torsos and the free surfaces of various shapes. To enable the exploitation of the geometric representation by a production technique in order to be able to materialize the numerical representation of surfaces. To use the knowledge and the ability of technicians so that the geometric representations can be manipulated with minimal mathematical background.

2. FUZZY GEOMETRIC NOTIONS

Fuzzy logic was introduced in 1965 by the Professor of University of California, Berkeley, Lotfi A. Zadeh [4] [5]. The use of fuzzy logic has been directed where it is not possible to express in quantitative terms; the reasoning is used to obtain a result and the classic rules don't give good results.

The other possibility of information is in the works of the fuzzy subset theory, which allows evoking the notion of uncertainty and imprecision notion. It permits:

- To introduce the knowledge in an automatic system,
- To extract the knowledge from numerical data,
- To refine its knowledge considering its mistakes [6],
- Calculus of two dimensions geometric entities [7],
- Calculus of convex envelopes [8],

The fuzzy sets and the fuzzy geometry function on uncertain or imprecise data, but they are fundamentally different because the fuzzy geometry algorithm provides two types of results [9]:

- Geometric results which are continuous and metric,
- Topologic results which are logic at 3 steps (inside, on, outside)

3. INTRODUCTION TO FUZZY VOLUMES

One of the objectives of fuzzy volumes is to simplify some applications: the research in 3D environment of volumic interference cases. Their alleviated modelization allows the treatment by simulation of the complex solid set, which evolves in the same environment, without considering the different geometric models used for solid modelization. Contrary to the title "fuzzy Volume", we do not talk about a volume in the actual sense, but of a set of peaks positioned outside or on the solid surface (volume targets) [10].

These peaks associated to the solid, form two envelopes; the first is a spherical form of proximity, represented by a minimum of parameters, center and radius. Its objective is to deactivate or activate the second envelope. The second is a shape envelope which is constituted of a peak network and positioned in the space, in order to approach the solid. These peaks are interconnected linearly to the center and must include the solid. The existence of a dynamic peak or a dynamic point allows the modification of fuzzy volumes with volume conservation, in order to adjust to certain situations where it is required to have a peak between two frozen peaks. For example, Segments that join peaks, as well as faces, are virtual, because we only use them during the treatment.

3.1 Notation used for a fuzzy volume set

- Fuzzy volumes; $A^f \ldots Z^f$.
- Solids ; $A \ldots Z$,
- Center : $a \ldots z$, (the centre a corresponds to A^f).
- The envelope of A^f is r^a.
- Face: $a_i^{\{P_j\}}$, $i\in\{1,2,\ldots,n\}$, n : number of planes of A^f.
- Peaks: $P_j^{\{a_i\}}$, $j\in\{1,2,\ldots,m\}$, m : total number of peaks.
- Dynamic point d^a of fuzzy volume A^f.
- Connection $P_j^{\{a_i\}} \to a_i^{\{P_j\}}$.

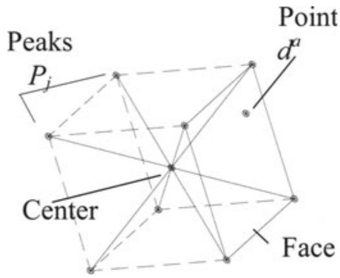


Figure -1. Representation of the fuzzy volume A^f

3.2 Principles

3.2.1 Condition of countenance

Consider solid A. We associate it with a fuzzy volume A^f having a center a, which is the solid center, and linearly joined to peaks P_i $i\in\{0,1,\ldots,m\}$, all expressed in a reference of origin O and axes x, y, z. The condition of countenance is : $A\subseteq A^f$ (than the volume $A^f \geq A$)

3.2.2 Fuzzy error

Each fuzzy volume associated with a solid A, generates a total volumic error $err^f\!=\!A^f\!-\!A$. The rate $1\text{-}(err^f/A)$ represents the fidelity of the fuzzy volume to the solid A. So the number m of peaks depends proportionally on the fidelity rate, and implicitly on the solid complexity, as an example, the peaks of the fuzzy volume (*Figure -1*) symbolize a cube. If we treat a cube solid having the same measurements as the associated fuzzy volume, we get a nil fuzzy error and a fidelity rate equal to 1.

3.2.3 Spherical envelope of proximity

A fuzzy volume can possess a spherical envelope of proximity, with a center *a* that corresponds to the fuzzy volume center, and a radius that corresponds to the : $max\ [a,P_i]\ (1 \leq i \leq m)$.

The principal objective of this envelope is to deactivate or activate the shape envelope (second envelope), because in some applications we treat some evolutionist solids with a complex shape having a distant position in the same environment. In order to have a low treatment time and memory space reserved for this application, we use the advantage that this envelope has only two parameters: the center and the radius.

3.2.4 Dynamic point

The dynamic points $\{{}_1d^a, {}_2d^a, \ldots, {}_ld^a\}$ are the $(m+l)$ peaks of the fuzzy volume A^f, and are able to take a position on the fuzzy volume surface. Their displacement is conditioned by the designation of an outside point P_{ext} , and the minimal distance research criteria, between $[d^a, P_{ext}]$. The outside point can be the center or the dynamic point of another fuzzy volume. The number of the dynamic points depends on the number of objects which can interfere with the fuzzy volume A^f. The behavior of the dynamic point in relation to the outside point P_{ext} is simple because we distinguish three possible cases that can guarantee the minimal distance criteria : On a peak, On the intersection of two facets, Inside the facet area.

3.3 The behavior of the dynamic point

It is a relative behavior to another point (peak), expressed by a position change according to criteria. In order to be able to put the dynamic point in a position which corresponds to criteria of the minimal distance, we proceed by steps. We start with a peak search for a set of the fuzzy volume, closest to the outside point, in order to affect the dynamic point. We have two algorithms. The first is an idiomatic algorithm search that consists in testing all peaks systematically. The other is a genetic algorithm search. The objective of the following step is the identification of segments that correspond to the facet intersection, and to have the dynamic point d^a in common. Now, it is necessary to test the existence of a point on the selected segments that will guarantee our starting criteria. The direct approach used is based on the Pythagorean theorem and a parametric configuration for the segment with an integral polynomial form [11]. In this paper, we use the Bézier model of degree 1.

3.3.1 Evaluation of the dynamic point belonging to a segment

Segments coding :

$a \leftarrow [P_0\, d^a]$, $b \leftarrow [d^a\, P_{ext}]$, $c \leftarrow [P_0\, P_{ext}]$, $d \leftarrow [P_1\, d^a]$, $e \leftarrow [P_1\, P_{ext}]$, $f \leftarrow [P_0\, P_1]$.

$a/f = ((c^2 - e^2)/2f^2) + 0.5$

First result: oddness of report a/f ;

$a/f < 0$	then	d^a is out of segment $[P_0\, P_1]$, or $]- , P_0[$
$a/f = 0$	then	d^a is confounded with the peak P_0
$0 > a/f > 1$	then	d^a is in interval $]P_0\, P_1[$ (*Figure - 2*)
$a/f = 1$	then	d^a is confound with the peak P_1
$a/f > 1$	then	d^a is out of the segment $[P_0\, P_1]$, or $]P_1, + [$ (*Figure -3*)

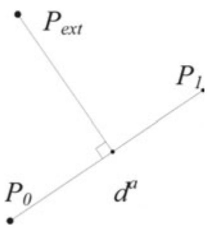


*Figure -2.*Represent point d^a belonging at a A^f segment

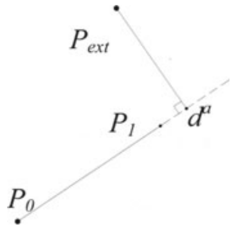


Figure -3 Represent point d^a out of a A^f segment

Second result: calculus of d^a if $(0 > a/f > 1)$, $d^a = F(a/f)$,

$$d^a = P_0 \cdot \left(\frac{1}{2} - \frac{c^2 - e^2}{2 \cdot f^2} \right) + P_1 \left(\frac{1}{2} + \frac{c^2 - e^2}{2 \cdot f^2} \right) \tag{1}$$

At this step, if there exists a dynamic point on one of the segments that satisfy our criteria, we proceed to search one of the two facets concerned with the intersection of the descended segment of the previous step. For this we use a direct approach, having as arguments the coordinates of at least three points that constitute the plane of the facet and the point P_{ext}. This approach is developed in the same way as the previous one.

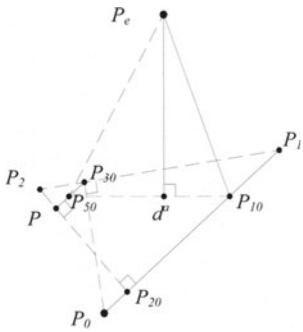


Figure -4. Representation of the *Pext* projection on a plane *P0,P1,P2*

3.3.2 Evaluation of the dynamic point belonging to a plane

From a plane identified by 3 points P_0, P_1, P_2 (*Figure -4*), we determine by a direct form d^a, the calculus of a minimal distance between the point P_{ext} and the plane. The principle is based on the preceding relation (1) of d^a that is applied in several steps. The procedure starts by the selection of a segment $[P_0,P_1]$ in order to search a parallel $[P_{30},P_{40}]$ in the same plane. Next calculate the nearest two points $\{P_{10},P_{50}\}$ at P_{ext}, belonging each to one of the parallel segments $[P_0,P_1]$ and $[P_{30},P_{40}]$. Finally we use these last two points and the outside point $\{P_{10},P_{50},P_{ext}\}$ to calculate d^a.

$$P_{20}=P_0\left(\frac{1}{2}-\frac{c^2-e^2}{2\cdot f^2}\right)+P_1\left(\frac{1}{2}+\frac{c^2-e^2}{2\cdot f^2}\right),\ P_{30}=P_1\left(\frac{1}{2}-\frac{c^2-f^2}{e^2}\right)+P_2\left(\frac{1}{2}+\frac{c^2-f^2}{e^2}\right),$$

$$P_{40}=P_2\left(\frac{1}{2}-\frac{g^2-h^2}{l^2}\right)+P_{20}\left(\frac{1}{2}+\frac{g^2-h^2}{l^2}\right),\ P_{10}=P_0\left(\frac{1}{2}-\frac{s^2-t^2}{f^2}\right)+P_1\left(\frac{1}{2}+\frac{s^2-t^2}{f^2}\right),$$

$$P_{50}=P_{30}\left(\frac{1}{2}-\frac{p^2-q^2}{r^2}\right)+P_{40}\left(\frac{1}{2}+\frac{p^2-q^2}{r^2}\right),\ \text{Result}: d^a=P_{10}\left(\frac{1}{2}-\frac{v^2-w^2}{y^2}\right)+P_{50}\left(\frac{1}{2}+\frac{v^2-w^2}{y^2}\right).$$

4. PROGRESSIVE SEARCH STEPS

The progressive search for the minimal distance between different fuzzy volumes is presented in several steps.

Step 1: Identification of the outside point P_{ext}.

Step 2: If point P_{ext} is outside the spherical envelope of A, $\{P_i\}$ and d^a are not activated and point P_{ext} remains in connection with the center a. In the contrary case, point P_{ext} is in the spherical envelope of A, $\{P_i\}$ and d^a are activated, and we pass to Step 3.

Step 3: Search the peaks $\{P_i\}$ nearest to point P_{ext}, and affect it to the dynamic point d^a ($d^a \leftarrow P_i$).

The search for peak P_i is based on a genetic algorithm of research [12]. In order to optimise the research, this algorithm allows a considerable reduction of the research population (peaks) during iterations.

Step 4: In some cases, step 3 can be insufficient, meaning that a peak exists closer to point P_{ext} than point d^a found in the previous step (that doesn't belong to peaks of volume A). We proceed then to another search that concerns segments binding d^a with other peaks. And step 3 will have served as a starting point for a second search. Link constraints between peaks and facets of a volume are represented very well by a table (*Table1*) that is used for the following treatment without detail.

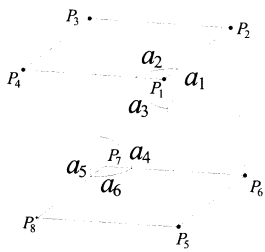


Figure -5. Graphic representation of a fuzzy volume of cube type.

$$d^a = P_i^{\{a_i\}} \rightarrow a_i^{\{P_i\}} \rightarrow \{P_j\}\ \mathrm{i} \neq \mathrm{j}$$

With : $\{a_l\}$ the set of facets having the peaks P_i in common, l the number of facets binding the peaks P_i, $\{P_l\}$ the set of summits set concerned by the peaks P_i

The research is executed between point d^a and peaks $\{P_j\}$, through the set planes of a_l having in common P_i. In order to determine the concerned peaks set $\{P_j\}$ we use this table with a simple manipulation shown in the following example.

We proceed as follows: 1. Localize segments that can be concerned by point d^a. 2. Add a column (Som.) at the end of the last facet, in which we affect point d^a in the intersection of the Som. column and of the line that corresponds to the peak found in the step 2 (in our example it is P_2). We calculate horizontally the sum of elements that we solely consider facets columns concerned by P_2 (a_1, a_2, a_4). Results are affected to the column Som. and interpreted as follows;

Table 1. Table of constraints links

	a_1	a_2	a_3	a_4	a_5	a_6	Som	Test	Rés
P_1	•	•	•				2	×	P_1
P_2	•	•		•			d^a		P_2
P_3		•		•	•		2	×	
P_4		•	•		•		1		
P_5	•		•			•	1		
P_6	•			•		•	2	×	
P_7				•	•	•	1		
P_8			•		•	•	0		

a_1	a_2	a_3	a_4	a_5	a_6	Som	Test	Rés
•	•	•				P_1		
•	•		•			P_2		
	•		•	•				
	•	•		•		1	×	P_4
•		•			•	1	×	
•			•		•			
			•	•	•	0		
		•		•	•	0		
	×							

- 0 : Indicate that the peak P_i or d^a has a volumic connection with elements having a 0 as indication of summons.
- 1 : Indicate that the peak P_i or d^a has a surfasic connection with elements having a 1 as indication of summons..
- 2 : Indicate that the peak P_i or d^a has a direct connection with elements having a 2 as indication of summons.

In this step it is the last result that is going to be exploited, either: $[P_2,P_1]$, $[P_2,P_3]$, $[P_2,P_6]$. The second approach is the selection of only one segment concerned by the existence of point d^a which can satisfy the condition:

$$[P_i,P_j] \perp [d^a,P_{ext}].$$

The procedure of selection used to this effect is the condition:
$d^a \in [P_i,P_j]$ **if** $0 \leq (5 - R) \leq 1$ ($R = (\text{dist}^2[P_i, P_{ext}] - \text{dist}^2[P_{ext}, P_j]) / (2.\text{dist}^2[P_i, P_j])$for every segment. We finish the step by the computation of point:
$d^a = P_i(0.5 + R) + P_j(0.5 - R)$.

Step 5: Often step 4 can be insufficient, which means that a peak exists closer to point P_{ext} than point d^a found in the previous step and that doesn't belong to segments. In this case we proceed to a search that concerns facets. In order to determine which facets are concerned by this step, we use the coordinates $[P_i,P_j]$, descended from the previous step and the link's table of constraints.

We continue with the preceding example in order to explain this step. Example 1 : Once the point d^a is found and the segment $[P_1,P_2]$, we put the previous result in the column Som., we deactivate lines that already underwent a test in the previous step as ; P_1, P_3, P_6. The first procedure is to select facets having in common $[P_1,P_2]$, by searching for a link in their corresponding lines, as : a_1 and a_2. The sum of columns a_1 and a_2 must be used to determine peaks concerned by the belonging test (column Test.) as : P_4, P_5, in order to get peak P_4 (column Res.). We localize the facet concerned by the intersection of the line P_4 and columns a_1 and a_2. We get the facet a_2, that possesses the nearest point to point P_{ext}. We finish the step by a calculus procedure of point d^a. The specificity of this procedure is owed to the rectangular shape of the facet. Because the existence of the parallel segment [P3,P4]⁄⁄[P1,P2], avoids us to calculate a segment [Px,Py]⁄⁄[P1,P2] necessary to determine d^a. as example, the triangular forms of facet incites us to pass by the calculus of a segment [Px,Py]⁄⁄[P1,P2].

$R_1 = (\text{dist}^2[P_i, P_{ext}] - \text{dist}^2[P_{ext}, P_j])/(2.\text{dist}^2[P_i, P_j])$, $d^{a1} = P_i(0.5+R) + P_j(0.5-R)$
$R_2 = (\text{dist}^2[P_i, P_{ext}] - \text{dist}^2[P_{ext}, P_j])/(2.\text{dist}^2[P_i, P_j])$, $d^{a2} = P_i(0.5+R) + P_j(0.5-R)$
$R = (\text{dist}^2[R_1, P_{ext}] - \text{dist}^2[P_{ext}, R_2])/(2.\text{dist}^2[R_1, R_2])$.

$$d^a = P_i(0.5 + R) + P_j(0.5 - R)$$

Step 6 : it is a logical test that indicates the position of P_{ext}, with regard to the fuzzy volume; it implies that it is inside or outside the volume (*Figure -6*).

if ($[a,P_{ext}] \geq [a,d^a]$)
then P_{ext} is outside **else** P_{ext} is inside.

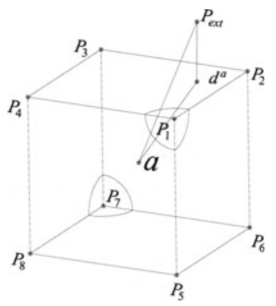


Figure -6. Representation of the P_{ext} , with regard the point d^a and the centre a.

5. APPLICATION

We treat in this application the pocket evidently in 3D (*Figure -8*) with only one islet. This allows us to observe the evolution of the tool in the stain space, represented by a marked geometric place. This geometric space includes all technological elements requisite for the stain realization.

We associate then fuzzy volumes to each element (*Figure -7*). This application is not detailed and is limited solely to the volumic interference detection state and the treatment after detection doesn't concern this paper.

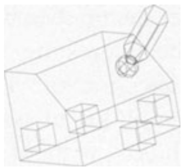


Figure -7. Fuzzy volumes association.

Figure -8. Pocket evidently.

General Algorithm
begin

Initialisation ; Initial Configuration of manipulator $(\theta_5, \theta_1, d_2, d_3, d_4)^T$,

$i \leftarrow 1$, Test $\leftarrow$ False, $N \leftarrow$ Total number of configurations

do Configuration i $(\theta_5, \theta_1, d_2, d_3, d_4)^T$

Test $\leftarrow$ Procedure : inference research

if (Test = True) then {interference existing}

else { $i \leftarrow i + 1$}

until ((Test = True) or ($i = N+1$))

end

5.1 Simulation

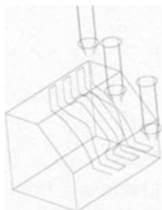


Figure – 9. Iso-parametric zigzag trajectory, 3 axes NC .

We simulate this example on two machining of different axes, and having in common the same machining cycle, step, tool and cut parameters. The first is a 3-axis site (TTT) that constrained the tool to displace itself in the task space in stationary orientation, which provokes in our case interference between the inactive tool part and the manufactured part (*Figure 9*) during the execution phase. The second case is a 5-axis site RTTTR having an orientation q_5 of the tool, which allows during the displacement to orient itself according to the normal trajectory in the iso-parametric planes, so there is no interference (*Figure 10*). Technically the addition of axes to a NC machining can be a solution to avoid collisions between objects, but we have to consider the economic aspect that influences this choice, and often leads technicians to look for other solutions.

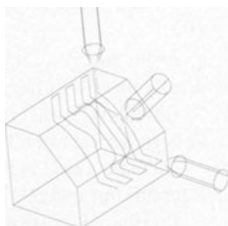


Figure – 10. Iso-parametric zigzag trajectory, 5 axes NC.

6. CONCLUSION

The fuzzy volume contribution in manufacturing is interesting because they intervene in important phases. The 3D simulation of a complex production site includes a set of technological elements such as the immobile units of manufacture, assembly or measurement. Or/and the other mobile units (as the mobile manipulators) allows the production to increase flexibility. It would be interesting to integrate the fuzzy volumes, which mainly concern the detection of volumic interferences (collisions). Their advantages offer the possibility of associating fuzzy volumes to :

- A set of volumic objects, without necessarily having, the same basis geometric model,
- An industrial environment,
- Complex objects.

However this survey includes some aspects that must be further investigated in order to automate the association of fuzzy volumes and to integrate them in a manufacturing system.

- A minimization aspect of the fuzzy error,
- An optimisation aspect of fuzzy volume.

7. REFERENCES

[1] Mortenson, M.E., (1985), Geometric modelling, John Wiley & Sons, New York.

[2] Goldman, R.N., (1987), The Role of Surfaces in Solid Modeling, in Geometric Modeling : Algorithms and new trends, G.E. Farin, Paris.

[3] Léon, J.C., (1991), Modélisation et construction de surface pour la CFAO, Hermès, Paris.

[4] Zadeh, L.A., (1965a), “Fuzzy sets”, Information and Control, Vol. 8, pp. 338-353.

[5] Zadeh, L.A., (1968), “Fuzzy algorithms”, Information and Control, Vol. 12, pp. 94-102.

[6] Dahmani, Y., Noubat, W. and Benyahou, A., (1999) “Optimisation de système d’inférence flous par la méthode

[7] Chauduri, B.B., (1991b), “Some shape definition in fuzzy geometry of space”, Pattern Recognition Letters, Vol. 12, pp. 531-535.

[8] Chauduri, B.B., (1991a), “Fuzzy convex hull determination in 2d space”, Pattern Recognition Letters, Vol. 12, pp. 591-594.

[9] Foufou, S., Brun, J.M. and Bouras, A., (1997), “Intersection de surface paramétriques : apport de la géométrie floue”, Revue de CFAO et d’informatique graphique, Vol. 12, N. 4, pp. 337-349.

[10] Benfriha K. and Aoussat A. (2002), “Volumes flous, principes et application au site d’usinage CN”, IDMME 2002, Clermont-ferrand, France.

[11] Léon, J.C., (1991), Modélisation et construction de surface pour la CFAO, Hermès, Paris.

[12] MITSUO GEN, RUNWEI CHENG. *Genetic Algorithms and Engineering Design*, Editions John Wiley & Sons, Inc., New York, 1997.

A NORMALIZED PRODUCT STRUCTURE MODEL FOR MANUFACTURING

Definition and implementation of the TMC concept, adaptation of the new technologies of material cutting

L. Deshayes, J-F. Rigal, P. Ghodous* and R. Gervais*
INSA de Lyon, France
** LIGIM - Université Claude Bernard, Villeurbanne, France*

Abstract: In the field of manufacturing a great number of industrial software tools integrate a cutting tool modeling. For some cases the cutting tool is considered as a resource for the various functions of the company. For other experts, it is for supplying the cutting parameters using data bases or experimental and numerical methods. The concept of Tool Material Couple (TMC) is one of these methods. To improve the performance of manufacturing systems, one has to facilitate collaborative work by developing tools for the concurrent engineering. The presented study is situated in this frame and proposes a normalized data model. This generic model based on the product structure model of the STEP standard uses the TMC concept to test its validity as consensual definition of the cutting tools and the cutting parameters. We also demonstrate that improvements of the concept TMC are necessary to consider the new cutting technologies, such as Hard Turning.

Key words: manufacturing, TMC, Concurrent Engineering, STEP standard, hard turning.

1. INTRODUCTION

This paper concerns the static description and the data of the cutting process using the Tool Material Couple standard (TMC) [1]. This TMC standard reached now an industrial maturity and has accumulated numerous experiment results. It is possible to consider its data as a common base for the various experts of manufacturing process. The purpose is to test the homogenization possibilities of different cutting approaches using the TMC.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 201–210.

First works [2] [3] have shown the interest of a concurrent approach in cutting process based on the TMC concept.

The first part of the article proposes a study of the computing systems of optimal cutting parameters (cutting software) and of a Computer Aided Manufacturing software (CAM software). This analysis allows to define the data and to build the product structure model associated to the two points of view. The second part proposes a product structure model taking into account both previous points of view as well as results of research on mechanical turning modeling. The proposed model is instantiated on a real and a simple case to validate the model. The last part shows that to improve a cooperation among different manufacturing experts, new experimental methods for considering cutting parameters have to be developed. The case of hard turning is considered and the study of the minimal cutting speed (Vcmin) determination shows clearly other components of the machining system to consider. This analysis shows that with new turning processes - such that hard turning - the evolution of the CTM's notions for a best process control have to be considered.

2. DIFFERENT CUTTING POINT OF VIEW

The methodology for developing application protocols (AP) from the international standard for the data representation and exchanges STEP [4] is used for modeling the point of view of different cutting experts. The figure 1a represents this methodology. The user needs are analyzed through two models: - The Activity and Application Model (AAM) describes the activities, the processes and the data flow of an application, -The Application Reference Model (ARM) is an information model written in acceptable and general terminology for the application. It can be represented by EXPRESS_G (graphic representation of EXPRESS) [5].

From these two models, the STEP integrated resources are specialized to satisfy the information needs of the associated model ARM. The result of this specialization is the AIM (Application Interpreted Model) model formalized in the EXPRESS language. Diagrams proposed afterward are represented in EXPRESS G. This graphic language allows to build diagrams characterized by a level which can represent that of the entity or that of the schema (including a set of entities).The representation of the entity is a box in fine full lines. See the entities **cutting tool** and **turning tool** on the example, figure 1b. The entities contain classified data according to three types:

- The types of simple data which can be INTEGER, REAL, (STRING). The symbol of a data type is a fine full line box with a double bar to the right, as the type STRING figure 1b.

- A type aggregated of data such as LIST (of element), ARRAY (of elements),… etc. On our example, the attribute name is a list (from 1 to 4 names trademark, common name …).

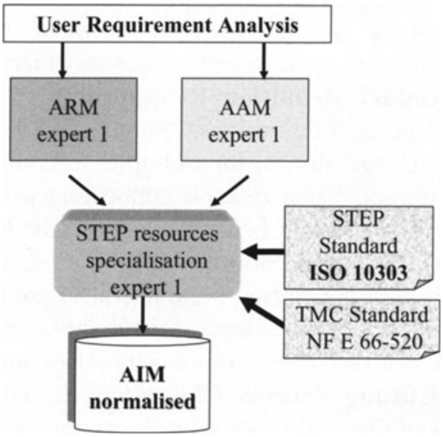


Figure -1a. STEP methodology

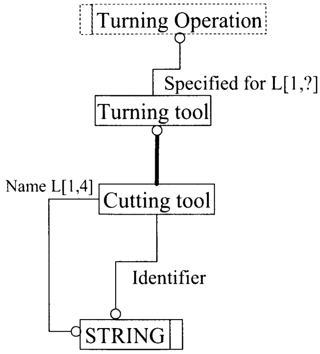


Figure -1b. EXRESS-G model example

The type builds of data SELECT represented in dotted line with two double bars to the left (type of data turning operation). So the turning tool is specified for a list of one to n types of turning operations which can be the roughing in general turning, facing etc. …

Finally, EXPRESS G allows to represent the relations between entities by means of three notations: the reference relation (continuous fine line), the relation of inheritance (continuous thick line), and the relation of optional reference (intermittent line). The direction of the relation is characterized by the symbol 'o'. So, in the figure 2, the entity **cutting tool** references the data type STRING for its attribute Name, and the **turning tool** entity inherits characteristics of the entity **cutting tool** entity. We pass on the modeling of the schema which would be only weighing down our study and hampering the legibility of diagrams. So, afterward schema are represented as entities.

2.1 CAM and Cutting expert product structure model

The cutting software and CAM activities have been analyzed permitting to build the AAM and ARM models. The structure model according to the STEP standard [7], is used to build the AIM structure model of each expert. These models are presented in figure 3a for the cutting expert and figure 3b for the CAM expert. To simplify the lecture we present only the Cutting Expert's model. We considered a cutting process like constituted of three product instances which are: the insert, the tool holder, and the manufactured part. These instances inherit the characteristics of the **Product** generic entity. The machine is not considered like an instance of the product but like an extern constraint on the TMC system. The entity **Product_Version** which describes the evolution during time is associated at the product entity. The **Product_Definition** entity characterizes the product during its life cycle. The **Constituent_material** model is associated to a product definition. Indeed, the mechanical characteristics of the material between two definitions of the product can be modified by different processes (thermal treatments for example). The **Product_Definition_Relationship** entity models the links between the various product definitions. Among these links the relation **Assembly_Component_Usage** allows, for example, to connect a tool with a tool holder. The **TMC** model is a relation connecting a tool with a material (of the work piece). This relation is used to choose the best cutting parameters. The **Product_Definition_Shape** entity connects a couple of products (tool/work piece) with the cutting parameters. These last has to correspond to the TMC model. Indeed, the shape of the manufactured product allows to select a given tool, and conversely a tool can realize only a certain type of operations. The **Cutting_Process_Characteristic** entity possesses the specific characteristics of the workpiece. This entity references the **Product_Definition_Shape**. The **Cutting_Process** model represented here in the form of an entity is connected by a relation of related / relating with the product definition shape entity. A cutting process changes a product shape into another defined by the process.

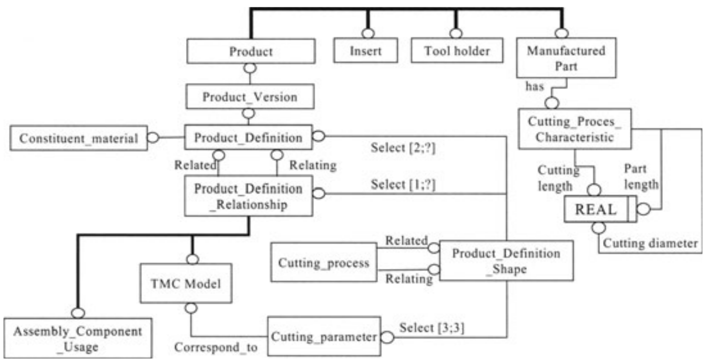


Figure -3a. Cutting Expert's product structure model

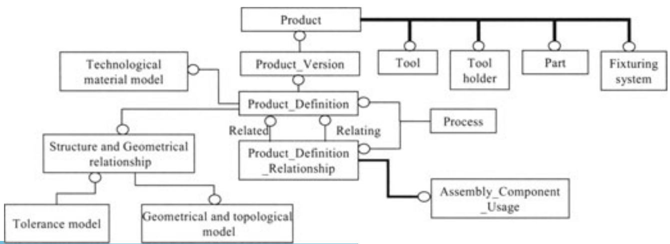


Figure - 3b. CAM Expert's product structure model

2.2 Cutting software and CAD-CAM software Data integration

Three problems have to be solved to insure the integration of both previous data models: - a semantic or vocabulary problem, -establish a correspondence of the data formats and, -a problem of data model structure. Transformation tables have been used [8] to make such correspondences and to create a generic model common to the two experts. To integrate these two models, the proposed generic model figure 6 and commented in the following paragraph facilitate the adaptation. In the studied CAD-CAM software, the TMC data (depth of cut ap, feed rate f and cutting speed Vc) are exported from the data base of the cutting software. No major incompatibilities have been observed. But for the tool descriptions this is more difficult because the **Tool** entity in the CAM model, figure 3b, does not have an explicit correspondence with the **Insert** entity from the cutting expert data model. This semantically correspondence in data modeling is implicit, but that discussion is beyond the scope of this paper.

3. PROPOSITION OF A GENERIC STRUCTURE MODEL FOR THE CUTTING FIELD

Data of two cutting process points of view have been analyzed in the previous paragraph. The objective being to propose a cooperative engineering platform, it is necessary to build a normalized model taking into account a maximum of common data. Each expert should be able to specialize these data considering its own model. The model comparison, allows to propose the generic structure model figure 4.

3.1 Proposition

The model, figure 4, possesses the characteristics of the product structure model of the STEP standard in which were added on one hand the common data to the two software described in the part 2, and on the other hand the useful data for the manufacturing engineering. The cutting tool suppliers point of view was also considered. To pursue the objective of the TMC concept improvement, we privileged the entities used in the standard [1]. The data concern the TMC model, the tool description, the manufacturing processes and the material data of the tool and of the part. The material model was developed by considering at the same moment the tool and the work-piece materials [9]. The **Process** entity defined in CAD-CAM software (figure 3b) on both **Product_Definition** and

Product_Definition_Relationship entities is resumed under the entity **Cutting_Process** by introducing a related/relating relation with the **Product_Definition_Form.** This last entity is similar to the **Product_Definition_Shape** entity of the cutting software structure model (Figure3a). This entity allows the representation of the **structure and geometrical relationship** schema of the product structure model in CAD-CAM systems and uses the **Tolerance model** and the **geometrical and topological model** such as modeled on the figure 3b. On this figure 4, it was also added the notion of **Elementary_Cutting_Feature** entity which corresponds to the list of the entities which can be realized by two definitions of product instances. The TMC model and the relations with the other entities of the structure model such that defined by the cutting software are preserved. The only difference being we let the cutting parameter referenced by the cutting process. Such modeling allows to consider more process information when selecting the appropriate cutting parameters.

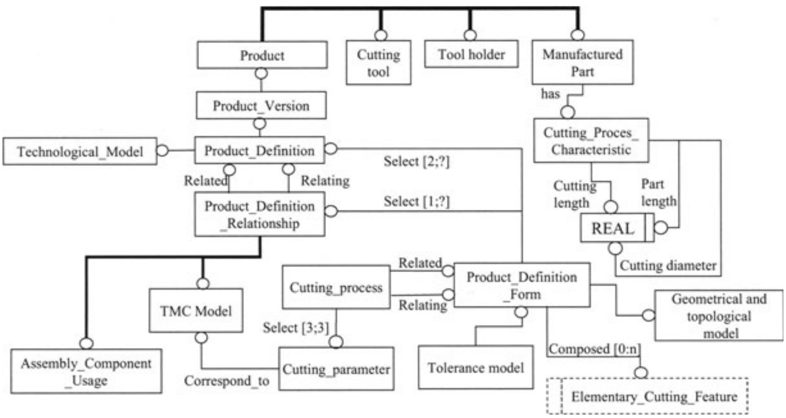


Figure - 4. Common structure model to different cutting expert

3.2 Application with an instantiated model

To illustrate the approach, a simple mechanical part is instantiated. The application concerns the turning process of the work piece defined figure 5. To facilitate the reading the application is limited to the outside profile (surface no 3, operation 20). The **Product_Definition_Form** (DF) before roughing of profile A allows to characterize the definition 1 of the part (*Def 1 part*) which corresponds to the geometrical characteristics before the operation. *DF before roughing A* also characterizes the Cutting_Feature associated to the profile A (CF_A). The relation of the model TMC (M_TMC) allows to propose the convenient cutting parameters (PC)

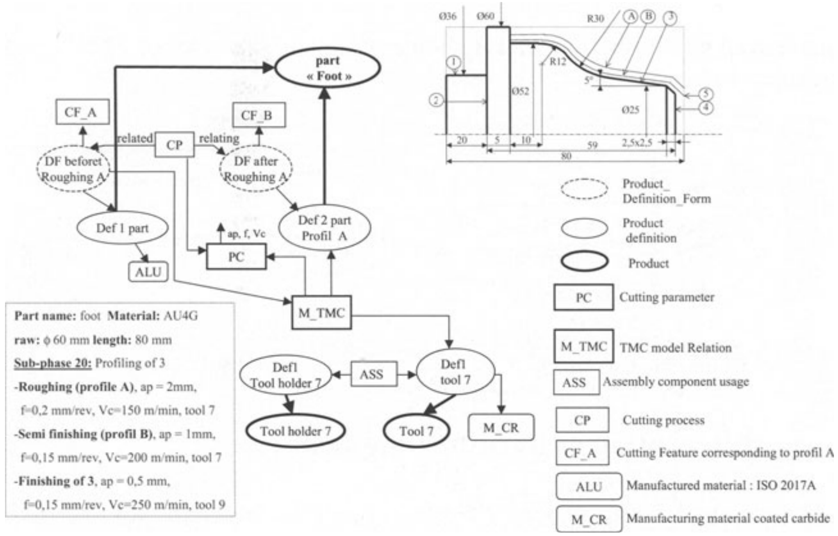


Figure - 5. Instantiated generic cutting structure model

associated to the entity CF_A and corresponding to the couple of product definition *Def 1 tool 7* and *Def 2 part Profile A*. The cutting process (CP) uses the cutting parameters and the *DF before roughing A*, to realize the **Product_Definition_Form** corresponding to the profile A (DF after roughing A). The **Product_Definition** corresponding to the profile A is then obtained. One proceeds in the same way for the profiling operation continuation. In every stage a new definition of the work piece is obtained. The assembly relation (ASS) represents the link between the tool and its tool holder. The entity material manufacturing coated carbide (MU_CR) indicates the material composition of the cutting tool. The entity (ALU) represents that of the work piece (ISO 2017A). This instantiation validates the generic structure model proposed. Let us note in particular that the entity TMC satisfies the requirements expected as a common concept among various experts.

4. CTM'S IMPROVEMENT FOR HARD TURNING

On this point of the TMC, the global objective of the work is to simplify the steps for obtaining cutting conditions by using on one hand the information methodologies presented above and in a second hand mechanical methods. The static modeling of the data allows to fix the concerned data and the connected attributes. As we are going to see it afterward in the case of the hard turning, the exploitation of the TMC method led to interpretation difficulties. We take like example, the minimal cutting speed determination whose choice is not easy for a lot of TMC due to the mechanical behavior.

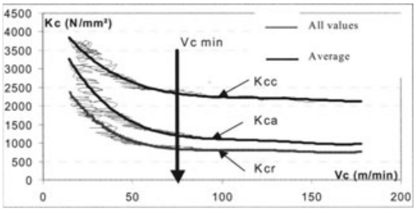


Figure - 6a. Vc min stage in the case of conventional turning

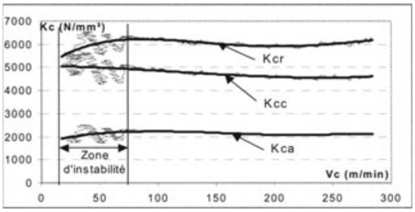


Figure -6b. Vc min stage in the case of hard turning

4.1 Integration of hard turning behavior

The minimal cutting speed (Vcmin) choice is important for the TMC stages [1]. This parameter is sometimes difficult to put in evidence. In [1] Vcmin is defined as being the speed from which the tool life wear is repetitive. Figure -a presents the specific cutting pressures (Kc) according to the cutting speed on a classical TMC or conventional turning. Two zones are distinguished. The first one for Vc's small values presents important variations. For the second one , for Vc's higher values, the variations are less important. The limit of this 2 zones is the minimal cutting speed Vcmin. The curve figure 6b, realized in the case of hard turning doesn't show significant uncoupling. Let us note that the diminution of Kc according to the speed is very weak on the speed range from 75 to 300 m/min. Nevertheless, it can be observed that the curves between 20 and 75 m/min are much more perturbed. than after 75m/min. Before 75m/min the KCC oscillations have an amplitude of 30 % from their mean value. This variation is reduced to about 6 % for greater cutting speed values. Due to this observation it can be considered that the Vcmin is of the order of 75 m/min. According to [9, 7] this observation becomes widespread in hard turning. The new definition for Vcmin can be a the minimal speed value for which the process is dynamically stable with a good tool life repetitiveness. In the case of hard turning the instability in the little speeds presents strong vibrations. This phenomenon is due to the dynamic of the whole manufacturing system. The tool, the workpiece and the whole machine tool are concerned and called the Part, Tool, Machine (PTM) system.

4.2 Relation between PTM and data model

Now the question about the dynamic management of the normalized data model figure 4, must be solved. The first modification concerns the semantic definition of the minimal cutting speed, became an acceptable cutting speed.

This modification can be easily passed on to the various experts associated because it is a neutral and generic model. On the other hand the notion of acceptable Vc possesses an important impact on the stages for obtaining its value. The notion of Kc's stability is added in implies to consider the set PTM. The modification of the normalized model, figure 4, considering these information are presented Figure 7. So the entity **PTM model** is introduced as under class of the entity **product definition relationship**. Now the **TMC model** entity connects two product definitions, while the PTM model entity connects three entities: the part, the tool and the machine. We can model this difference by introducing the machine entity and defining the PTM model on the machine entity. This modeling round off the model defines figure 4 without change of the global model. On the other hand two methods of cutting parameter determination are considered: the method normalized by the TMC concept and the future method due to the PTM concept. Nowadays the TMC method is much more mature than the PTM. Thus the improvement of the TMC method has to give directions for elaborating the bases of the PTM to represent new technologies such as the hard turning. But, In practice, the changes introduced by this new entities provoke a redefining of the data bases. This dynamics is very badly managed in STEP. Methods of opened communications using the WEB technologies connected to a shared data base system should be used during the implementation.

5. CONCLUSION AND PERSPECTIVES

The comparison of the product structure model of two manufacturing expert allowed to propose a generic model integrating the point of view of the CAD-CAM, but also that of the TMC standard. The model built on this base allows to validate the use of the TMC concept as a set of consensual definitions in the various domains of the cutting process. A computer

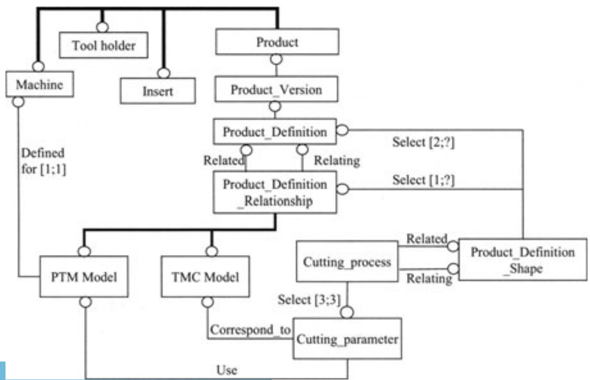


Figure - 7. Integration of the PTM model to the generic structure model (figure 5) development will validate completely the study. Data bases corresponding to the models above will be implemented (under Access or Oracle). Web technologies like XML and EIP (Enterprise Information Portal) will be used to exchange and share the information via web pages. The consideration of new technologies such as the hard turning pulls an overtaking of the notion of Tool Material Couple to introduce that of the Part Tool Machine system. The PTM modeling is still at the beginning of developments and is far from equaling in maturity that of the TMC. So, to keep a good industrial reality we suggest on a first approach to build PTM's notion and exploiting it in addition to the TMC. It is envisaged that in the future, both models coexist. The TMC model alone for the conventional manufacturing, and the TMC and PTM models together for the new manufacturing technologies such as hard turning, high speed cutting etc. The elaboration of the PTM model leads to a new research of significant TMC parameters. It is the objective fixed of the development of two additional methods: cooperative modeling (to define common concepts to the various manufacturing experts) and analysis of the involved mechanical and experimental models.

6. REFERENCES

[1] **NF E 66-520**, "*Working zones of cutting tools – Couple Tool-material – part1 : general presentation*". ISSN 0335-3931. Sept 1997.

[2] **L. DESHAYES, J. F. RIGAL** "*Capitalisation des connaissances dans un contexte d'ingénierie coopérante pour la gestion de données de coupe*", Journée thématique primeca, pp 15-22 , Mai 2001

[3] **L. DESHAYES, C. DARTIGUES, P. GHODOUS, J.F. RIGAL** « *Distributed and standard system for cutting data management* » CE 2001; Anaheim, USA 2001, pp. 310-316

[4] **P. GHODOUS, D. VANDORPE**, *"Integration of design and manufacturing data model using STEP"* , Proceedings of the 1st IDMME Conference, Nantes, France, 1996, pp. 445-458

[5] **ISO 10303-11**, STEP Product Data Representation and Exchange "*Part 11, Description Methods: The EXPRESS Language Reference Manual*", IOS, NIST, 1994.

[6] **Délivrables D1.2** et **D4.1,** « *identification des paramètres significatifs du COM* » et « *Modèles normalisés de coupe* », Rapports projet région Rhône-Alpes 2IUF

[7]. **ISO 10303-43**, STEP Product Data Representation and Exchange "*Part 43, integrated Generic Ressources: Representation structures*", NIST, 1994

[8] **R. GERVAIS**, "*Modélisation des informations pour la conception et fabrication coopérante*", rapport de DEA, Juillet 2001, 50 p.

[9] **M. REMADNA, "**Le comportement du système usinant en tournage dur. Application au cas d'un acier trempé usiné avec des plaquettes CBN (Nitrure de Bore Cubique)", Thèse INSA- Lyon 2001, 243 p.

PART ORIENTATON IN LAYER-BASED MACHINING

Y. H. Chen, Z. Y.Yang and W. S. Sze
Department of Mechanical Engineering, University of Hong Kong

Abstract: Determination of build orientation for layer-based manufacturing (LM) can be considered as a multi-criteria optimization problem. The preferred build orientation should have the tendency to maximize surface quality, minimize build time and build cost simultaneously. For a given part model, different build orientation will result in variant surface quality, support design, number of stock layer, removed material volume and part stability. In this paper, determining the build orientation is modeled as a fuzzy decision making problem. Fuzzy sets are employed to rate the contribution made by alternatives. The preferred orientation is then chosen according to its rank in the result of fuzzy synthesis evaluation.

Key words: layered manufacturing, build orientation, fuzzy set

1. INTRODUCTION

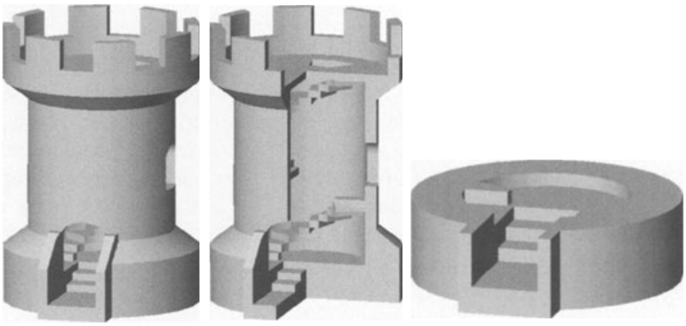
Nowadays, LM technologies have been widely used by product designers, tool designers, and manufacturing engineers. A crucial step in LM is choosing a preferred direction for the model to be built in, i.e. *build orientation*. Build orientation affects the part accuracy, surface quality, build time, and build cost. Because not all the desired effects can be achieved in one build orientation, some trade-off in the determination of build orientation must be made. Part accuracy, build time, support structure, and part stability are the main factors to be considered.

In current commercial LM systems, the build orientation is often chosen manually based on experience. In order to make this operation

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 211–220.

automatic and robust, much effort has been invested [1],[2],[3],[4],[5]. Although these methodologies have proved to be effective for traditional LM processes, none of them can be applied to layered-based machining.

Layer-based machining has attracted many research efforts recently because it is a combination of additive and subtractive method. Figure 1(a) shows a castle model. The interior is rather complicated as in Figure 1(b). Through layer-based machining (Figure 1(c)), a large castle model can be built by a large variety of materials. Horváth and Vergeest *et al.* [6] used an electronically/mechanically controlled flexible blade to manufacture free-form front surface layer by layer to eliminate the staircase effect that is inherent in current LM systems. However, the shape of a part is limited to simple geometry. Song and Chen [7] developed a robot based layered machining system for rapid prototyping applications. As the robot based layered machining system is described in a previous paper [7], this paper is focused on build orientation determination.



(a) A Castle model (b) Section view (c) Sliced part

Figure – 1. Model shows the improved accessibility by slicing

In traditional LM processes, the main consideration for build orientation includes: part accuracy, build time, build cost, the number of support, support-contact area, trapped volume, and part stability. In our system, all factors except trapped volume will be considered. In addition, tool accessibility is taken into account because the stock layer thickness is much larger than that in traditional LM (typically 10mm vs. 0.1mm).

2. FUZZY SETS

Fuzzy decision making has been used extensively in solving engineering problems [8],[9]. In this paper, fuzzy sets are employed to rate the active factors with respect to their “satisfaction degree”. Let x denotes the singletons that make up the universe of discourse X. Various combinations of these singletons make up sets, say A, on the universe. If the universe of discourse X is discrete and finite, fuzzy set $\tilde{A}$ can be expressed as:

$$\tilde{A} = \left\{ \frac{\mu_{\tilde{A}}(x_1)}{x_1} + \frac{\mu_{\tilde{A}}(x_2)}{x_2} + \cdots \right\} = \left\{ \sum_i \frac{\mu_{\tilde{A}}(x_i)}{x_i} \right\}, \qquad (1)$$

where $\mu_{\tilde{A}}(x_i) \in [0, 1]$ is the membership of x_i, which represents the degree of membership of x in $\tilde{A}$. An important application of fuzzy transform is *fuzzy synthetic evaluation* (FSE in brief). The following steps are typically included in FSE [10]:

1. Define the universe of factors:
   $U = \{u_1, u_2, \ldots, u_m\}$;
2. Define the universe of evaluations:
   $E = \{e_1, e_2, \ldots, e_n\}$;
3. Define weight vector, whose elements are the weight assigned to each factor:
   $W = \{ w_1, w_2, \ldots, w_i, \ldots , w_m \}$, where $\sum_i w_i = 1$;
4. For an alternative a_i, construct the single-factor fuzzy relation matrix $\tilde{R}_i$:
   $\tilde{R}_i = [r_{ij}]$,
   where $r_{ij} = \mu_{eiuj}$ is the membership of u_j in e_i;
5. Calculate the evaluation fuzzy vector v_i for alternative a_i:
   $$v_i = W \otimes \tilde{R} = \{v_{i1}, v_{i2}, \ldots, v_{ij}, \ldots, v_{in}\}^{\mathrm{T}}, \qquad (2)$$
   where $\otimes$ represents the composition operation;

The alternative a_i is evaluated as the evaluation with respect to $max(v_i)$ according to maximum membership rule.

3. MAJOR FACTORS AND THEIR FUZZY SET REPRESENTATION

The contributions of each alternative to the factors are defined in this section. Fuzzy membership functions are used to represent the various ill-defined boundaries of categories in the evaluation set.

3.1 Part stability

Part stability factor is defined by the following equation:

$$F_{ps} = [\frac{A_{bp}}{A} + (1 - \frac{h_{cg}}{D}) + (1 - \frac{d}{D})] / 3, \qquad (3)$$

where A is the area of the object convex hull surface with the maximum area, D is the diameter of the part, which is treated as a polyhedron, A_{bp} is the area of base plane, h_{cg} is the height of *center of gravity* (CG). d is the distance between the projection of CG onto the base plane and the boundary of the convex hull of base plane. If the projection of CG is within the convex hull of the base plane such as CG'_2, d is set to 0. From the definition, $F_{ps} \in [0,1)$.

3.2 Inaccessible area

Inaccessible area factor is defined as:

$$F_{is} = 1 - \frac{A_{ia}}{A_t}, \qquad (4)$$

where A_{ia} is the inaccessible area, A_t is the total surface area of the part.

Dexel encoding method [8] is employed to find out A_{ia}. DexelEncode function encodes 3D parts into the dexel model as below:

DexelEncode (part, gridsize, direction, dexmat)
Input: part geometric model
gridsize the resolution of encoding
direction in which the part is encoded
Output: dexmat the dexel model of part
box = the interior of a rectangle containing the projection of part onto the plane perpendicular to the direction (eg. y-axis).
image = {(i,j)| (i,j) ∈ ray matrix, (i*gridsize, j*gridsize) ∈ box}
for (each (i,j) ∈ image) do
{
(S1,…,S2n) = y-values of intersections of the part with the ray through (i*gridsize, j*gridsize) normal to x-z plane, in decreasing sequence
}
dexmat[i,j] = [(S_1, S_2), …, (S_{2n-1}, S_{2n})].

The part is encoded in two perpendicular directions (P and Q) that are perpendicular to build direction for each build orientation. We get two dexel lists,
$[(S_1^1, S_2^1), \ldots, (S_{2n-1}^1, S_{2n}^1)]$ and $[(S_1^2, S_2^2), \ldots, (S_{2n-1}^2, S_{2n}^2)]$. Then we browse the two lists and record the number of occurrence of

$$S_i^j - S_{i+1}^j < d_c, \qquad (5)$$

into an integer IA, where j =1, 2; i = 2k, $k \in (1, n\text{-}1)$; d_c is the cutter diameter. Then A_{ia} is calculated as:

$$A_{ia} = 2 * IA * gridsize^2. \qquad (6)$$

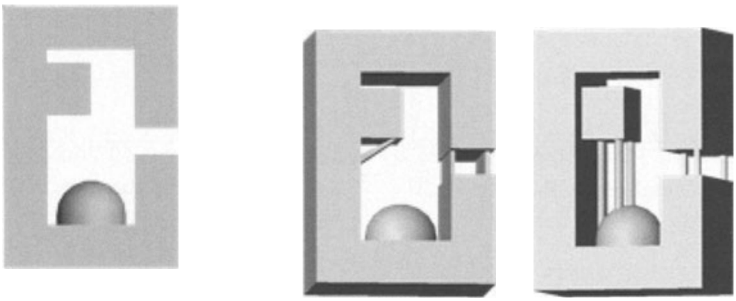
$F_{is} \in [0,1)$.

3.3 Support area

Because of the stiffness of raw material, not all the overhangs need support. As illustrated in figure 2(a), overhangs can be classified into two kinds according to their geometric property: *Complete-Overhang* and *Semi-Overhang* (cantilevered overhang). The former is *support-needed overhang*, while the latter can be divided further into *support-needed overhang* and *support-needless overhang*. Figure 2 (b) gives some possible support structures. Support structure for *Semi-Overhang* is aimed at preventing the part from deformation caused by gravity and cutting force.

Support area factor F_{sa} is defined as:

$$F_{sa} = 1 - Sum_{NS}/A_t\,, \tag{7}$$



(a) Model with overhangs (b) Possible support structures

Figure –2. Overhangs and possible support structures

where A_t is the total surface area of the part. Sum_{NS} has the physical meanings to indicate how much complete overhang area the sum of all the support-needed area is equivalent to. $F_{sa} \in [0,1]$.

3.4 Number of stock layer

Number of stock layer factor F_{sl} is defined as:

$$F_{sl} = 1 - h_{bo}/D\,, \tag{8}$$

where h_{bo} is the part height in build direction, D is the diameter of the part. $F_{sl} \in [0,1)$.

3.5 Removed material volume

Removed material volume factor F_{rm} is defined as:

$$F_{rm} = V_{mb}/V_{ab}\,, \tag{9}$$

where V_{mb} is the volume of the minimum bounding box of the part, V_{ab} is the volume of axis-parallel minimum bounding box (APB in brief) when the part is oriented. $F_{rm} \in [0,1)$.

3.6 Fuzzy representation of factors

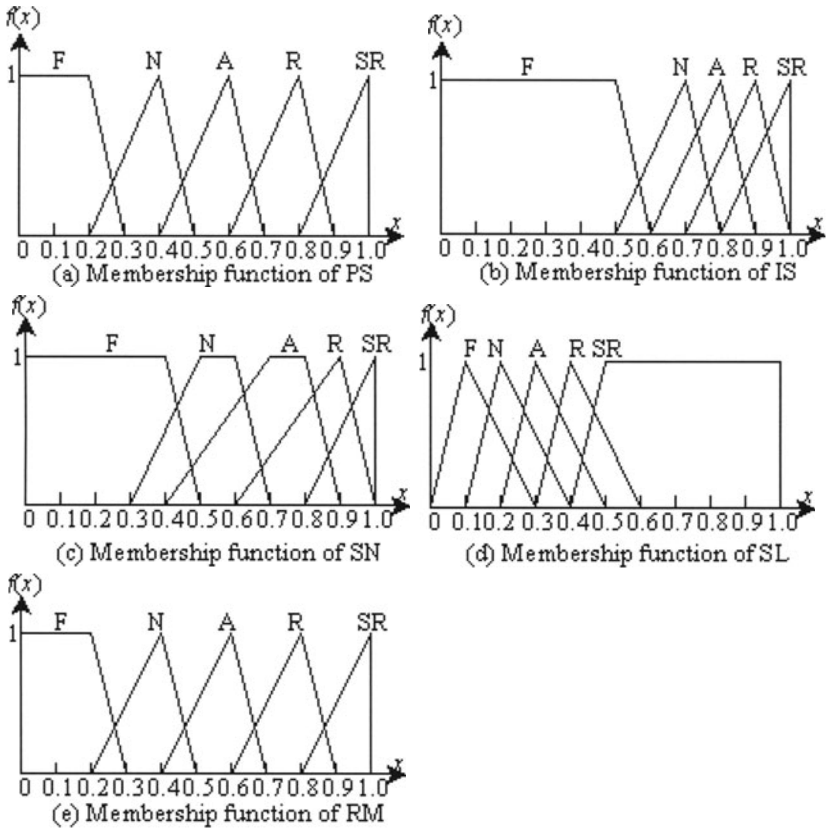


Figure -3. Membership functions of fuzzy variables

The five factors: Part stability (PS), Inaccessible surface area (IS), Support needed area (SN), Number of stock layer (SL), and Removed material volume (RM) are then evaluated as five linguistic categories: “Strongly recommended” (SR), “Recommended” (R), “Acceptable” (A), “Not recommended” (NR), and “Refused” (F). The membership functions of these evaluation categories are subjectively defined based on analysis and experiences. The graphical representation of the membership functions for each factor is shown in figure 3.

Then the mathematical equations of the membership functions can be given as following:

$$f(x) = \begin{cases} (x-a)/(b-a) & a \le x \le b \\ 1 & b \le x \le c \\ (x-d)/(c-d) & c \le x \le d \\ 0 & otherwise \end{cases}, \qquad (10)$$

Using these equations, the evaluation of a given factor can be represented as a limited fuzzy set. For example, if F_{ps} = 0.65, then F_{sl} is evaluated as the following fuzzy vector:

e_{sl} = {0, 0, 0.5,0.25,0}, which means the membership values of F_{ps} in "acceptable" and "recommended" are 0.5 and 0.25 respectively.

3.7 Fuzzy synthetic evaluation

Fuzzy relation matrix $\widetilde{R}_i$ for alternative a_i is then constructed by fuzzifying the factors:

$$\widetilde{R}_i = (e^i_{\mathrm{ps}},\ e^i_{\mathrm{is}},\ e^i_{\mathrm{sn}},\ e^i_{\mathrm{sl}},\ e^i_{\mathrm{rm}})^{\mathrm{T}}, \qquad (11)$$

where e^i is the single factor evaluation fuzzy vector.

then evaluation vector $\widetilde{V}_i$ for alternative i can be calculated:

$$\widetilde{V}_i = W \otimes \widetilde{R}_i = (v_1, v_2, \ldots, v_5), \qquad (12)$$

where W is given in (15), $\otimes$ denotes composition operation, $\widetilde{V}_i$ is the fuzzy vector containing the membership values for alternative a_i in each of the evaluation categories.

3.8 Fuzzy ranking

The preferred build orientation determination is made on the basis of fuzzy ranking. Suppose we have two evaluation fuzzy vectors, $\widetilde{V}_i$ and $\widetilde{V}_j$. We can use operations based on extension principle to calculate the truth value of the assertion that $\widetilde{V}_i$ is better than $\widetilde{V}_j$ with the following expression:

$$T(\widetilde{V}_i \geq \widetilde{V}_j) = \sup_{x \geq y} \min(\mu_{\widetilde{V}_i}(x), \mu_{\widetilde{V}_j}(x)), \qquad (13)$$

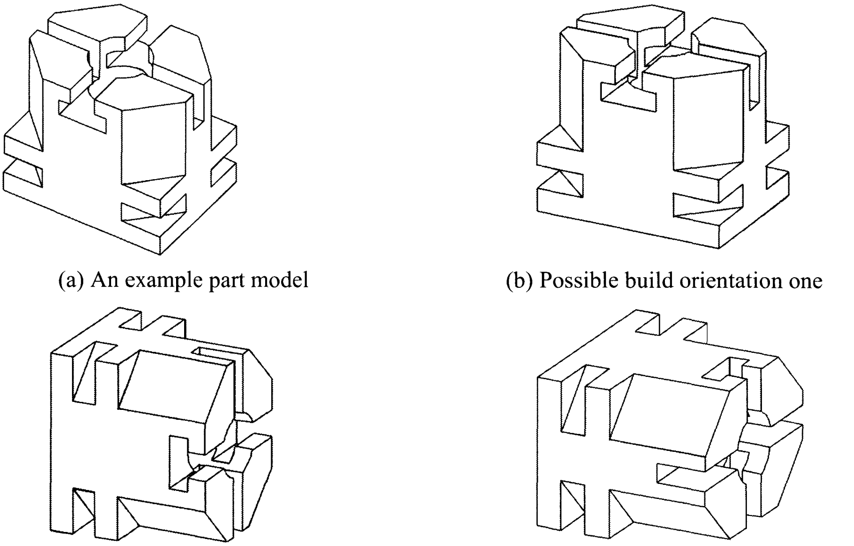
(13) is the extension of inequality $x \geq y$ according to extension principle. It represents the degree of possibility in the sense that if a specific pair (x, y) exists such that $x \geq y$ and $\mu_{\widetilde{V}_i}(x) = \mu_{\widetilde{V}_j}(y)$, then $T(\widetilde{V}_i \geq \widetilde{V}_j) = 1$.

The definition for two fuzzy sets can be extended to the more general case of multiple fuzzy sets:

$$T(\widetilde{V} \geq \widetilde{V}_1, \widetilde{V}_2, \ldots, \widetilde{V}_k) = T(\widetilde{V} \geq \widetilde{V}_1) \text{ and } T(\widetilde{V} \geq \widetilde{V}_2) \text{ and} \ldots \text{and } T(\widetilde{V} \geq \widetilde{V}_k) \qquad (14)$$

4. EXAMPLE

The part in figure 4 is used as an example to demonstrate the methodology presented in this paper. Three candidate build orientations are chosen as figure 4 (b), (c), and (d). The following are some of the parameters:



(a) An example part model

(b) Possible build orientation one

(c) Possible build orientation two

(d) Possible build orientation three

Figure –4. A part model and its candidate build orientations

$d_c = 10\text{mm}$; girdsize = 0.1mm;

$D = 567.8908$; $A = 140000$; $A_t = 896326.0$;

$A_{bp}{}^{(a)} = 120000$; $A_{bp}{}^{(b)} = 51000$; $A_{bp}{}^{(c)} = 82800$;

$h_{cg}{}^{(a)} = 126.7357$; $h_{cg}{}^{(b)} = 200$; $h_{cg}{}^{(c)} = 150$;

$A_{ia}{}^{(a)} = 0$; $A_{ia}{}^{(b)} = 29617.6$;$A_{ia}{}^{(c)} = 28592.8$;

$V_{mb} = V_{ab}{}^{(a)} = V_{ab}{}^{(b)} = V_{ab}{}^{(c)} = 42000000$;

Table –1. Fuzzy evaluation for each candidate build orientation

Candidates BO / Factors	BO(a)	BO(b)	BO(c)
F$_{ps}$	0.8781	0.6684	0.7758
F$_{is}$	10.967	0.0000	0.9681
F$_{sa}$	0.9027	0.8136	0.8593
F$_{sl}$	0.3837	0.2956	0.4717
F$_{rm}$	1.0	1.0	1.0

The five factors for each candidate are calculated based on these data and the results are tabulated in Table 1.

Then we obtain the single-factor evaluation matrices for each candidate by fuzzifying the factors as follows:

$$R_a = \begin{array}{c} \begin{array}{ccccc} F & NR & A & R & SR \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0.219 & 0.3905 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0.937 & 0.5135 \\ 0 & 0.0815 & 0.5815 & 0.837 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ PS \\ IS \\ SA \\ SL \\ RM \end{array}, \qquad (15.\ a)$$

$$R_b = \begin{array}{c} \begin{array}{ccccc} F & NR & A & R & SR \end{array} \\ \begin{pmatrix} 0 & 0 & 0.316 & 0.342 & 0 \\ 0 & 0 & 0 & 0.33 & 0.835 \\ 0 & 0 & 0.864 & 0.712 & 0.068 \\ 0.22 & 0.5325 & 0.953 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ PS \\ IS \\ SA \\ SL \\ RM \end{array}, \qquad (15.\ b)$$

$$R_c = \begin{array}{c} \begin{array}{ccccc} F & NR & A & R & SR \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0.879 & 0 \\ 0 & 0 & 0 & 0.319 & 0.8405 \\ 0 & 0 & 0.407 & 0.8643 & 0.2965 \\ 0 & 0 & 0.0943 & 0.6415 & 0.717 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ PS \\ IS \\ SA \\ SL \\ RM \end{array} \qquad (15.\ c)$$

Then we put the matrix above into (12) and get the following synthesis evaluation vectors for each candidate:

$V_a = (0.0651, 0.0651, 0.0651, 0.1102, 0.1049)$,

$V_b = (0.0651, 0.0651, 0.1102, 0.1049, 0.0651)$,

$V_c = (0.0651, 0.0651, 0.0651, 0.1102, 0.0651)$,

According to maximum membership rule, build orientation *a* and *c* are evaluated as "Recommended", while *b* as "Acceptable".

From (13) we get,

$$T(\widetilde{V}_a \geq \widetilde{V}_c) = 0.1102,\ T(\widetilde{V}_c \geq \widetilde{V}_a) = 0.0651,$$

Therefore, build orientation *a* is preferred over *c*.

5. CONCLUSION

A methodology employing fuzzy set theory to select the preferred build orientation in layer based machining system has been developed. The benefits of experience were brought into a decision-making procedure by using fuzzy set to represent the factors that influence the selection of build orientation. The implementation of this methodology is a functional module of our Robot-based Layered Manufacturing system RoLM, which is an OpenGL based software package developed for adaptive layer based robot machining.

ACKNOWLEDGEMENTS

This project is supported by a grant from Hong Kong Grant Council with account number HKU567/96E.

REFERENCES

1. Lan, P.T., Chou, S.Y., Chen, L.L., and Gemmill, D., 1997, Determining fabrication orientation for rapid prototyping with stereolithography apparatus. *Computer-Aided Design*, **29**(1), 53-62.
2. Xu, F., Wong, Y.S., Loh, H.T., Fuh, J.Y.H, and Miyazawa T., 1997, Optimal orientation with variable slicing in stereolithography. *Rapid Prototyping Journal*, **3**(3), 76-88.
3. Alexander, P., Allen, S., and Dutta, D., 1998, Part orientation and build cost determination in layered manufacturing. *Computer-Aided Design*, **30**(5), 343-356.
4. Pham, D.T, Dimov, S.S., and Gault, R.S., 1999, Part orientation in Stereolithography. *The International Journal of Advanced Manufacturing Technology*, **15**, 674-682.
5. Masood, S.H., Rattanawong, W., and Iovenitti, P., 2000, Part build orientation based on volumetric error in fused deposition modelling. *The International Journal of Advanced Manufacturing Technology*, **16**, 162-168.
6. Horváth, I., Vergeest, J.S.M., Broek, J.J., Rusák, Z., and Smit, B., 1998, Tool profile and tool path calculation for free-form thick-layered fabrication. *Computer-Aided Design*, **30**(14), 1097-1110.
7. Chen, Y.H., and Song, Y., The development of a layer-based machining system. *Computer-Aided* Design, **33**(4), pp. 331-342.
8. Chen, Y.H., Wang, Y.Z., and Leung, T.M., 2000, An investigation of parting direction based on dexel model and fuzzy decision making. *International Journal of Production Research*, **38**(6), 1357-1376.
9. Güngör, Z., and Arikan, F., 2000, Application of fuzzy decision making in part-machine grouping. *International Journal of Production Economics*, **63**, 181-193.
10. Ross, T.J., 1995, *Fuzzy Logic with Engineering Applications* (New York: McGraw-Hill).

CONTRIBUTION TO THE GENERATION OF TOOL PATHS IN A CAM SYSTEM

A. Larue and B. Anselmetti
LURPA, ENS de Cachan, France

Abstract: The flank milling of complex forms is a very effective process from the point of view of productivity and surface quality. Many works deal with research on the optimal positioning of the tool which is considered as a rigid body in order to minimize tool path errors. The purpose of our work is to integrate the compensation of the tool distortions in this optimal positioning calculation. In flank milling with long tools, the distortion of the cutter generates a significant wave (that can reach 0,6mm) on the machined surface due to the effects of the helical angle and the radial force which varies during the cutter rotation. After detailing an analysis of the force evolution and the associated model calculation, we will present a test protocol, that can be implemented in industry, in order to characterize the model parameters as a function of the couple tool-workpiece material. Then we will present a test to assess our prediction model of the straightness defects of the machined surface according to all machining parameters. These results make it possible to make up for defects by applying a translation to the tool in 3-axis and by applying a translation combined with a rotation in 5-axis milling.

Key words: Identification, Force model, Compensation, Tool paths

1. INTRODUCTION

The flank machining of complex forms is currently a strong industrial issue. The objective is to machine a high quality surface while minimizing manufacturing times. A lot of the research deals with the generation of an

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 221–230.

optimal positioning of the tool which is considered as a rigid body in order to minimize tool paths errors [1], [2], [3]. These approaches ignore the cutting process's intrinsic elements, in particular the tool distortions which can be significant (about 0.6 mm for a 20 mm diameter long cutter and a raised material flow in medium-hard steel). Our objective is to take into account the tool deformations in tool path calculations for flank milling of free forms. This article qualitatively analyses the flank milling process and proposes a method to assess the cutting forces and the tool distortions. A test protocol, that can be implemented in industry is proposed to characterize the various model parameters as a function of a given couple tool-workpiece material. The defects are then compensated by a 3 axis tool translation.

2. ANALYSIS OF THE CUTTING FORCE

In order to highlight the tool distortion, a plane was machined with a high speed steel cutter of 20 mm diameter and 90 mm of length, with a progressive radial engagement from 0.5 to 3 mm and a constant feedrate per tooth fz = 0.2 mm/rev. For example, for a radial engagement of 3 mm, the distortion generates a variation of localization of 0.6 mm with a wave giving a variation of straightness of 0.2 mm (figure 1).

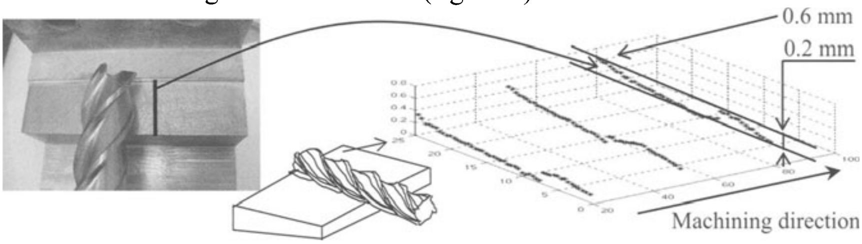


Figure -1. Evaluation of the machining deviation amplitude.

To model the tool distortion, it is necessary to understand the evolution of the cutting force. Figure 2 shows the cutting process :

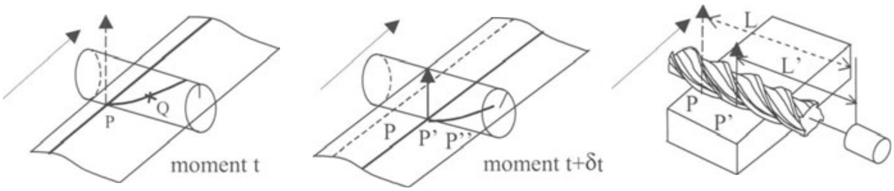


Figure -2. Displacement of the generating point.

The cutting edge of the milling tool is a helix. The surface is thus obtained by a generating point which moves along the milling tool generatrix during the cutter rotation. At moment t, the generating point is at P at a distance L from the housing. At moment t+δt, the point is at P' at a distance

L'. An elementary cutting edge placed at point Q which is inside the material (figure 3) cuts an elementary chip section which depends on the engagement angle at this point. An elementary cutting pressure is thus applied at this point. This force generates an elementary inflection of the cutter. The distortion at point P is thus the sum of the elementary distortions for the whole length in mesh of the cutting edges simultaneously engaged in the workpiece material.

The contact surface between the cutter and the workpiece material was developed in figure 3 which represents the cutter in three angular positions. The evolution of the chip section along the cutting edge, and the normal density of the force for each point Q are identified according to the angular position β at this point. α is the total engagement angle of the cutter in the workpiece material.

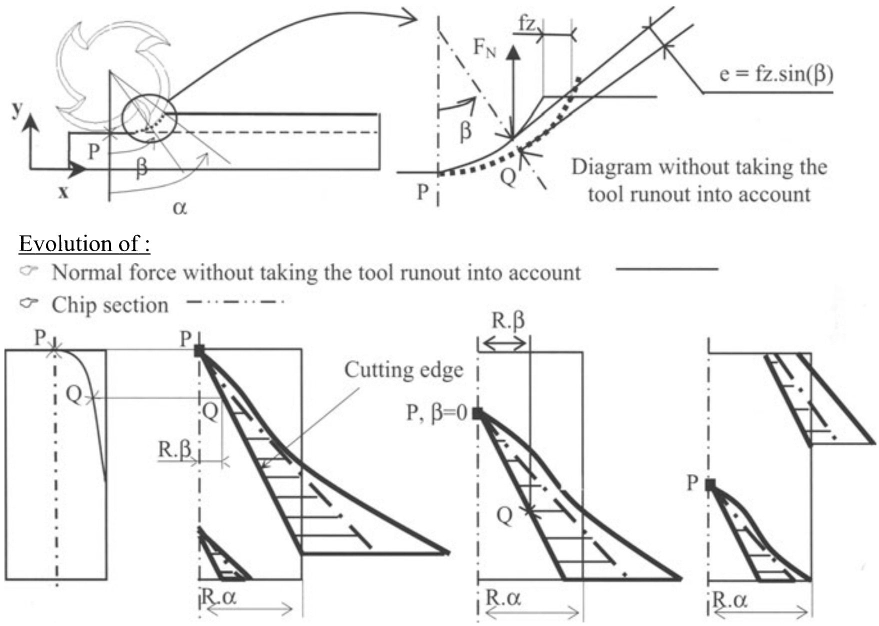


Figure -3. Radial Force evolution on helical developped profile.

During rotation, distortion is variable, in particular because of the distances from the generating point P and points Q to the housing variation. Moreover, there can be one or 2 teeth completely engaged. This variation of distortion as a function of point P sometimes creates a significant wave in a machined surface section.

3. DISTORTION MODEL OF THE MILLING TOOL

There are numerous works concerning the static force model study [4], [5], [6]. Our modeling of the forces makes use of the basic principles of the Kline and DeVor model [7], [5], [8]. This approach allows for the prediction of the cutting force and is based on a static equilibrium of the tool fitted in the spindle cone.

3.1 Influence of the various factors on distortion

We suppose that the defects observed are mainly due to the cutter distortion. This assumption was validated by checking. The machined parts are rather massive and firmly clamped to be considered as a rigid body. The fixture is also very rigid (vice or table fixture). The spindle was subjected to a static force equivalent to the machining force [8] : the distance variation between the spindle and the fixture (~ 0.01 mm) is is very low in relation to the observed defects. Similarly, the static distortion of the tool holder is also negligible. The distortions of components other than the milling tool will thus be ignored. The cutter will be considered as a beam fixed in the milling tool holder. A test also showed that the flexion was not deviated by the helical shape of the cutting edges.

3.2 Modeling of the distortions

The method that we propose differs from the Kline and DeVor approach as re-envisaged by Seo [6] : Our calculation estimates the generating point of the cutting edge distortion by adding up the distortions due to the elementary forces distributed along the cutting edge.

The model illustrated by figure 4 determines the range of deformation due to normal forces on the machined surface for each elementary length of the cutting edges in mesh.

The cutting pressure is assumed to follow the model (eq.(1) ; eq.(2)) :

$\Rightarrow$ Radial force : $Kr = Kr_o .\ ep^{-0.3}$ (ep = depth of cut) (1)

$\Rightarrow$ Tangential force : $Kt = Kt_o . ep^{-0.3}$ (2)

Coefficients Kr_o and Kt_o depend in particular on the tool angles and the machined material. The exponent –0.3 is a term which is considered as constant for a wide material range. After the tool is broken down into elementary discs whose thickness is ea, the two components dFt and dFr of the elementary force can be calculated according to the angular position β at point Q considered (eq. (3) – eq. (4)).

$$\Rightarrow \text{Radial force}: \; dFr = Kr.ep.ea = Kr_{o.}\,(fz.\sin(\beta))^{1-0.3}.ea \tag{3}$$

$$\Rightarrow \text{Tangential force}: \; dFt = Kt.ep.ea = Kt_{o\,.}\,(fz.\sin(\beta))^{1-0.3}.ea \tag{4}$$

The normal component to the machined surface dN can be calculated :

$$dN(\beta)= dFr(\beta).\cos(\beta) + dFt(\beta).\sin(\beta) \tag{5}$$

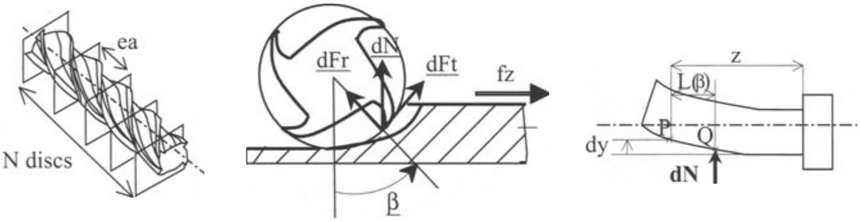


Figure -4. Details of the model used.

The distortion dy at point P due to force dN at point Q can be calculated by considering that the tool is a fixed beam of constant Igy moment of inertia :

$$dy = \frac{1}{E.Igy}.dN(\beta).\left[\left(\frac{(z-L(\beta))^3}{3}\right).\left(\frac{(z-L(\beta))^2.L(\beta)}{2}\right)\right] \tag{6}$$

The radial engagement being small, there is only one tooth in mesh on each disc. At each point P, the theoretical y_{th} is the sum of dy generated by all points Q of the cutting edges in mesh [8].

4. COUPLE TOOL-WORKPIECE CHARACTERIZATION

4.1 Description of the test protocol

The force model requires the identification of Kr_o, Kt_o, Igy et E. The term E.Igy is identified thanks to a static test of the milling tool using a dynamometer [8].

The test is based on the machining of a plane. The part is measured on a three-dimensional coordinate measuring machine. A simulation model developed under Matlab software calculates the tool distortions for each

point P on the surface by adjusting the parameters Kr_o, Kt_o to reproduce the measured form as accurately as possible.
The test thus consists in flank machining a plane starting from a tilted raw plane of a known value (figure 5) giving a radial engagement ranging from 0.5 to 3mm. The milling tool engagement angle varies throughout the facing. On the two raw part sides, the two zones for which the radial depth of cut is 0.3 mm make it possible to easily readjust the measured dots cloud. The pure flank milling without end-milling is obtained by a groove in which the tool tip emerges. The calibration stage of the force model coefficients is then not disturbed by a parasitic friction of the tool point during the machining.

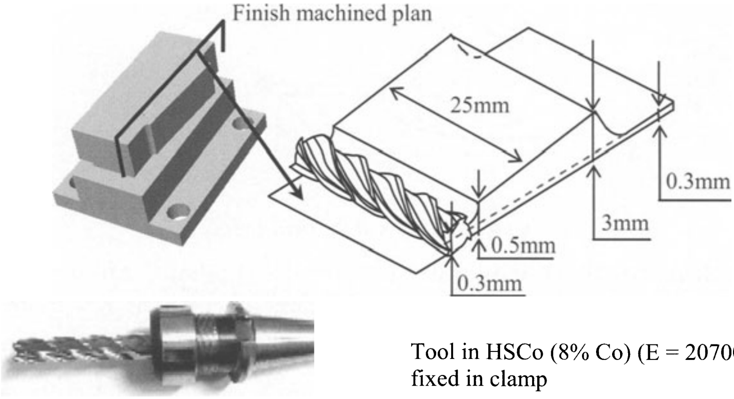


Figure -5. Test of tool – workpiece characterization.

Measurement is carried out in a rough-planing reference – mark built on the surfaces made with the tool to test and of the cutting conditions which must give negligible distortions of the tool (figure 5).

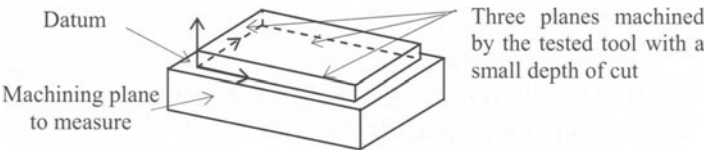


Figure -6. Rough-planing reference - datum for the machined surface measurement.

The measurement, after processing in the calculation model, gives the difference between the points on the machined surface and the theoretical surface built relative to the reference-mark of rough-planing (figure 6). The model coefficients are obtained by minimizing the sum of the squared differences between the measured dot cloud and the simulated dot cloud [8].
Figure 7a represents the simulated dots cloud (deflection ranging from 0 to 0.7 mm). Figure 7b shows the distribution of the differences between the

points measured and the points simulated (±0.06 mm). The most significant variations are on the part zone where radial engagement is largest.

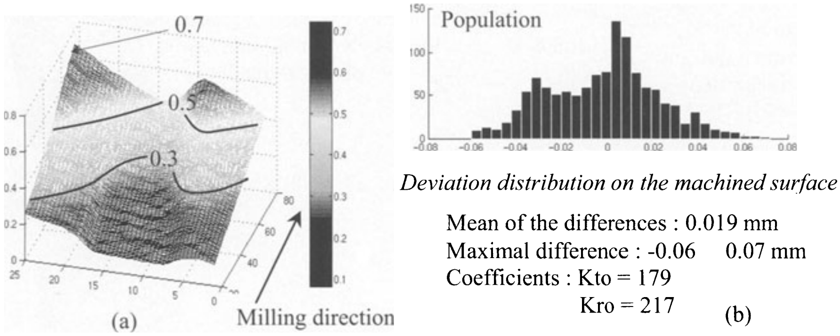


Figure -7. Numerical results for the HSS tool (tool diameter : 20mm, tool length : 94mm).

Figure 8 shows the results obtained with a shorter carbide tool.

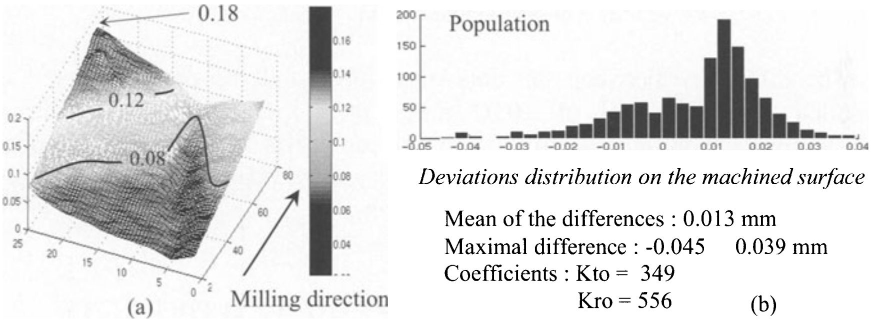


Figure -8. Numerical results for the carbide tool (tool diameter : 20mm, tool length : 54mm).

This test protocol allows for the estimation of the force model coefficients and for the prediction of the machining defects with a sufficient precision next to the amplitude of the measured distortions. Figure 8 shows that most of the variations for the robust carbide tool are included in an interval of ±0.03 mm, which is acceptable as regards our precision target.

4.2 Validation procedure of the prediction defect

The test consists in realising a circular interpolation. In this case, for a given radial engagement, the engagement angle is smaller than for the plane. The measurement consists in identifying several generators between the altitudes z = 0 and z = 30.

The tests were performed for the two tools. The curves in figures 9 and 10 show the quality obtained by this simulation.

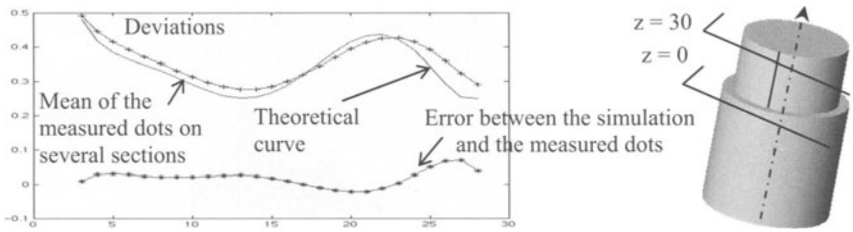


Figure -9. Validation of the identification procedure for a high speed steel tool.

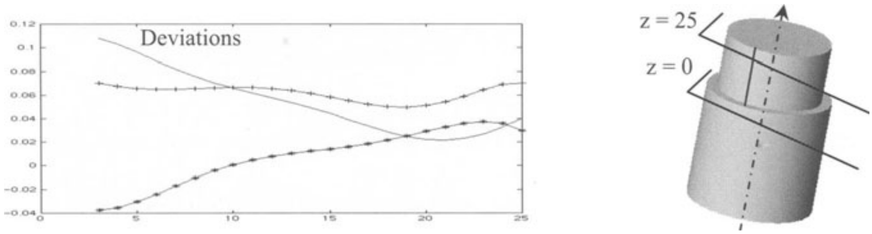


Figure -10. Validation of the identification procedure for a carbide tool.

The difference between the dots measured and the dots simulated is included in an interval of ±0.02 mm over the first 20 millimetres. Considering the initial objectives, our modeling is validated. The model must however still be refined, because the top cylinder is still rather badly evaluated (approximately 0.05 mm deviation).

5. COMPENSATION OF PREDICTED DEFECTS

The mirror technique [6], [9] allows to partially make up for the defects generated by the working tool deflection. In order to achieve this, in each section, an average value of the distortion was calculated to generate a tool translation following the normal to the surface. The milling test of a plane starting from the rough surface used for characterization of the force model coefficients was realised by calculating the compensated tool path as shown in figure 11. After computation of the mean deformation to compensate in each section, the defects are interpolated with a non uniform cubic Bspline curve, which is then discretized to generate toolpath. The test results are illustrated in figure 12.

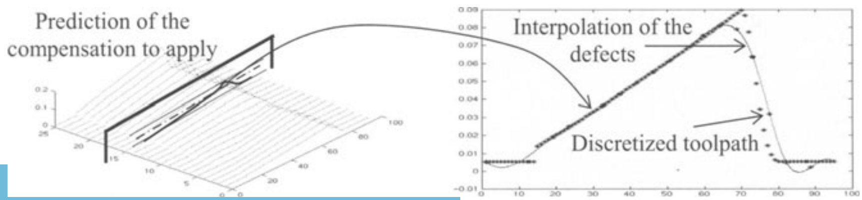
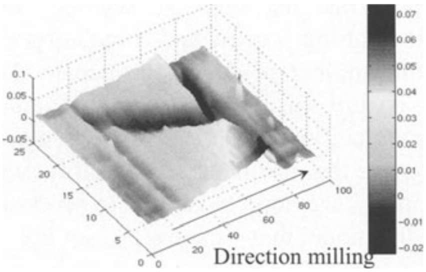
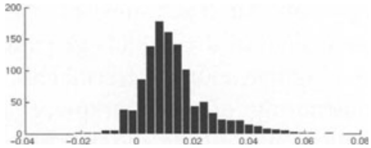


Figure -11. Generation of the compensated tool path.

Difference between the expected plan and the real measured point

Difference distribution





Difference e to 95% (with compensation) :
0.019 mm < e < 0.172 mm
Difference to 95% (with compensation) :
-0.004 mm < e < 0.046 mm

Figure -12. Differences between the measured points and the theoretical plane (carbide tool).

The position defect after compensation strongly decreased compared to normal milling. We observed that the 95% difference between the measured points and the theoretical plane is really smaller when we compensate. 77% of the surface points after compensation respect an interval of ±0.01 mm, but the whole surface is tilted because of the tool deflection. It now seems difficult to improve the model without going over to 5 axes in order to compensate for the tool paths simultaneously by translation and rotation. Moreover, in this test, the transition zones have not been taken into account in the simulation. To improve this result, we will soon set up a 5-axis compensation and will manage the transition calculation at the depth of cut changes. The integration of our algorithm in a CAM system requires a tool engagement map definition correlated with a part's feedrate map which takes into account the machine tool dynamic. These maps allow to estimate the surface deviation. In 3-axis milling, we define a mean profile to calculate tool paths and locally control the feedrate. In 5-axis, we define tool paths with positioning of two points of a tool axis on the mirror surface. A distortion of this surface can be used to modify the part feedrate when its distribution is not continuous.

6. CONCLUSIONS AND PROSPECTS

Machining tool path generation is rather difficult in many ways. Three great classes of problem appear : the tool optimal laying on the machined surface, the generation of a numerical control program taking into account the machine controller's real capacities and the generation of a numerical control program taking into account the distortion aspects generated by the cutting process. We have specifically worked on the latter. We have shown that the distortion aspect due to the mechanical stresses during machining is particularly disadvantageous. No current CAM system has integrated this type of problem. To contribute to such an integration, we set up and evaluate a fast step to qualify a couple tool – workpiece material in order to predict the distortions undergone by a part during its machining. The method is based on the machining of a test plane. This first stage allows accurate assessment of the defects generated by the cutting process and calculation of the compensation. Nevertheless, a certain number of defects related to the non-rigidity of the workpiece and its fixture, also related to vibrational phenomena and the shock effects on attack and on exit of the cutter remain to be taken into account. In order to improve the machined parts quality, we set up a 3-axis method compensation. Finally, we hope to be able to propose a 5-axis compensation method based on the model that we have just set up.

REFERENCES

1. A. AFFOUARD. "Introduction de l'interpolation polynomiale dans le cadre de l'usinage à 5 axes par le flanc", DEA de production automatisée LURPA ENS CACHAN, 2001.
2. F. MONIES, J.M. REDONNET, W. RUBIO, P. LAGARRIGUE. *"Improved Positioning of a Conical Mill for Machining Ruled Surfaces : Application to Turbine Blades"*, J. of Engineering Manufacture, Vol. 214, N°7, 2000, pp. 625-634.
3. TÖNSHOFF. *"Optimal Tool Positioning for Five-Axis Flank Milling of Arbitrary Shaped Surface*s", Production Engineering, Vol. 7, N°1, 2000.
4. F. ABRARI, M. A. ELBESTAWI. *"Closed form formulation of cutting forces for ball and flat end mills"*, Int. J. of Machining Tool Design Research, Vol. 37.1, 1997, pp. 17-27.
5. W. A. KLINE, R. E DEVOR, J. R. LINDBERG. *"The prediction of cutting forces in end milling with application to cornerong cuts"*, Int. J. of Machining Tool Design Research, Vol. 22.1, 1982, 7-2.
6. T. SEO. *"Intégration des effets de déformation d'outil en génération de trajectoires d'usinage"* Thèse de Doctorat, Ecole Centrale de Nantes, 1998.
7. R. E. DEVOR, W. A. KLINE, W. J. ZDEBLICK. *"A Mechanistic Model for the Force System in End Milling with Application to Machining Airframe Structures"*, Proceedings of eighth North11American Manufacturing Research Conference : USA, Vol. 8, 1980, pp. 297-303.
8. A. LARUE, B. ANSELMETTI, B. SOULIER. *"Influence des déformations des fraises sur la qualité des surfaces usinées",* 7ème Colloque PRIMECA sur la Conception Mécanique Intégrée, La Plagne, 2000, pp. 334-341.
9. J. Y. HASCOET, J. J. LEE, A. DUGAS. *"Machining simulator with dynamic error analysis on free 5urface machining"*, third. Int. Conference on Integrated Design and Manufacturing in Mechanical Engineering, Montréal, Canada, 2000.

A TOOL-WEAR INVESTIGATION IN HARD TURNING OF VARIOUS WORK MATERIALS WITH CBN TOOLS

G. Poulachon[1], P. Jacquet[1], S. Pheulpin[1], E. Seguin[1] and I.S. Jawahir[2].

1 ENSAM, LaBoMaP, Cluny, France
2 University of Kentucky, Lexington, USA

Abstract: An extensive study has been undertaken to investigate the tool-wear mechanisms of CBN cutting tools in finish machining of the following hardened steels: X155CrMoV 12 cold work steel (AISI D2), X38CrMoV5 (AISI H11) hot work steel, 35NiCrMo16 hot work steel and 100Cr6 bearing steel (AISI 52100), treated at 54 HRC. A large variation in tool-wear rate has been observed in machining of these steels. The generated tool flank grooves have been correlated with the hard carbide density of the workpieces. A crater wear study has also been performed and, it is shown that the appearance of an adhered third body could induce a chemical wear in the tool.

Key words: Hard turning, carbides, microstructure, tool-wear

1. INTRODUCTION

Because of the increased productivity and flexibility required in manufacturing, machining hardened steel with a cutting tool instead of grinding is much more attractive. It is possible to decrease the machining cost by using hard turning instead of grinding because a turning center is less expensive than a grinding machine. Also, machining operations can produce many different features with a single cutting tool as opposed to the need for changing grinding wheels in grinding. The only drawback of the turning

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 231–240.

method is the high cost of the CBN turning insert. According to König et al. [1] the tool-life varies from 1 to 25 times for different hardened steels treated under the same hardness and machined at the same cutting parameters. Therefore, linking the tool-life with the cutting material microstructure is significant in hard turning process.

2. STATE OF THE ART

König et al. [1] and Narutaki and Yamane [2] examined the wear mechanisms by diffusion, abrasion, and adhesion. All these mechanisms are prevalent during the wear process of the tools. According to previous research [1], [3-6], there is a chemical and diffusion process which influences the tool-wear process. This diffusion wear is often linked with the built up layer (BUL). The existence of a BUL on the rake face and sometimes on the clearance face has been shown by [3-6]. The BUL can have many implications on tool-wear. The first is reduction of tool-wear according to Luo et al. [3] because it is like a coating on the tool face that prevents the tool from abrasion due to hard carbides in the workpiece. Barry et al. [4] used three different workpieces with the same hardness and different rates of inclusions. After extensive analysis of the BUL, the authors show that an increase in aluminum content leads to an increase in crater and flank wear.

It is also known that annealed alloyed steel contains many different types of carbides. It can be a factor to explain the increase in tool-wear when there is a decrease in hardness. The flank groove width increases with the increase in workpiece hardness. The amount of carbides increases with the decrease in hardness, due to high tempering temperature. The increase in groove width can be explained for this type of steel by the fact that for high workpiece hardness there are only massive alloy carbides and consequently the groove width is closer to its average diameter. For lower hardness, there is also the formation of small hard carbides due to tempering. Grooves are formed by abrasion as well as precipitated alloy carbide particles. Ohtani and Yokogawa [8] show this behavior and also highlight the fact that tool-life increases with the increase in hardness (in the hardness range lower than HRC 50).

The influence of adhesion can be observed especially on BN grains [3-4], [8]. Indeed, with diffusion, oxidation, or chemical reaction, the grains can be removed more easy because the boundaries between the binder and BN are broken. The adhesion on BN grains followed by abrasion with high carbides from the workpiece can tear BN grains which are not completely dissolved by diffusion or chemical reaction and leave a ridge. Thus, the tool-wear increases with the removal of massive grains. This action can explain the lower tool-life of tool insert with high amount of CBN or with higher BN grain size. As a result, adhesion wear is a major wear parameter and it influences the wear of the ceramic binder with a well known mechanism of adhesion and abrasion, but also wear of BN grains with a mechanism of adhesion and wrenching of full grains.

3. EXPERIMENTAL SET-UP

3.1. Machine tool and workpiece description

The experiments were performed on a Mazak CNC turning center with 20 kW power. The tool geometry is as follows: cutting edge angle, $K_r = 91^{o;}$ rake angle; γ_o = -6°, inclination angle; λ_s = - 6°. The machining operation performed is tube turning. The samples are tubes of 110 mm outer diameter, 200 mm length and 20 mm thick. The tube thickness has been chosen to keep the microstructure in the whole sample uniform. The tubes are clamped between soft jaws and a tail stock to ensure stability in the machining operation.

3.2. Cutting parameters

The cutting parameters are chosen to be as close as possible to the industrial cutting conditions. The hard turning operation is generally used as a finishing operation, as the alternative to grinding.

A **depth of cut (a_p)** of 0.2 mm has been selected to enable the tool-chip contact only on the chamfer.

Two **feed rates (*f*)** 0.08 and 0.12 mm/rev have been selected because they can produce a good surface roughness for a 0.8 mm radius tool tip.

Two **cutting speeds (V_c)** have been selected; 230 and 180 m/min.

This cutting parameter is the most dominant on the tool-life in comparison with the two previous parameters and it is strongly correlated with the cutting temperature. A new model developed by Poulachon [9] shows that the optimal cutting speed for machining of 100Cr6 bearing steel treated at 54 HRC was around 120 m/min. Therefore, this cutting parameter has been increased in order to reduce the cutting time in the experiment.

4. ANALYSIS

4.1. Flank wear study

It is generally known that in machining, the flank wear is the most common criterion. Indeed, the value of the cutting edge retreat has an influence on the workpiece dimensional accuracy. As a consequence, this criteria has been the first used to measure the tool-wear rate and also to classify the different tool-wear rates according to the types of steels machined and the cutting conditions used in the experiments. Nevertheless, the tool-wear is a loss of the cutting material along the cutting edge, therefore, this criterion has been chosen to measure the tool damage. Thus, the flank wear *VB* value is only a geometric dimension and it does not show the tool-wear rate, but it is possible to obtain the volume of tool material removed as a function of *VB* and some geometric values of the tool as shown in Figure 1. Barry et al. [4] used the same method. The value *S* seems to be more reliable to show the progression of the tool material volume removed. By integration of *S* along the cutting edge, it is possible to obtain the volume *V* of the material removed. After the calculation of *V* and *S,* another parameter has been designed; the *length* of the workpiece rubbing is given by the following formula: *length = machined surface/feed rate*.

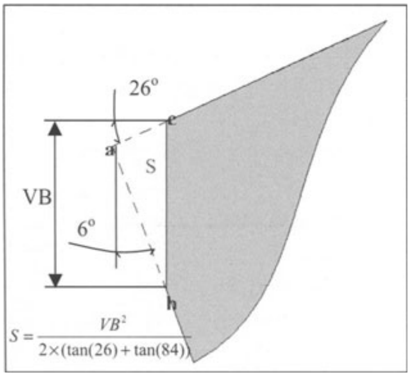


Figure -1. The calculation of the material removed from the tool due to tool-wear.

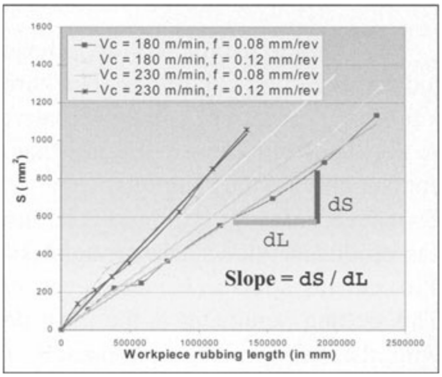


Figure -2. Tool material removal rate during the machining of X38CrMoV 5 steel.

The integration of *length* along the cutting edge gives the surface of the workpiece that has rubbed the flank surface of the tool. Locally, it is possible to draw the graph *S* as a function of the *length* as shown in Figure 2. The curves obtained are straight lines that cross the origin. Barry et al. [4]

obtained a similar shape of curves. Thus, each flank wear can be defined by one coefficient that is the straight-line slope "A_i".

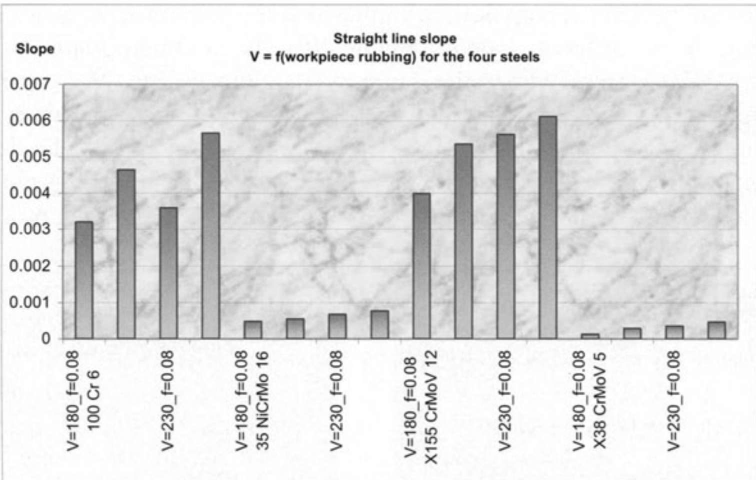


Figure -3. Comparison of tool-wear rates in machining of four different hard steels with CBN tool.

Figure 3 shows the straight-line slopes for the four steels obtained from the tool-wear experiments. From the results shown in Figure 3, the steels can be classified into two different groups according to the straight-line slopes. The first group induces large tool-wear rates. This group has two work materials; 100Cr6 bearing steel and the X155CrMoV12 steel. The second group is the group of steels with lower tool-wear rates. This group has two work materials: 35NiCrMo16 steel and the X38CrMoV5 steel.

5. ANALYSIS OF STEELS THAT INDUCE LARGE TOOL-WEAR RATES

5.1. Flank wear shape for the X155CrMoV12 steel

The surface of the tool face shows many grooves that can be observed with the optical microscope as shown in Figures 4(a). The grooves are formed in the cutting speed direction. Those grooves seem to be the result of extensive abrasive wear. The grooves on the flank surface appear at the beginning of machining and they never disappear. The global depth of these grooves is constant; it is not a function of the flank wear *VB* as shown in Figure 4(a). The frequency of the grooves along the cutting edge is measurable. It is of interest to see that there are two different types of grooves. One type is a large groove with an average size close to 10µm. The other one is a small type with an average of 1 µm. It also shows that the two types of grooves are superposed; it implies that the phenomenon is not local, and has two different roots. Figure 4(b) is a micrograph of the X155CrMoV12 steel where the large white clusters are M_7C_3 primary carbides, with an average size close to 15 µm. The major groove type is close to the big carbide cluster size. The minor grooves seem to be dug by the small clusters, which are about ten times smaller than the big clusters. There are M_7C_3 secondary carbides appearing during the tempering process.

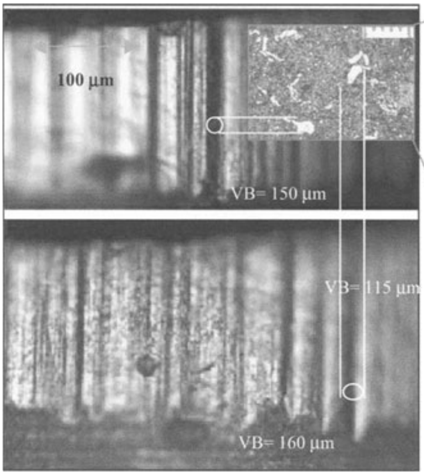


Figure-4(a). Flank wear in machining of X155CrMoV12 steel (cutting parameters: Vc = 180 m/min, f = 0.08 mm/rev).

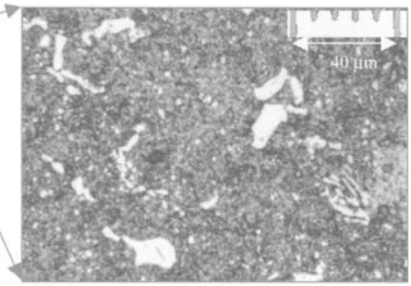


Figure-4(b). Microstructure of the X155CrMoV 12 steel, HRC 54.

At this stage of the study, the only thing that can be affirmed is that the grooves with the major groove sizes are made by high M_7C_3 carbide clusters. Some groove cannot be completely focused with the optical magnification. It signifies that the depth of those grooves is larger than the depth of focus 5 µm.

6. ANALYSIS OF STEELS THAT INDUCE A LOWER TOOL-WEAR RATE

6.1. Flank wear pattern for the X38CrMoV5 steel

This steel is the lowest abrasive according to the flank wear criteria. The structure is martensitic with cementite and a small amount of MC carbide type appears after the tempering. The size and the amount of carbides in the structure cannot be measured. The wear of the tool seems to be more linked to the martensitic matrix abrasive action. The pattern of the tool-wear in Figure 5(a) shows that the tool flank has several grooves. The average of the groove size is close to 25 µm. However, even when the grooves seem to have a size larger than the X155CrMoV12 grooves, the depth seems to be smaller. The grooves seem to be dug by the grains of the workpiece matrix that have a size close to 25 µm as shown in Figure 5(b). The profile of the grooves shown in Figures 5(a) has not been measured, but it can be affirmed that the dark lines on the flank wear is because the light is not well reflected. Moreover, the depth of focus is close to 5 µm and the grooves are completely focused and they induce the depth of the grooves smaller than 5 µm. Also, a third layer appears on the worn flank face. This layer seems to be homogenous like the one seen with the X155CrMoV12 steel.

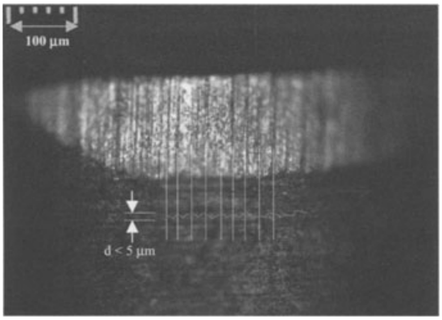


Figure-5(a). Flank wear for the X38CrMoV5 steel (cutting parameters: Vc = 180 m/min, f = 0.08 mm/rev).

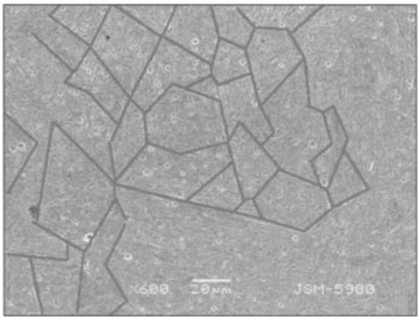


Figure-5(b). 35NiCrMo16 steel microstructure with HRC 54.

6.2. Flank wear shape for the 35NiCrMo16 steel

A preliminary study shows that for the fourth set of experiments, the pattern of the flank wear *VB* is similar, but the flank wear value *VB* is different. Its structure is martensitic without any carbides. Consequently, the tool-wear is fully linked to the martensitic matrix abrasive action and the anticipated chemical wear takes place.

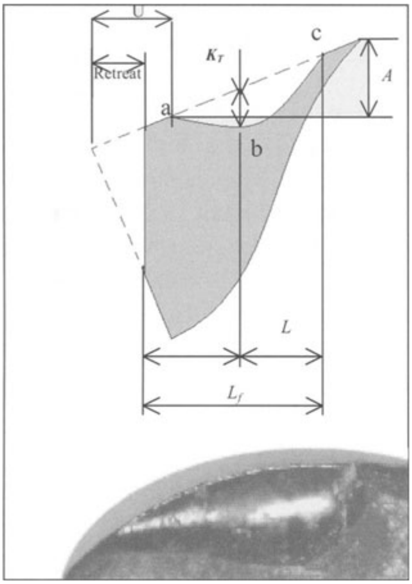


Figure-6. Cross section of a tool along the orange line.

To conclude, the X155CrMoV12 and 100Cr6 bearing steels are shown as the more abrasive steels according to Figure 3, because they contain the largest and hardest carbides. The flank wear observed seems to result mainly from the abrasive wear of the two types of carbides. The flank wear due to fatigue of the tool material can also be retained as the origin of tool-wear. The third layer, seen as the layer in gold color, can induce another wear behavior like diffusion-abrasion, or adhesion-abrasion. In the first case, the third layer reacts with the tool material. Barry et al. [4] show that the diffusion between BN grains and Aluminum element can appear. Another explanation has been given by König et al. [10] where they show the re-crystallization of the binder in a new structure that leaves the CBN grain less tied up. The second behavior, adhesion-abrasion is only a mechanical wear behavior. In each case, the material deposed on the flank wear is expected to come from the matrix of the workpiece, which is more ductile than the carbides. The 35NiCrMo16 and X38CrMoV5 steels classified in the second group come from a matrix composed with the same elements. It is mainly martensitic grains with few carbides. According to this study the appearance of large and high carbides in the workpiece is worse for the cutting tool than for an homogeneous martensitic structure, even if the macro-hardness is the same for all the machined materials.

7. CRATER WEAR STUDY

The crater wear also shows a quantity of material removed by the chip during machining. However, the measure of the K_T is alone in inadequate to characterize tool-wear because it gives only a geometric value and it does not give the volume of the crater. A method to generate the pattern of the crater has been used in order to observed the progression of the tool material removal rate. Figure 6 shows the parameters used. The crater shape has been generated only for the largest section of this one as shown in Figure 6.
According to Figure 6, three points need to be defined; *a*, *b*, *c*. As a result, two different graphs, Figures 7, and 8 were plotted. Figure 8 was plotted after analyzing the data of all 16 wear tests. It shows that the ratio L_f/L is not a function of cutting speed, feed rate, or the workpiece material, indeed, it is independent of these parameters. With the results shown, the crater pattern is known in three points, and the equation that links the values L, L_f, and K_T are known.

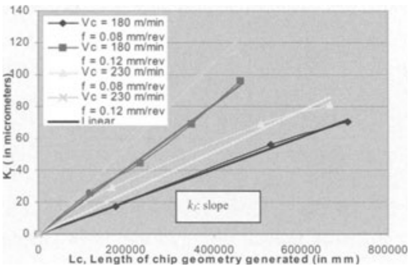


Figure-7. Progression the crater depth KT in machining of X155CrMoV12 steel.

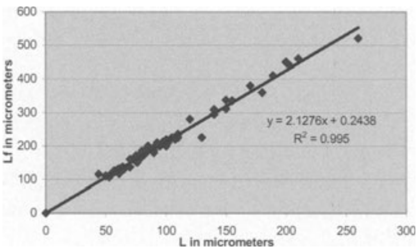


Figure-8. Ratio Lf /L

An approximation of the crater curve as two degree polynomial where the k_i is constant defined by the experiments was performed.

8. CONCLUSIONS

This study has shown the influence of the steel microstructure on tool-wear in CBN tools. The major influencing parameter on tool-wear is the presence of carbides in the steel microstructure. The commonly known crater wear and the flank wear are linked, however the ratio between the flank wear and the crater wear is fixed by the microstructure of each steel. Indeed, even if the 100 Cr6 and the X155 CrMoV 16 steels have flank wear rate very close, the crater wear is different. The position and the size of the dead zone seem to be always a constant independent of the cutting conditions. Another important aspect is the influence of cutting speed for steels with only martensite grains. For these steels, the increase in the cutting speed has a greater impact on the tool-wear rate. The lower effect on steel containing carbides could be explained by the fact that these carbides are not affected by the cutting temperature (they can still be observed in the white layer of the X155 CrMoV16 chip).

9. REFERENCES

[1] W. KÖNIG, R. KOMANDURI, H.K. TÖNSHOFF, G. ACKERSHOTT "*Machining of hard materials*", Annals of the CIRP, Vol. 33/2/1984, pp. 417-427.

[2] N. NARUTAKI, Y. YAMANE, "*Tool wear and cutting temperature of CBN tools in machining of hardened steels*", Annals of the CIRP, Vol. 28/1/1979, pp. 23-28.

[3] S.Y. LUO, Y.S. LIAO, Y.Y. TSAI, "*Wear characteristics in turning high hardness alloy by ceramic and CBN tools*", Journal of Materials Processing Technology, Vol.88, 1999, pp. 114-121.

[4] J. BARRY, G. BYRNE, "*Cutting tool-wear in the machining of hardened steels*", Wear, Vol. 247, 2001, pp. 152-160.

[5] M. ZIMMERMANN, M. LAHRES, D.V. VIENS, B.L. LAUBE, "*Investigation of the wear of cubic boron nitride cutting tools using auger electron spectroscopy and X-ray analysis by EPMA*", Wear, Vol. 209, 1997, pp. 241-246.

[6] S.A. KLIMENKO, Y.A. MUKOVOZ, V.A. LYASHKO, A.N. VASHCHENKO, V.V. OGORODNIK, "*On the wear mechanism of cubic boron nitride base cutting tools*", Wear, Vol. 157, 1992, pp. 1-7.

[7] S. ENOMOTO, M. KATO, S. MIYAZAWA, T. ONO, "*Characteristics of tool-life of CBN cutting tool in turning chromium-molybdenum steels of various hardness*", Bulletin of Japan Society of Precision Engineering, Vol.21, No.3, Sept 1987, pp. 209-212.

[8] T. OHTANI, H. YOKOGAWA, "*The effects of workpiece hardness on tool-wear characteristics*", Bulletin of Japan Society of Precision Engineering, Vol.22, No.3, Sept 1988, pp. 229-231.

[9] G. POULACHON, Phd Dissertation, ENSAM Cluny, France, December 2nd 1999, p.164.

[10] W. KÖNIG, M. INDING, "*Fine turning and drilling of hardened steels*", METALWORKING IDR 3/90, pp. 139-142.

ADVANCED NUMERICAL SIMULATION OF THE CRIMPING PROCESS OF ELECTRIC CONNECTORS

M. Abbas, J-L. Batoz, T. Courtin and G. Dhatt
FCI France, Corporate Research Center, La Ferté-Bernard

Abstract: Crimping is a classical technology process to ensure the electrical and the mechanical link between a wire and a connector. Numerical modelling of the process is helpful to choose and to optimize the dimensions of the crimping part of the connector. In this paper, we discuss a 2D simulation of the crimping process, using implicit and explicit finite element methods (ABAQUS/Standard and ABAQUS/Explicit) and we compare the results with experimental data from the industrial process of crimping (geometry, shape, surfaces and punch force). This non-linear problem involves large elasto-plastic strains and multiple contact conditions, with friction between the strands and the grip. One of the major difficulties of the simulation is due to the definition of all possible contact couples between strands. The explicit method is preferred for the modelling of multi-contact problems, in spite of the quasi-static process of crimping. Thus, some simulations with the implicit method have been performed to compare the results and tune the simulation parameters of the explicit approach (space and time discretization). Subsequently, parametric studies are performed to show the effect of the friction ratio or the position of the strands in the wire.

Key words: Electrical connectors, crimping process, finite element method, large elasto-plastic strains, contact with friction.

1. INTRODUCTION

Crimping is the operation which links a wire with the contact by the folding of two wings around the wire (see Fig. 2 and 3). The development of the electronic systems in cars increases the number of connections and

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 241–250.

requires more reliability of wiring harnesses. Moreover, to reduce the weight of cars, the section of the wire decreases. Mechanical simulation of the crimping process is important to guarantee a high degree of reliability in the connector: knowledge of the electrical contact resistance depends on the geometry and on the stress state of the connector after crimping.

A limited number of papers has been published on the subject of numerical simulation of the crimping process. Villeneuve & al. [4] and Berry [2] used a dynamic explicit formulation, mainly because of the multi-contact problems. Their results present the advantage of numerical simulation and show good agreement between the final geometry obtained by the experimental method and by numerical simulations. Morita & al. [3] try to explain the problems of springback by using a visco-plastic model.

This paper presents some results obtained by a 2D finite element method of the crimping process. We use an explicit formulation in spite of the quasi-static aspect of the crimping process mainly because of the numerous and severe contact conditions. Meanwhile, some implicit simulations were carried out to allow the tuning of the computational parameters of the dynamic simulation, especially for the springback evaluation.

2. PROBLEM DESCRIPTION

2.1 Geometry and terminology

A 3D view of the crimping part of the connector with a seven strand wire is shown in Fig. 2 (before complete crimping).

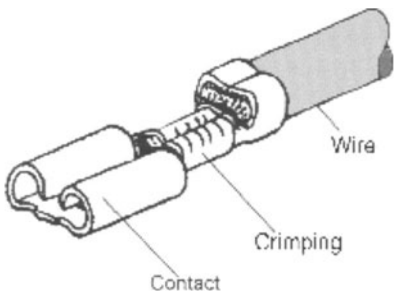


Figure -1. A simple electrical connector

The "U"-form of the connector is called the wing or grip. For the B-shaped crimp, we use a double-curved punch and a curved die (Fig. 3). The die is fixed and the punch moves downward to fold the wing around the wire.

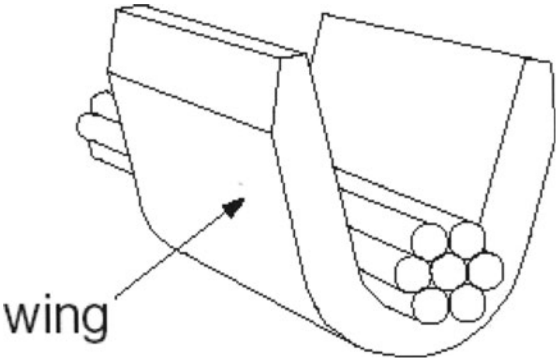


Figure - 2. 3D view of the crimped part of an electric connector

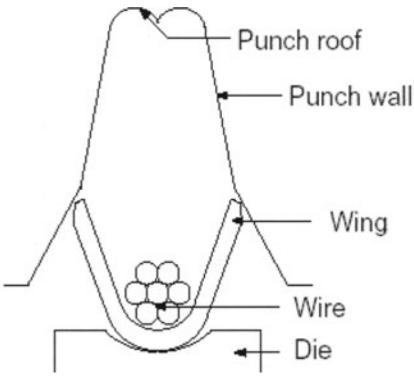


Figure -3. Baseline geometry for the 2D simulation

Table 1 presents the geometry of the wires considered in the present study.

Table -1. Description of the wires

Section – Commercial (mm^2)	Number of strands	Diameter of strands (mm)	Section area (mm^2)
0.35	7	0.25	0.34
0.50	7	0.30	0.49
0.60	12	0.25	0.59
0.75	19	0.22	0.72

2.2 Material

The grip and the wire are made of copper materials. An elasto-plastic constitutive law with isotropic hardening is considered for the simulation (Fig. 4)

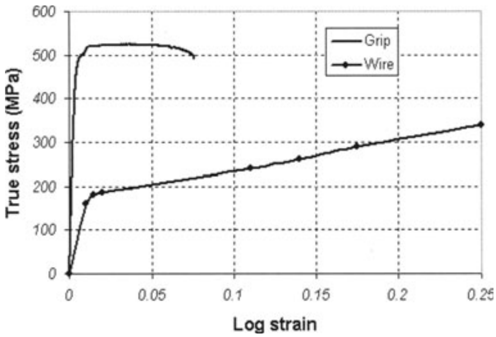


Figure -4. Uniaxial stress-strain curves

These curves are obtained by tensile tests but the crimping involves large compression strains and we need to extend the hardening curve to large strain values by slope extrapolation.

3. SIMULATION MODEL

3.1 Analysis techniques

Crimping is a very non-linear forming process involving

– large elastoplastic strains;
– multiple deformable body contact conditions with friction;

In the present study, the simulations were performed using the commercial package ABAQUS (Standard and Explicit). Explicit solvers are very efficient for forming problems, especially when they involve complex contact conditions. However, the problem is quasi static and some simulations were performed with the implicit solver to check and to validate the results obtained by the explicit approach. The influence of the punch speed given special consideration. A classical Von Mises plasticity model with isotropic hardening law is considered for the material behaviour.

3.2 Finite element type

The crimping process is a 3D problem. But to reduce the computing time, we simplify the model and study a 2D cross-section. The depth of the grip part is 3.4mm and we did the plane model hypothesis. As we can observe in Fig. 5, the plane stress assumption is better because it allows the out-of-plane extrusion and we can observe this extrusion on real crimped connectors. The plane strain model is too stiff and the simulation does not converge because of the incompressibility locking of the model. Between these two choices, we consider also the generalized plane strain assumption which allows a limited extrusion between two rigid planes. The influence of mesh density was studied in a number of tests. No mesh adaptivity was considered. Triangular (T3) and quadrilateral elements (Q4) with linear approximations have been used.

3.2.1 Friction

Coulomb's model of friction is considered with a friction coefficient between 0.10 and 0.30. Different coefficients have been used between the wing and the punch wall and between the wing and the die.

3.2.2 Punch velocity

The velocity of the punch is about 0.5 m/s during the process: crimping can be considered as a quasi-static process. Using a dynamic explicit solver, we have to increase the punch velocity to find a compromise:

- high velocity leads to low computing time. The explicit method is conditionally stable and the longer the time length observed by simulation, the higher the number of increments needed is;
- low velocity is physically better since the phenomenon is quasi static;

Rules to simulate a quasi-static process with a dynamic solver can be extracted from a number of publications:

- the tool velocity must be less than 1% of the sound velocity in the material. For copper, the sound speed is about 6,000 m/s;
  low velocity is physically better since the phenomenon is quasi static;
- the kinetic energy during the complete simulation must be less than 1% of the total energy of the system;

From these considerations and with information extracted from [1], [2] and [3], we tried several velocities using ABAQUS/Explicit and we compared the results with the ones obtained using the implicit solver.

3.3 Results

3.3.1 Definition

The compression ratio t is defined by this equation and Fig. 5:

$$t = \frac{S_{crimped}^{final} - S_{wire}^{initial}}{S_{wire}^{initial}}$$

$S_{wire}^{initial}$ is the initial wire area

$S_{crimped}^{final}$ is the final crimped area

We define two other geometrical parameters (see Fig. 5): the crimping height Hc and the crimping width Wc.

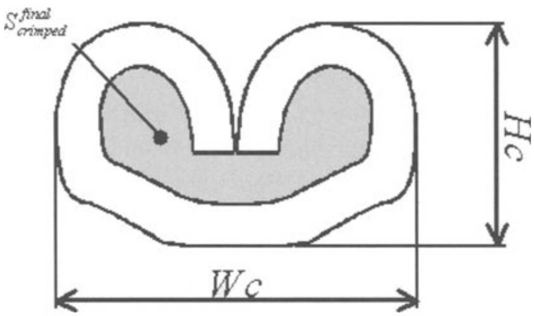


Figure -5. Geometrical parameters at the final stage of crimping

The crimping ratio is very important and is the most frequently used factor to determine the quality of crimping. That ratio is usually around 15%.

3.4 Implicit simulations

The results of the two formulations (plane stress and generalized plane strain) are compared in Fig. 7.

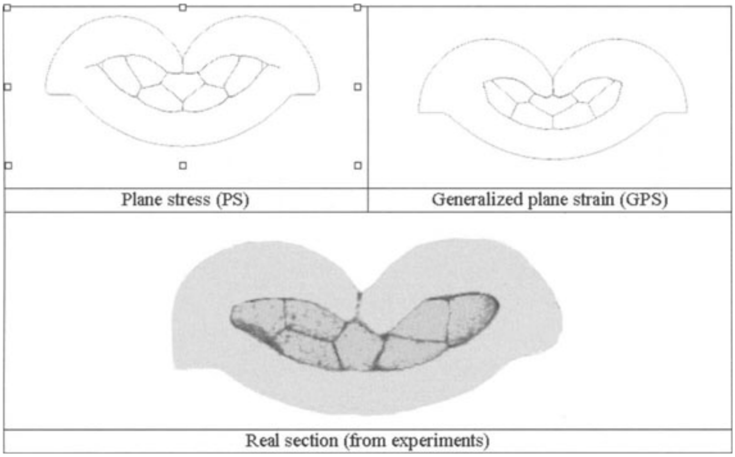


Figure -6. Deformed cross-section - Effect of formulation

These deformed shapes (starting from the situation of Fig. 3) have been obtained with 1,566 quadrilateral elements to discretize the wing and 700 elements to represent the strands. The punch displacement is controlled (5 steps) and the CPU is about 6h30 for plane stress and 3h30 for generalized plane strain on a Unix workstation (alpha 866MHz bi-processor).

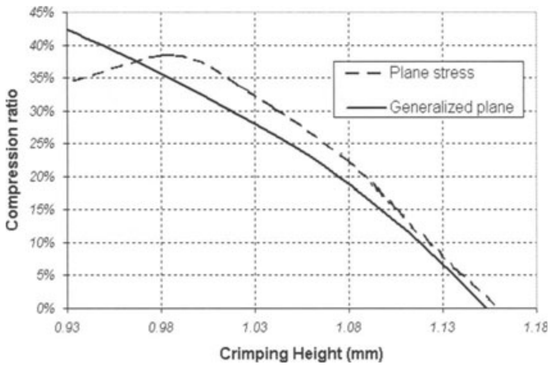


Figure -7. Compression ratio and crimping height for two elements formulation

The generalized plane strain method is too stiff and the final compression ratio is found to be too high. The plane stress formulation which permits out-of-plane extrusion appears to be the best model to simulate 2D crimping. For the 0.35mm² wire, we see in the Fig. 9 a close correlation between the experimental values and those of the simulation.

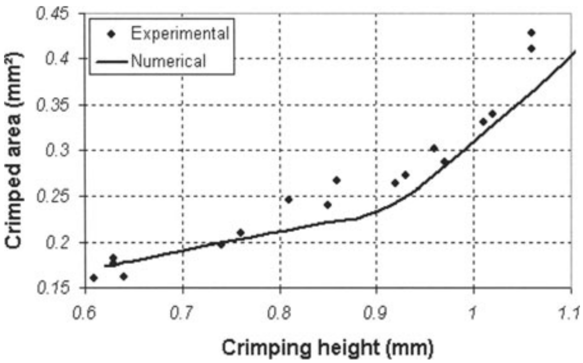


Figure -8. Compression ratio and crimping height for the 0.35mm² wire

3.4.1 Partial conclusion

The implicit method is possible but it is difficult to achieve convergence during tuning of the model. In order to minimize the rigid body movements of the strands, we have to add some springs and to focus attention on the geometry (no gap between parts) and on the friction coefficients. This method requires a lot of computing time but allows to simulate springback and to find adequate parameters for the explicit method (as the punch velocity for example).

3.5 Explicit simulations

3.5.1 Punch speed

Using the dynamic explicit approach, we need to reduce the amount of kinetic energy. The value of the punch velocity is very important. Therefore, we perform different analyses at different velocities. It appears that inertial effects are not significant if the punch velocity is less than 15 m/s. The results presented in the following figures have been obtained without mass scaling, with Δt around 2.0x10-9s and for a punch speed of 10m/s.

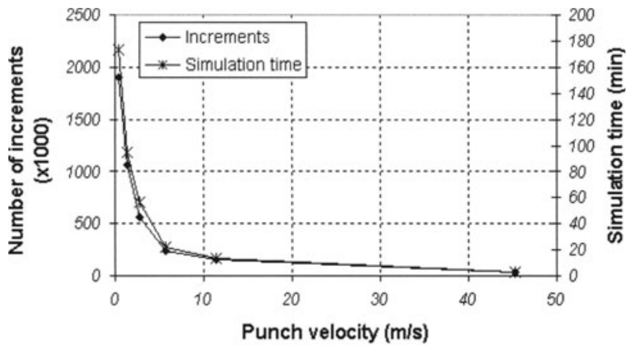


Figure -9. Computing time and cost of the explicit approach

Fig. 10 shows the cost of computing using the explicit approach. The computing time decreases hyperbolically with the punch speed.

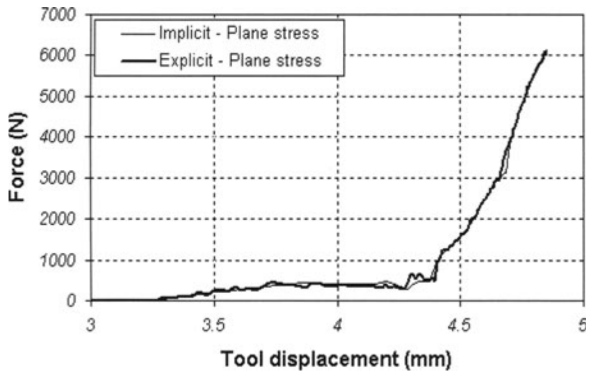


Figure -10. Comparison of explicit and implicit methods comparison

In Fig. 11, we compare the punch load punch displacement curves using the explicit and implicit approaches. It is demonstrated that both approaches give quite similar results. The CPU is about 40min for the explicit approach with a punch velocity of 10m/s.

3.5.2 Effect of global friction coefficient

We studied the effect of uniform friction coefficient on the final deformed shape. In Fig. 14, we observe that an excessively high friction coefficient leads to over-compression of the wings does not allow them to properly surround the wire.

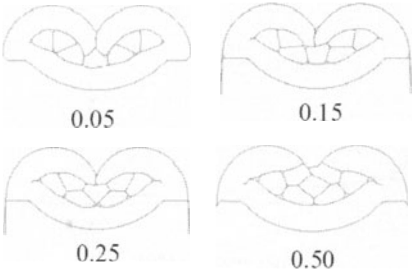


Figure -11. Effect of global friction coefficient

4. CONCLUSIONS

The numerical simulation of the crimping process of electric connectors using advanced FE packages such as ABAQUS is possible. Good results can be obtained using explicit dynamic approaches as found in ABAQUS/Explicit. We have shown that a good representation of the out-of-plane extrusion is possible using 2D plane stress elements. The strand configuration shows less than 5% error on the final configuration. The friction coefficients are important to obtain a nicely crimped shape. The study of the springback phenomena is also an objective as well as complete 3D simulations in order to achieve a full understanding of the process and before performing optimization of the process parameters.

5. REFERENCES

[1] "ABAQUS Theory Manual", Version 6.2, Hibbit, Karlsson & Sorensen Inc., 2001.

[2] D.Berry. "Development of a crimp forming simulator", ABAQUS Users' conference, 1998.

[3] T.Morita, Y.Saitoh, M.Takahashi & al. "Numerical model of crimping by finite element method", IEEE Holm, 1995, pp. 154-164.

[4] G.Villeneuve, D.Kulkarni, P.Bastnagel et D.Berry. "Dynamic finite element analysis simulation of the terminal crimping process", IEEE Holm, 1996, pp. 156-171.

SURFACE INTEGRITY IN HARD TURNING

J. Rech and C. Lescalier*
LTDS, ENISE, Saint-Etienne, France
**ENSAM, Metz, France*

Abstract: Highly-stressed steel components, e.g. gears and bearing parts, are appropriate applications for hard turning. Therefore the process effects on significant engineering properties of work materials have to be carefully analyzed. Roughness, residual stresses and white layers as parts of surface integrity, are function of the machining parameters and of the cuttability of the cutting edge, i.e. of the tool wear. The aim of this work was to study the influence of feed rate, cutting speed and tool wear on the effects induced by hard turning on case-hardened 27MnCr5 gear conebrakes and to point out the technical limitations in mass production.

Key words: hard turning, surface integrity, residual stresses, white layers

1. INTRODUCTION

Traditionally, the machining of hardened steel components has been the domain of grinding. In recent years, c-BN (cubic boron nitride) tooling has proven to be a viable alternative, providing both environmental and cost benefits. It offers the possibility of greater process flexibility, reduced machining time, lower energy consumption, swarf recycling possibilities and the optional use of coolant. Despite these obvious advantages, industrial realization of hard machining has not risen in comparison with the potential range of applications. The clearly unsatisfactory industrial acceptance of hard machining technology can be attributed partly to insufficient knowledge of the component behaviour of hard machined technical surfaces and partly to the uncertainty concerning the attainable accuracies-to-size. Particularly, the presence of residual tensile stresses at the surface of a component and the

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 251–260.

formation of a hard white layer have usually been considered as detrimental to the component's performance (rightly or wrongly).

Highly-stressed steel components, e.g. gear and bearing parts, are good applications of hard turning. Investigations into residual stresses, white layer and surface roughness (surface integrity) as a function of machining parameters and tool flank wear were undertaken on gear conebrakes of the same composition and thermal history.

2. EXPERIMENTAL PROCEDURE

Field tests were undertaken on gear conebrakes to define the limitation of the hard turning process in large scale production. Gear conebrakes are made of case-hardened steel 27MnCr5.

The cementation provides an 850 Hv0.3 hardness on the external surface and affects about 0.6 mm of the bulk material. The composition of the case-hardened layer is just about constant at a regular depth of 0.3 mm. In this zone, the carbon concentration of the layer is modified : about 1% carbon. As a consequence, the machined material is close to a 100MnCr5.

The case-hardening also modifies the residual stress profile and induces compressive stresses. The machined depth of cut is 0.15 mm because of the previous turning operations (before heat treatment). The surface discovered by the hard turning operation should still have a 850 Hv0.3 hardness and a tangential external residual stress of -400 MPa, if the cutting process has not affected the machined surface (which is not usually the case).

Continuous dry turning tests were performed on a high rigidity lathe. Finish cutting was conducted at a cutting speed V_c in the range of 50 to 250 m/min, a feed rate f in the range of 0.05 to 0.2 mm/rev and a depth of cut of a_p = 0.15 mm. The c-BN tool inserts were ISO code TNGA 160408 S (chamfer normal rake angle γ_n = -20°, chamfer width : 0.1 mm, honing edge radius r_β= 0.03 mm) coated with a 1 μm TiN layer. The tool holder was a MTJNR27050-16. The cutting edge inclination angle of the insert is λ_s = -6°, and the normal rake angle is γ_n = -6° . An EDS analysis of the c-BN insert has shown that it is mainly composed of about (in weight) 60% BN, 35% TiCN and some few elements as Al, W and Co.

Tool flank wear was measured after each test by an optical microscope. Two and three dimensional surface roughness values were measured on the conebrakes using a stylus profilometer. The two dimensional surface roughness was measured using a trace length of 4.8 mm, a cut-off length of 0.8 mm, and a gauss filter. The values were recorded at five equally spaced locations around the circumference of the specimen. Mean values are plotted in figure 1.

3. SURFACE ROUGHNESS BEHAVIOUR

In finishing operations, as in hard turning, the end of tool life is usually based on a predetermined damage level of the machined surface. So as to explain and predict the surface roughness, investigations have been conducted to determine the effect of parameters such as feed rate f, cutting speed V_c and flank wear VB. Three parameters of surface roughness have been measured : R_a, R and R_{max}.

The general topography of the machined surface consists of long straight grooves in a direction parallel to the cutting velocity. These grooves are caused by the micro-geometry of the cutting edge. Examination of the machined surfaces (figure 1a) using three-dimensional topography reveals the dependence of surface roughness on tool radius as well as on feed rate. This kind of surfaces is forbidden for a lot of applications (oil-tightness). This kind of problem does not concern gear conebrakes, but can be avoided with additional abrasive processes as lapping.

The theoretical expression of the surface roughness parameter R_a is :

$$R_a = \frac{f^2}{18.\sqrt{3}.R_\varepsilon} \quad (f\text{: feed rate (mm/rev)}, R_\varepsilon\text{: insert radius = 0.8 mm}) \qquad (1)$$

This model is based on a perfect geometrical model, made of circles (radius 0.8 mm) with a pitch of f mm. In this expression, flank wear is ignored, which is justified by the fact that all experiments were conducted with new inserts (no flank wear).

The measured experimental values are shown in figure 1b and 1c. One can observe that the feed rate is the main parameter that influences the surface roughness, compared to the influence of the cutting speed. Figure 1b also indicates that the evolution of the experimental values is almost the same as predicted by eq. (1), irrespective of the cutting speed (in the range 50 to 150 m/min).

Experimental curves show that the cutting speed has a smaller influence for finishing operations, especially at a low feed rate. The results for cutting speed higher than 150 m/min (200 and 250 m/min) have been presented to show that an evolution significantly different can be observed. These results are not suitable for large scale applications, because the inserts are quickly worn after just a few seconds. As a consequence, it can be concluded that, in the suitable cutting conditions area, the surface roughness parameter R_a is not significantly influenced by the cutting speed, but is mainly influenced by the feed rate.

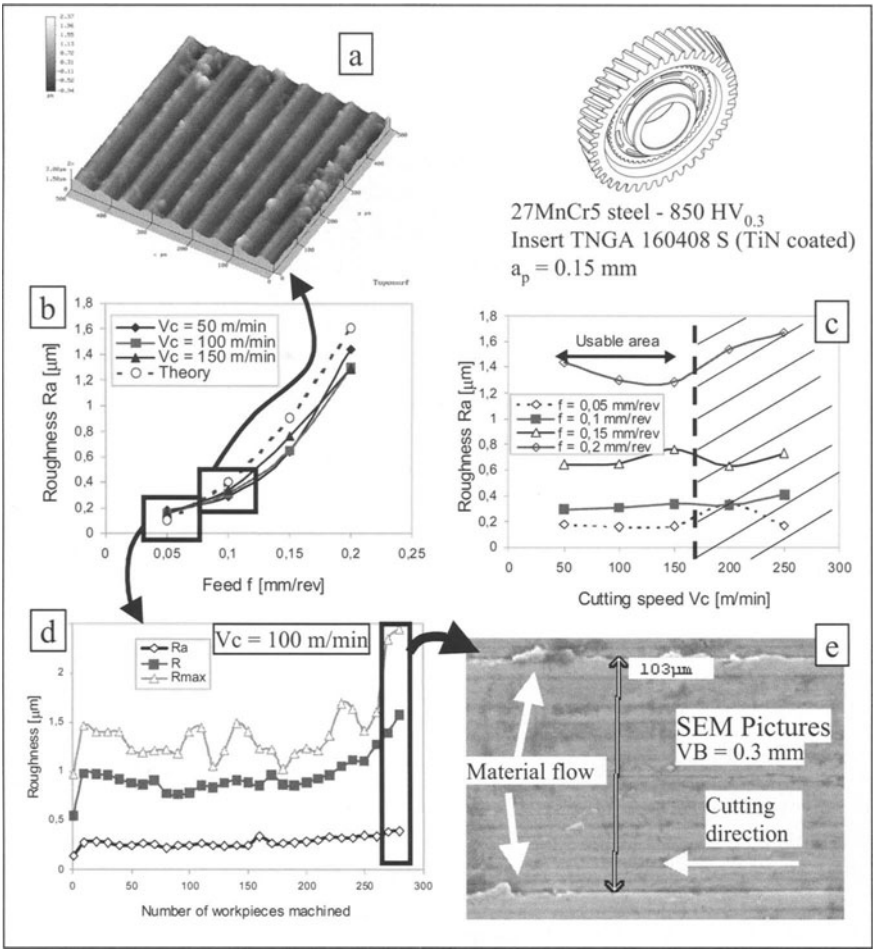


Figure -1. Influence of cutting parameters on the surface roughness parameters after finish hard turning - Steel 27MnCr5 (850 Hv0.3) a_p = 0.15 mm – Insert TNGA160408S + TiN

Observation of the surface machined at lower feed revealed the existence of a severe plastic flow (identical pictures as those observed in figure 1e). Despite the fact that a good surface finish was obtained using a small feed (where the height of the feed marks becomes smaller), a close examination of the machined surface shows that extensive material flow existed. A typical SEM picture of material side flow can be observed in figure 1e. As presented in [1], the material side flow is defined as a displacement of the workpiece material in a direction opposite to the feed direction such that burrs form on the feed mark ridges. Workpiece material in the cutting zone is subjected to temperature and pressure conditions high enough to cause a complete plastification. Chip material flows in a direction perpendicular to that of the chip. This material sticks to the newly machined surface and causes damage to the machined surface quality, even if the surface roughness is kept within the desired tolerance. In addition, the adhered material is hard and abrasive, such that it wears any surface that comes into contact with the produced surface.

A lower feed increases the area in which the chip thickness was lower than the minimum chip thickness. Hence, instead of cutting, a large part of the material was ploughed and led to material side flow. This is confirmed by the evolution of the experimental R_a curves which are higher than the theoretical curve for a 0.05 mm/rev feed rate. This fact is borne out by the damage of the surface roughness at low feed rate. As a consequence, it seems that a minimal feed rate exists. This minimum value is in relation with the minimum chip thickness and the honing edge radius (0.03 mm). In this case, the minimum feed rate seems to be between 0.05 and 0.1 mm/rev.

Additionally, figure 1d shows the effect of tool wear on the surface topography during the machining of hardened steel. It represents the evolution of the surface roughness parameters (R_a, R, R_{max}) with the number of workpieces produced, i.e. with the flank wear [1]. Due to the thermal effect of tool wear, the material in the cutting zone becomes so viscous that it fills the grooves and flows in a uniform and homogeneous way to the side of the cutting tool, forming high ridges.

Furthermore, one can observe in figure 1d that the surface roughness parameter R_a is not very sensitive to the flank wear and, as a consequence, to the occurence of material side flow. On the contrary, R and R_{max} are much more sensitive to the material flow. This testifies to the necessity of having both types of parameters : one that indicates the mean surface roughness of the surface (justified by the functionality of the surface) and an other that indicates the damage of the tool (flank wear and material side flow).

4. RESIDUAL STRESSES AND WHITE LAYER BEHAVIOUR

The level and profile of residual stresses are some of the major criteria on gear conebrakes, because of the high shearing load involved on these surfaces. Residual stresses as a function of cutting speed, feed rate and flank wear were investigated. Using a X-ray diffractometer, combined with a chemical polisher, depth profiling of residual stresses was performed. Figures 2 and 3 present the results of these investigations.

Residual stresses are a function of three factors : mechanical, thermal and metallurgical factors. In hard turning, these factors are governed by tool composition, tool wear, tool geometry, machining parameters and the interaction of the tool material in the workpiece material. The changes in the physical properties of the workpiece surface due to hard turning are attributed partly to the mechanical stresses and partly to the temperatures arising during cutting. In order to compare the changes caused by hard turning, it is essential to analyse the corresponding chip formation mechanism. This mechanism has already been discussed by several authors [2]. In the area around the tip of the cutting edge, the compressive stress levels must be very high, as this is the only way of ensuring that the workpiece will plastifiy to a sufficient degree to allow chip formation. The high level of mechanical stress being exerted on the surface of the workpiece tends to induce compressive residual stresses.

Thermal stresses result mainly from the friction between the wear land *VB* and the workpiece. The high direct stress levels cause high tangential stress which, in conjunction with relative motion between cutting edge and workpiece, result in high levels of friction power. It must therefore be assumed that most of the heat flows into the workpiece. Additionally, it is evident that the wear land friction alone can result in extremely high temperatures which, however, do not penetrate deep into the workpiece. Since [2], the temperature distribution, measured with a CCD infrared camera in orthogonal c-BN hard turning of a steel 100Cr6 (62 HRc - V_c = 184 m/min - a_p = 2 mm - f = 0.1 mm/rev - insert c-BN), shows that the maximum temperature is located on the flank face and reaches between 800 and 1100°C. When these temperatures exceed the γ - α transition temperature (transition temperature is dependent on the heat rate), martensite produced by friction develops, which is recognisable as a white layer in micrographs. The formation of martensite as a result of friction causes residual tensile stress which surperimposes itself on the residual compressive stress resulting from mechanical compression.

In all experiments, the profiles of residual stresses in the axial direction had similar characteristics to those in the circumferential direction. Therefore only the circumferential residual stresses mesurements have been presented.

A typical profile of residual stresses in the circumferential direction is shown in figure 2c. For a new insert and a cutting speed of 100 m/min, the external residual stress is about -250 MPa. Beneath the surface the profile decreases to -800 MPa at a distance of 0.07 mm, and then increases to the level of the bulk material (-400 MPa). This profile is very interesting for the resistance to fatigue, as the level is always compressive. Similar profiles have been described by [2].

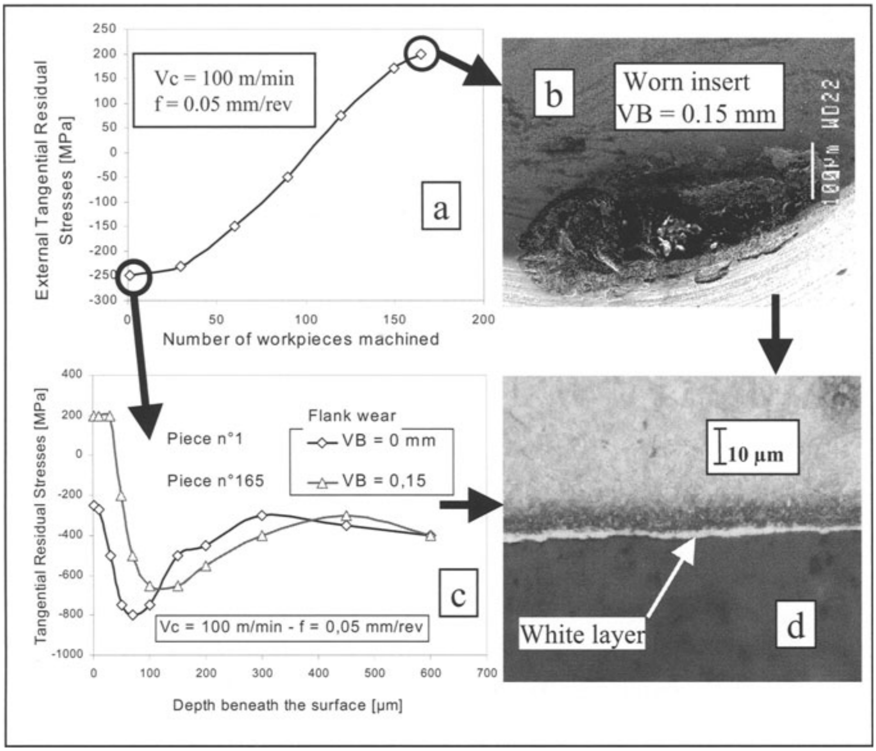


Figure -2. Influence of flank wear on the profile of residual stresses after finish hard turning - Steel 27MnCr5 - 850 Hv0.3 - Insert TNGA 160408 S (TiN coated), a_p = 0.15 mm.

4.1 Residual stresses and flank wear

Additionally, figure 2c shows the influence of flank wear on the profile of residual stresses. One can observe that the external residual stress increase with flank wear and, at the same time, the maximum residual compressive stress shifts further below the surface. Flank wear increases the level of friction energy and thus the cutting temperatures. This indicates that a new cutting tool generates a surface residual compressive stress, whereas a worn tool tends to generate a residual tensile stress at the surface.

Furthermore, the microstructure of workpiece n°165 produced with the cutting conditions described in figure 2 reveals the presence of an affected zone, which is about 0.01 mm thick. After etching using a 5% nital solution for 5 s, a white layer is observable, followed by a dark layer, and finally the bulk material. White layers are found in many removal processes such as grinding (since [4]). White layers seem to be detrimental to product performance, and therefore require either a post-finishing process or must be avoided.

Since [3], white layers are a product of friction and heat, generated either by the cutting tool wear or by the cutting speed. White layer formation is mainly a thermal process involving phase transformation of the steel, possibly plastic strain activated. Actually, the microstructural evolution during white layer formation is not fully understood. Nevertheless, it is possible to say that, for hypereutectoid steel as 100MnCr5, martensite, the starting microstructure in hard turning, is a metastable structure that will decompose to ferrite and cementite when heated (a tempering process). However, the high heating rate encountered during cutting (2.1×10^6 °C/s according to [2]) may prevent martensite from decomposing (no time to respond). Instead, it may transform diffusionlessly to austenite (reverse martensitic transformation) with no carbide dissolution needed, as martensite has a high carbon concentration. This is in contrast to conventional hardening where the starting structure of soft steel is ferrite with cementite that requires soaking time for cementite dissolution into austenite. Carbides in affected zones show no differences from the bulk, further suggesting an absence of carbide dissolution during white layer formation. Since [2], the X-ray diffraction analysis of a white layer has shown that the volume fraction of austenite is higher than in the bulk area. This indicates the important phase transformations that have occurred during the cutting. The substantial increase of austenite at hard turned surfaces is probably due to an insufficient cooling rate or a lack of tempering when the machined surfaces encounter heating (rehardening) and self quenching during cutting.

As a conclusion, it has to be pointed out that tool flank wear is a major parameter in finish hard turning, which can lead to an early change of insert overriding predetermined forecasts based on other parameters (roughness). As an example, for a surface roughness criteria $R < 1.5$ µm, 280 workpieces can be machined, and at the same time, a criteria of $\sigma < 0$ MPa will stop the production after 100 workpieces. The main problem for a large scale production plant is the quick quantification of these residual stresses.

4.2 Influence of cutting parameters

Figure 3b shows the evolution of the external tangential residual stress with the cutting speed for different feed rates. One can observe that the cutting speed tends to increase the external residual stress, irrespective of the feed rate in the range of 50 to 150 m/min. On the contrary, the evolution of the curves changes above 200 m/min. It has to be noted that the values of residual stresses produced with a cutting speed of 200 and 250 m/min are useless for large scale applications. In these cutting conditions, the flank wear rate is so high that it is impossible to produce a workpiece in economically and technically acceptable conditions (presence of white layers). As a consequence, one can conclude that, in economical cutting conditions (50-150 m/min), the cutting speed tends to increase the level of external residual stresses. A similar conclusion has been proposed by [2-3] by indicating that cutting speed has almost the same effect as flank wear (described previously).

Figure 3a also shows the effect of feed rate on the residual stresses in the circumferential direction. In the range of 0.05 to 0.1 mm/rev, the residual stresses near the surface shifted towards compression as feed rate was increased. Conversely, in the range of 0.1 to 0.2 mm/rev, the residual stresses near the surface shifted towards tension as the feed rate was increased. These results are confirmed irrespective of the cutting speed (in the suitable range of cutting speed). The limit of feed rate (0.1 mm/rev) can be explained by the relation between the small chip thickness at low feed rate and the cutting edge honing radius (0.03 mm). With low feed rates under 0.1 mm/rev, the friction energy conducted in the machined surface should be higher, because the chip thickness is too small and a part of the chips may be squeezed below the clearance face, as described by [1].

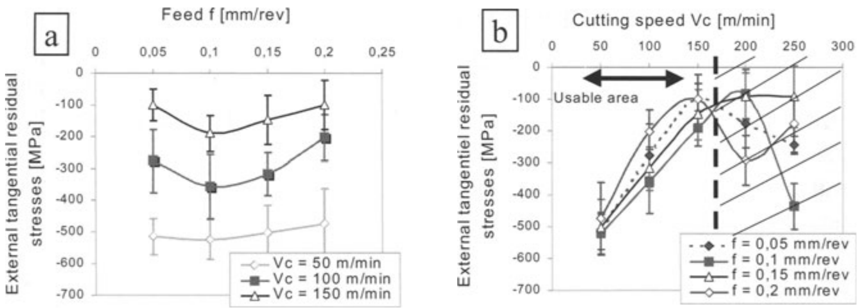


Figure -3. Influence of the cutting conditions on the external residual stresses after finish hard turning - Steel 27MnCr5 - 850 Hv0.3 - Insert TNGA 160408 S (TiN coated), $a_p = 0.15$ mm.

Nevertheless, compared to the evolution of the residual stress level with the cutting speed, one can conclude that the feed rate doesn't have a major influence on the residual stresses. Furthermore, at a depth below 0.03 mm, there was no change in the residual stress profiles (profiles not presented). These results confirm previous work of [5], which indicates that the feed rate did not significantly affect the residual stresses in the deep subsurface.

5. CONCLUSIONS

For highly loaded parts, e.g. gears, the physical and technological material properties are of major importance in the ability of a surface to perform required service functions; hardness, microstructure, residual stresses profile and level (surface integrity) are among such properties.

Finishing cutting processes have a great influence on the surface integrity because of the thermomechanical material removal mechanisms.

Hard turning process is interesting regarding its capacities to produce a low surface roughness (R_a < 0.2 μm) during a long cutting time and also to induce compressive residual stresses when machining at low feed rate and low cutting speed. Feed rate is the major parameter that influences the surface roughness, whereas cutting speed is the major parameter that influences the residual stresses level.

Hard turning process can be subject to some restrictions, especially due to the generated helical surface topography (notexistant with the cylindrical grinding processes) and the occurrence of material side flow at very low feed rates or with worn tools. Another restriction of hard turning is the influence of the flank wear which shifts the residual stresses towards tension and also tends to induce white layers. In hard turning operations, residual stresses level and white layer are the main criteria for inserts change, far before roughness or accuracies-to-size, although these parameters are very difficult to follow in a production plant.

The restrictions of hard turning could be solved by the association with a subsequent abrasive process (as lapping), which is supposed to remove the helical topography and the material side flow, and, at the same time, which should shift the residual stresses level towards compression [6].

REFERENCES

1. Kishawy H.A., Elbestawi M.A., Effect of process parameters on material side flow during hard turning, International Journal of Machine Tools and Manufacture, 1999, n°39, 1017-1030.
2. Poulachon G., Aspects phénoménologiques, mécaniques et métallurgiques en tournage c-BN des aciers durcis, PhD thesis, (Cluny, France, dec, 1999).
3. König R., Komanduri R., Tonshoff H.K., Ackershott G., Machining of hard materials, Annals of the CIRP, 1990, Vol.39/1, 417-427.
4. Shaw M.C., Vyas A., Heat affected zones in grinding steel, Annals of the CIRP, 1984, Vol.43/1, 279-282.
5. Matsumoto Y., Hashimoto F., Lahoti G., Surface integrity generated by precision hard turning, Annals of the CIRP, 1999, Vol.48/1, 1999, 59-62.
6. Rech J., Influence du toilage sur l'intégrité des surfaces issues de tournage dur, Proceedings of the workshop ' superfinition des pièces mécaniques', Paris, march 2001.

TECHNOLOGICAL PARAMETER CALCULATION AND STRETCH FORMING PROCESS CONTROL OF 3D PROFILE PARTS

S. Oding and N. Stourov
Applied Mechanics of Voronezh State Technical University, Russia

Abstract: The use of the complex shaped profile parts is very frequent in automobile and aircraft industries. A general manufacturing process is the stretch forming of a rectilinear workpiece over a die. This presentation proposes a simplified method to model and simulate the stretch forming process of 3D profiles. The profile is represented as a curved beam complying with the Bernoulli's hypothesis. This simplified model, taking into account the forming limits related to the defects appearance (fractures, wrinkles…), allows finding the laws of the profile end displacements and the forces to be applied. The simplified simulation method is being confirmed by the comparison with the results of the finite element analysis. The highly reduced calculation time and the ordinary necessary computer resources authorize the joint utilization of this simulation software and software controlling the stretch forming presses.

Key words: technological plasticity, stretch forming, 3D profiles, technological defects, process control

1. INTRODUCTION

The modern automobile industry is characterized, on the one hand, by gradual complication of a car exterior, on the other hand, by reduction of time and costs during production. Thus there is the constant search for new and retrofit of present manufacturing processes. One of the major problems during metal forming is the reduction of tool preparation time, of material costs for test parts and staff release due to the automation of the whole work cycle.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 261–270.

The stretch forming of three–dimensional profile parts is one of such processes. This operation consists in the tension of an originally rectilinear workpiece over a curvilinear surface of a die (tool) [2] (figure 1). The workpiece can additionally be twisted about its axis. Recently special–purpose equipment with CNC systems was created which realizes simultaneously the bend in two planes and the twist of profiles. However there remains the real problem of control data and die preparation.

The most effective method of part–program creation and of die surface design for stretch forming presses is that of mathematical modeling of the given process taking into account the workpiece's material behavior, the strain–stress state nature, and force and kinematical capabilities of a machine. As a result of calculation, the optimal loading program is determined which permits to avoid the appearance of defects (fracture of workpiece, wrinkles, and invalid springback). This program is transformed then into movements of press [1].

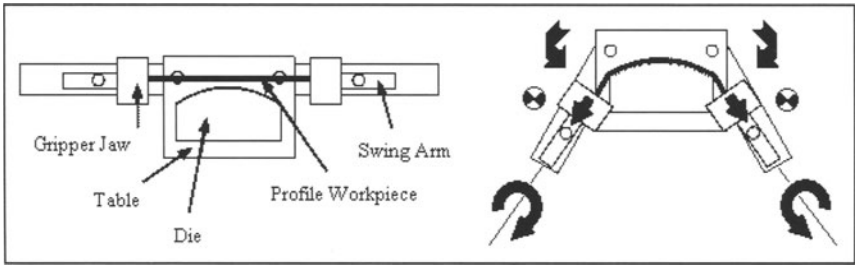


Figure -1. Profile stretch forming (©ACB Pressure Systems)

In the present work we propose a calculation method for the complex strain–stress state that appears while stretch forming three–dimensional profiles. This method is based on simplifying hypotheses: hypothesis of flat cross sections under bending and twist, assumption of material isotropic hardening during plastic deforming. The package PS3F is being created to solve the integral problem of technological data preparation and stretch forming equipment control.

2. NUMERICAL SIMULATION ALGORITHM

2.1 Geometrical description

The surface of the die is described in a fixed three–dimensional rectangular Cartesian coordinate system. Its origin is in top of the die. The axis Z is perpendicular to the table of the press (figure 2).

The shape of the part to be produced determines the die surface. The part (and workpiece) is described by a space curve connecting centers of gravity of cross sections. Furthermore this curve is named "profile (workpiece) axis". We suppose that the profile adjoins the die surface leaving no gap, cross sections are perpendicular to the workpiece axis and the die surface.

The position of workpiece particles (points) in the space is characterized using a mobile coordinate system. Coordinate ξ_i describes the distance read out lengthwise the workpiece axis from the die top up to the cross section with index *i*. The cross section is described in a flat rectangular Cartesian coordinate system. At a current point of workpiece the X-axis of this system coincides with the tangent to the workpiece axis, Y-axis – with the major normal, Z-axis – with the binormal. The origin of this coordinate system coincides with the center of gravity of the cross section. The cross section is described in a plane by one or several contours. Contours are composed of points and lines connecting them.

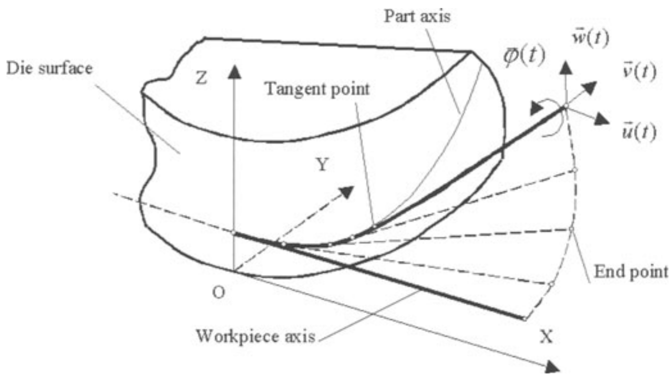


Figure -2. Schematic model of 3D profile stretch forming

2.2 Material model

The material of workpiece is assumed isotropic elastoplastic. In the elastic range the material follows Hooke's law:

(1) $de_{ij} = ds_{ij} / 2G$,

where e_{ij} is the strain deviator, s_{ij} is the stress deviator, G is the shear modulus. In the plastic range the material is hardened as isotropic according to the power law:

(2) $\sigma_e = A(\tilde{e}_0 + e_p)^m$,

where σ_e is the equivalent stress, e_p is the cumulative plastic strain, A, $\tilde{e}_0$, m are the parameters of the material flow curve approximation. Bauschinger's effect describes an anisotropy of material properties when changing loading sign (for example, stretching – compression). Within the framework of the given problem Bauschinger's effect is characterized by a ratio between yield limits in stretching and compression:

(3) $\beta = \sigma_{compr} / \sigma_{trac}$,

where σ_{compr} is the conditional yield limit in compression, σ_{trac} is the stress reached when stretching. In a general case the ratio β depends on material properties and cumulative plastic strain. The following approximation gives this parameter's value for different materials [4]:

(4) $\beta = \beta_m + (\beta_0 - \beta_m) \cdot \exp(-C \cdot e_p)$,

where β_m, β_0, C are the empirical parameters of Bauschinger's effect. This allows the expression of equivalent stress in the compressed zone to be written as:

(5) $\sigma_e = A(\tilde{e}_0 + e_p)^m - \Delta\sigma$,

where $\Delta\sigma = \sigma_{Trac}(1-\beta)$.

2.3 Strain-Stress State

The principal hypothesis for calculation of strains is the hypothesis of flat cross sections. So the flat cross sections perpendicular to the workpiece axis remain flat and perpendicular to curvilinear profile axis after bending and twist. Besides, the cross section does not distort in the plane ($\gamma_{yz} = 0$). The deformed state at a point of workpiece cross section is characterized by the tensor:

(6) $[E] = (\varepsilon_x, \varepsilon_y, \varepsilon_z, 0.5\gamma_{xy}, 0.5\gamma_{xz})$,

where indexes *x, y, z* relate to the axes of the workpiece cross section coordinate system.

According to the assumption of material incompressibility the linear components are related by the expression:

(7) $\varepsilon_y = \varepsilon_z = -0.5\varepsilon_x$.

For the determination of ε_x we suppose that the forming process is broken down into 3 phases:

a) Stretching and twist of the rectilinear profile workpiece before it touches the die surface;
b) Instantaneous bend at the tangent point;

c) Supplementary stretching, bending and twisting of the profile moving in contact with the die.

Thus,

(8) $\varepsilon_x = \Delta e_I + \Delta e_{II} - \chi \cdot y + \Delta e_{III}$,

where Δe_I is the strain of the 1st phase given by the loading program, χ is the profile axis curvature at the current point, y is the distance between the cross section current point and axis of cross section rotation while bending, Δe_{II} is the correction taking into account the fact that the axis of cross section rotation does not pass through the cross section center of gravity while bending, Δe_{III} is supplementary stretching of the 3rd phase.

We calculate shear strains using an expression of the small strain tensor:

(9) $\varepsilon_{ij} = \left(\partial u_i / \partial x_j + \partial u_j / \partial x_i\right)/2$.

The twist angle φ represents the angle between a cross-section's principal axis and the line of intersection "cross-section – bending plane". It is also necessary to take into account the curvilinear shape of the profile axis characterized by the value of torsion ν. Therefore we are introducing the concept of a total twist angle $\theta = \varphi + \nu \cdot ds$, where ds is the length of a profile axis elementary section limited by two cross-sections. The movements of points of the cross section plane which is turning as a single whole around its center of gravity are equal:

(10) $$\begin{cases} v = r \cdot \sin(\varphi_0 + \theta) - r \cdot \sin \varphi_0 \\ v = r \cdot \cos(\varphi_0 + \theta) - r \cdot \cos \varphi_0 \end{cases},$$

(11) $$\begin{cases} v = y \cdot \cos\theta + z \cdot \sin\theta - y \\ w = z \cdot \cos\theta - y \cdot \sin\theta - z \end{cases}.$$

where r is the distance between a cross section's point and the cross section's center of gravity.

Considering $\partial u/\partial y$ and $\partial u/\partial z$ as small values with respect to $\partial v/\partial s$ and $\partial w/\partial s$ we obtain:

(12) $\{\gamma_{xy} = \partial v/\partial s + \partial u/\partial y \cong \partial v/\partial s,\ \gamma_{xz} = \partial w/\partial s + \partial u/\partial z \cong \partial w/\partial s,$

Finally,

(13) $$\begin{cases} \gamma_{xy} = (d\varphi/ds + \nu)(z \cdot \cos(\varphi + \nu \cdot ds) - y \cdot \sin(\varphi + \nu \cdot ds)) \\ \gamma_{xz} = (d\varphi/ds + \nu)(-z \cdot \sin(\varphi + \nu \cdot ds) - y \cdot \cos(\varphi + \nu \cdot ds)) \end{cases}.$$

The stress state at a point is characterized by the tensor:

(14) $[\Sigma] = \left(\sigma_x, \tau_{xy}, \tau_{xz}\right)$.

The determination of stress components at a cross section point is made in one or two steps. In the beginning we perform an elastic estimation according to equation (1), that gives σ_{ij} and equivalent stress according to the formula:

(15) $\sigma_e = \sqrt{\sigma_x^2 + 3(\tau_{xy}^2 + \tau_{xz}^2)}$.

Then we validate the condition of the elastic zone using the Huber-von Mises yield criterion:

(16) $\sigma_e \le \sigma_S$,

where value σ_S characterizes the boundary between elastic and plastic zones at the given phase of deformation. It this limit is exceeded we update strain–stress state according to the power law expression (2). The stretch forming process takes place in conditions of the complex alternating–sign loading. To describe it, we apply equations from the theory of plastic flow [3]:

(17) $de_{ij} = \frac{1}{2G} ds_{ij} + \frac{3}{2} \frac{d\bar{e}^p}{\sigma_e} s_{ij}$,

where $d\bar{e}^p$ is the intensity of plastic strain increments. We convert these equations for our case according to the form of strain and stress tensors and solve them together with the equation (5). The volume change in compliance with the incompressibility condition is:

(18) $\varepsilon_0 = (\varepsilon_x + \varepsilon_y + \varepsilon_z)/3 = 0$,

(19) $\sigma_0 = \sigma_x/3$,

that gives the equality $de_{ij} = d\varepsilon_{ij}$.

(20) $\left\{ \sigma_x = \frac{\sigma_e}{d\bar{e}^p} \cdot d\varepsilon_x^p, \tau_{xy} = \frac{1}{3} \frac{\sigma_e}{d\bar{e}^p} \cdot d\gamma_{xy}^p, \tau_{xz} = \frac{1}{3} \frac{\sigma_e}{d\bar{e}^p} \cdot d\gamma_{xz}^p. \right.$

(21) $\left\{ d\varepsilon_x^p = d\varepsilon_x - \frac{1}{3G} d\sigma_x, d\gamma_{xy}^p = d\gamma_{xy} - \frac{1}{G} d\tau_{xy}, d\gamma_{xz}^p = d\gamma_{xz} - \frac{1}{G} d\tau_{xz}. \right.$

(22) $d\bar{e}^p = \sqrt{(d\varepsilon_x^p)^2 + \frac{1}{3}\left[(d\gamma_{xy}^p)^2 + (d\gamma_{xz}^p)^2\right]}$

(23) $\sigma_e = A(\tilde{e}_0 + e_p)^m - \Delta\sigma$,

where $e_p = \sum_{i=1}^{N} d\bar{e}_i^p$ is the cumulative plastic strain, i is a loading phase number.

The internal forces and moments must comply with the equilibrium equations:

(24) $N_k = \int_A \sigma_x \cdot dA$,

where N_k is the axial force, A is the cross sectional area;

$$(25) \quad M_x = \int_A \rho \sqrt{\tau_{xy}^2 + \tau_{xz}^2} \cdot dA,$$

where M_x is the twist moment, ρ is the distance from a cross section point up to the cross section's center of gravity.

3. VERIFICATION BY FEM

During the calculation process, the PS3F code forms the data for profile part and die surface 3D model creations. Also the output data contains the profile end movement trajectory. For that, along with applied stretching force and twist moment values, it computes the movement of end cross-section points. This information transfers to the finite element analysis when generating the model and boundary conditions.

Profile is meshed by cubic 20-node elements. Contacting surfaces of the die and profile are meshed by 8-node contact elements. A selected model of material plastic properties corresponds to the PS3F material model.

In order to compare the results of PS3F and ANSYS, we picked out the points in the upper, the most loaded, part of cross-sections located uniformly lengthwise to the profile contour. Results relate to one symmetrical half of the profile (figure 3, figure 4).

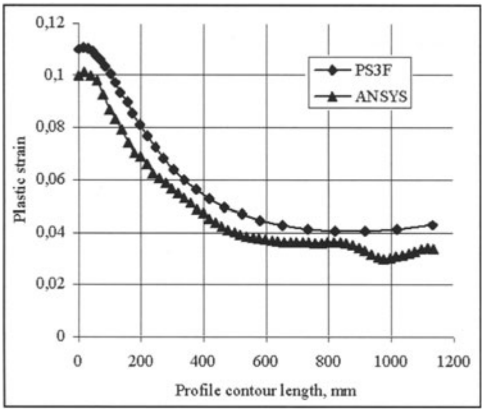


Figure -3. Cumulative plastic strain

Comparison of both calculation results shows their good conformity in the limits of typical technological demands. Slightly higher values given by PS3F appear from the hypothesis of cross-section shape and dimensions conservation during the forming process. These results allow us to draw a conclusion about the correct, on the whole, strain-stress state estimation proposed be the PS3F algorithm. Thus, the initial assumption about the proximity of the strain-stress state in this process to the uniaxial one is confirmed. Quantities determined in the PS3F calculation process allow estimation of the strain-stress state of profile material and the probability of technological defects appearing. Profile end trajectory at all forming steps serves the input data when setting the press movements and generating finite element boundary conditions.

4. FORMING PROCESS CONTROL

The principal aspect determining the forming process control strategy is the connection between the press actuator displacements and the current information about the workpiece's strain-stress state. In that case we accept the mathematical model of the forming process as the control object. The control criterion for a given process is the manufacturing of a part with necessary accuracy and characteristics. The problem of control consists in searching the workpiece end trajectory, which will ensure part production with minimal necessary plastic strains and acceptable probability concerning the appearance of defects. The main defects observed are fracture of workpiece, wrinkles, and springback.

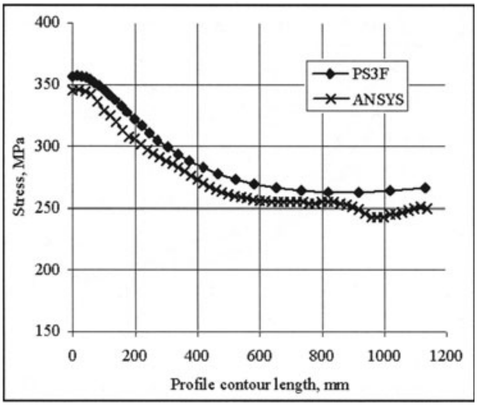


Figure -4. Equivalent stress

Plastic working processes may be characterized by the controllability, i.e. the possibility of change of material plastic flow conditions under the influence of variable parameters. Input parameters of the process are conditionally divided into the following groups:

– Uncontrolled (material constants, friction conditions, workpiece allowances, die geometry);

– Controlled (loading program, workpiece end trajectory).

The uncontrolled parameter value is set before the start of simulation and cannot be altered. That value may also possess some statistical straggling as, for example, properties of material and lubricant. Other parameters depend on equipment design features (workpiece length, gripper jaw allowances of workpiece etc.). The modeling algorithm actively modifies controlled parameters. The target is to satisfy defect criteria and limits, which form the basis of the physical object and process models of the system being investigated. Figure 5 illustrates the profile forming control procedure.

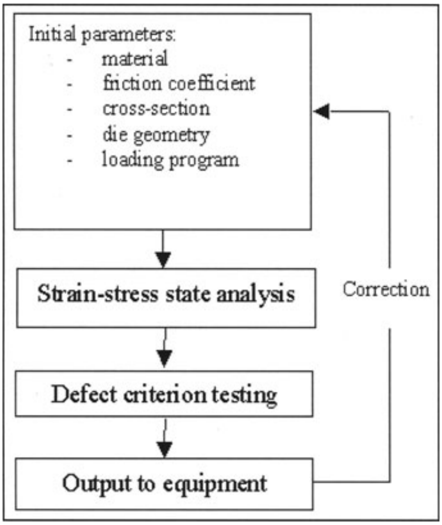


Figure -5. PS3F algorithm flowsheet

Compliance of all defect criteria and realization of the part contour that coincides with the theoretical one after springback is considered as the simulation process termination. The output information is composed of the corrected die contour, the workpiece length and the profile end trajectory.

The next stage will be recalculation of the profile end trajectory into the press actuator displacements. Each press has its force and kinematical limits. Thus, the displacements calculated during the simulation could never be carried out. That problem means that the part cannot be produced on the given press and that another press must be chosen. If the problem doesn't disappear after the equipment change, the simulation input parameters should be modified.

5. CONCLUSION

In the presented work the stretch forming process of three-dimensional profile parts was analyzed. A numerical algorithm was created to calculate the strain-stress state and estimate the appearance of defects. This algorithm basis consists of the set of mathematical models of material, die and part geometry, forming process. Calculation output data is the profile end trajectory ensuring the optimal loading program. Present calculation was tested and validated using the finite element simulation. The proposed algorithm allows the creation of industrial software to control the CNC of presses.

The described methods of the cooperation between PS3F and ANSYS/Structural enable to make more accurate strain-stress state estimation for complex and very important parts by means of FEM simulation. The use of the trajectory calculated by PS3F greatly reduces the ANSYS/Structural calculation time. The search for the optimal loading program with plasticity and contact using only ANSYS is not real in the industrial conditions because of the immense calculation time.

Today many aircraft factories utilize the technological simulation packages for sheet and 2D profile stretch forming processes. Their algorithms and control concept are similar to PS3F and there is good conformity between predicted optimal forming process parameters and the quality of produced parts. The automobile industry now needs modern manufacturing methods to answer the growing complexity of car design. One of the most efficient solutions is to introduce the vast experience in stretch forming simulation of aircraft parts for the purpose of cost reduction in car mass production.

REFERENCES

[1] S.S. ODING. "Technological possibilities of parts forming on the stretch forming presses with CNC", Doctoral thesis, Toula, 1989.

[2] S.S. ODING. "Synthesis of part programs for parts forming of sheet, profile and multi-layer panels by bending tension equipment with CNC", Forge and punch production, N°9, 1993, Moscow, pp. 13-16.

[3] L.M. KACHANOV. "Fundamentals of the Theory of Plasticity", Mir Publishers, Moscow, 1974.

[4] G. BACKHAUS. "Anisotropy of hardening. Theory in comparison with experiment", Mechanics of Solids, Proceedings of URSS Academy of Science, Moscow, Vol. 6, 1976, pp. 21-31.

APPLICATION OF THE STABILITY LOBES THEORY TO MILLING OF THIN WORKPIECES
Experimental Approach

L. Arnaud and G. Dessein
Laboratoire Génie de Production, ENIT, Tarbes, France

Abstract: The optimisation of cutting conditions in High Speed Machining (HSM) requires the use of a vibratory approach in order to avoid a fast deterioration of the tool and of the spindle, as well as a loss of quality of the surface finish. We suggest a transposition of the method of stability lobes to the case of milling thin parts, which is very typical in the aeronautical manufacturing context. After having modelled the dynamic behaviour of a blade and of the cutting efforts in side milling, we describe the zones of machining instability. An experimental validation allows us to emphasise the transition from stability to instability, in accordance to our theoretical results. The experimental profile is then compared with a computed profile. A decomposition of the different situations of contact between the tool and the part permits us to show the influence of back cutting in the model. Tests of machining then permit the quantification of its role. The objective of this work is the definition of a quick methodology for determining the optimal cutting conditions in a given industrial machining configuration.

Keywords: Stability lobes, milling, chatter

1. INTRODUCTION

The optimisation of the cutting conditions for the machining of turbo machine blades in High Speed Machining (HSM) requires control of the vibratory phenomenon. Vibrations of machines, that withstand high forces and are optimised in term of mass, can be controlled, notably by the appropriate filter. Vibrations of the part-tool couple are more difficult to eliminate. Many research publications have aimed at searching for an efficient strategy in order to avoid tool vibrations, but the proposed solutions have not yet been completely validated on industrial applications. Our objective is to apply and

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 271–280.

to experimentally validate the approach of stability lobes to the case of milling of thin walled parts with a constant surface roughness and an optimised depth of cut, and to identify the back cut effect.

2. THEORETICAL APPROACH

The analysis of chatter during the machining of metals has shown the mechanism of regenerative vibration to be the main cause of vibrations. The analytic model resulting from these works shows stable and unstable zones. Studies concerning the nonlinear aspects of the process have allowed a better understanding of the limits of the model and have shown the nonlinearities to be mainly located in the tool-part contact. Initially developed for turning machining, this approach has been largely used for many types of milling. Stability lobes have been largely validated experimentally, although through simple experiments [1]. In order to enrich the initial analytic model, many numerical models have been suggested [2] [3], and have notably allowed us to make the connection between stability lobes and step by step time simulation. The elimination of regenerative chatter is obtained using various methods :

- The diminution of the chip section (reduction of the advance for example) provides an elementary way for obtaining stable conditions, and is therefore the most employed technique when there is a lack of information.
- The modification of cutting conditions, in consideration of the stability lobes of the actual machining, allows the preservation of a short machining time, but requires experimentation or preliminary modelling.
- The utilisation of variable tooth pitch milling also allows us to eliminate vibrations for tools having several teeth, and for given cutting conditions [4].

It is also possible to perform a real time compensation of the vibrations, which requires the detection, the processing then the creation of the action of compensation itself. The detection can be made using an accelerometer, by force measurement, by displacement measurement, by microphone [5] or by monitoring the intensity of the spindle drive [6]. The data processing, either realised by spectral analysis or time analysis, either sophisticated or very simple, has in any case to be calibrated and very quickly carried out in order to prevent too strong vibrations appearing. The action of compensation can be performed by modification of feed speed [7] or of the frequency of rotation [8].

Others strategies aim at eliminating the source of vibrations itself, i.e. to obtain a better compliance between the vibrations of the tool-part couple and the spindle frequency. To that purpose, the sinusoidal variation of the spindle speed [9] is a solution according to given cutting conditions. Multilevel random variation [10] functions may also provide a solution for a range of cutting conditions.

There are several particularly active areas of research :

It has been shown in the 3S laboratory of Grenoble (France) [13], with which the authors have collaborated for this work, that numerical modelling allows an increasingly better prediction of the dynamic effects of machining,

- Approaches based on real time compensation need to take the problem of regenerative vibrations into account by allowing sophisticated variations around the nominal cutting conditions [10].

- CAD-CAM systems can presently take some forced vibration models into account and also dynamic machine capacities [11]; they can also in some cases take into account regenerative vibrations. Finally, others processes such as high speed sawing by blade or by disc also suffer from regenerative type vibrations [12].

As a first step, we have applied the theory of stability lobes to the side milling of our flexible parts, the tool being considered as rigid.

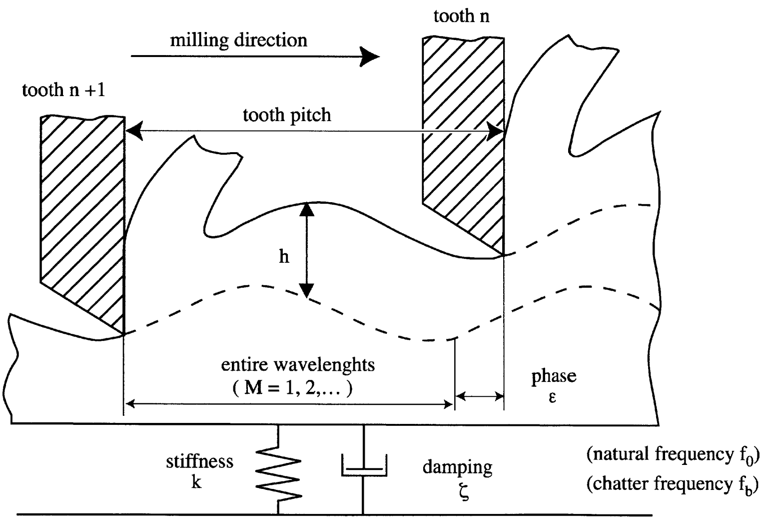


Figure -1. Cutting process with regenerative vibrations

We consider the milling process as defined in figure 1, consider a tool with *p* teeth, turning at *N* revolutions per minute with an advance by tooth of *h* and a width of cut *b*. Only effects related to the degree of freedom of the system are taken in account here. The part is characterized by its stiffness *k*, its damping ζ and its natural frequency ω_0.

We are first looking for the conditions in which the oscillations of tooth *n+1* have a larger or smaller amplitude in comparison with tooth *n*.

We considers that the cutting force is linked to the width of cut *b* and to the specific cut coefficient in the direction of vibration K_s as equation (1).

$$F_n = -\, b\, K_s\, (U_n - U_{n-1}) \tag{1}$$

In a permanent regime and in the linear area, a harmonic vibratory regime can be observed. In the complex domain, the displacement of tooth n depends on the cutting force and on the dynamic stiffness of the system :

$$U_n = Z_{(f_b)}\, F_n \tag{2}$$

The two previous equations allow to express :

$$U_{n+1} = (b\, K_s)\, (Z^{-1} + b\, K_s)^{-1}\, U_n \tag{3}$$

The condition of stability limit can be written as follows :

$$|\, U_{n+1}\, | = |\, U_n\, | \tag{4}$$

By taking account the previous equation, one can obtain :

$$\mathrm{Re}\, \{b_{\mathrm{limit}}\, K_s\, Z\} = -\, 0.5 \tag{5}$$

Let us consider now that the dynamic stiffness of the system can be written as :

$$Z^{-1}_{(f_b)} = k\, \left(1 - (f_b f_0^{-1})^2 + 2\, i\, \zeta\, f_b f_0^{-1}\right) \tag{6}$$

The axial depth at the limit of stability can then be expressed as :

$$b_{\mathrm{limit}} = -\left(k\, \left(1 - (f_b\, f_0^{-1})^2\right)^2 + 4\, \zeta^2\, (f_b\, f_0^{-1})^2\right) \left(2\, K_s\, \left(1 - (f_b\, f_0^{-1})^2\right)\right)^{-1} \tag{7}$$

We have now to express it according to the frequency of spindle *N*. The phase between the tooth frequency and the chatter frequency shows *M* entire chatter periods, and a remaining fraction denoted ε :

$$f_b\, (N\, p)^{-1} = M + \varepsilon\, (2\, \pi)^{-1} \tag{8}$$

The phase between passage of two consecutive teeth, can be directly expressed as follows :

$$\varepsilon = \mathrm{Arg}\, (U_{n+1}) - \mathrm{Arg}\, (U_n) = \pi - 2\, \mathrm{arcTan}\left[2\zeta\, f_b\, f_0^{-1}\, \left(1 - (f_b\, f_0^{-1})^2\right)^{-1}\right] \tag{9}$$

We can then deduce the link between the spindle frequency and the chatter frequency :

$$f_b\, (N\, p)^{-1} = M + (2\pi)^{-1}\, \left(\pi - 2\, \mathrm{arcTan}\left[2\zeta\, f_b\, f_0^{-1}\, \left(1 - (f_b\, f_0^{-1})^2\right)^{-1}\right]\right) \tag{10}$$

The determination of the chatter frequency according to the spindle speed from the previous equation and the previous expression of the depth limits allows us to obtain the stability lobes.

As a second step, we have used software developed in the 3S laboratory [13]. This software is dedicated to the time domain study of the vibratory behaviour of the milling system. The only phenomenon taken in account is that of self-maintained vibration. Only a macroscopic scale of modeling of the cutting process is considered and used for simulation. At each instant the tool-part interaction is calculated. An implicit time integration scheme is used to solve the nonlinear system. The mechanical behaviour of the whole part-machine system is modeled by truncated modal base whose coefficients have been determined by spectral and finite element analysis. The calculation of cutting forces uses two models to take into account vibratory aspects. In effective cutting, the model of [14] is applied; during back cutting, the model described in [15] is applied. The simulation is performed on side milling of flexible parts with a deformable tool, the part being considered as non deformable throughout the totality of the volume of milling, all solid motions being allowed. The main results provided by the software are the state of the manufactured surface and the vibrations of the part and of the tool. We have represented in figure 5 and figure 6 the results obtained on our experiment.

3. EXPERIMENTAL APPROACH

3.1 Experimental procedure

In order to check for consistency between our study and a given industrial case, we have chosen to manufacture a flexible part having a behaviour similar to a blade and allowing the milling of several rails. By using the configuration given in figure 2 and the technique described in [1], we can limit movements of the part to a single degree of freedom.

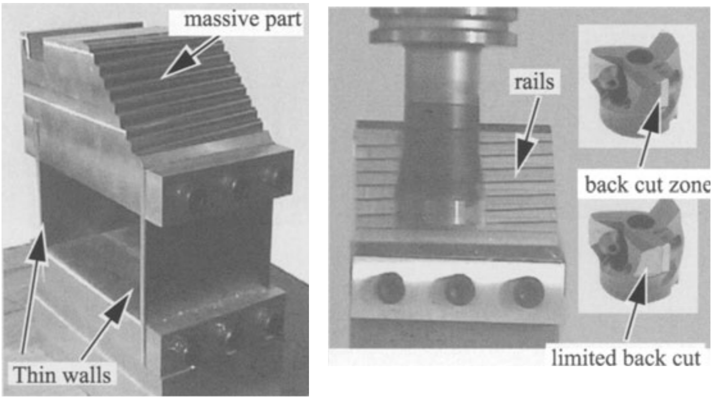


Figure -2. Tested parts, machined rails and tool drills

In order to experimentally cover the different zones of the stability lobes, we use a mill with a diameter of 40 mm (4 teeth, $\kappa = 90$ and $\gamma_p = 0$) considered as infinitely rigid in comparison to the part (see figure 2). The absence of a helical form allows us simplified modelling and does not require the utilisation of the different tools planes [3]. The maximum value of rotation of the tool is

8000 rev/min. The test part has been dimensioned to obtain a first natural frequency of around 180 Hz. The first natural frequency of the tool-spindle system is approximately 1700 Hz, which should not affect our results. The machined rail allows the variation of the depth of cut from 0 to 5.5 mm and allows the identification of the stability-instability transition, keeping the other cutting conditions constant (width of work $A_r = 2$ mm and advance $f_z = 0.15$ mm/tooth). Each rail corresponds to a different spindle speed. Tests are firstly performed with the tool in a configuration of a strong back cutting effect (initial tool), then in a limited back cutting configuration (figure 2), corresponding to a reduction of the friction surface of the tool, limited to the thickness of the tip.

3.2 Experimental results on the stability transition - instability

After performing the tests, machined surfaces which are function of the vibratory behaviour of the part are obtained. For some rails, two zones clearly appear: the stable area, where the surface roughness is consistent with the cutting process, and the unstable area where the surface can be of poor quality (see figure 3).

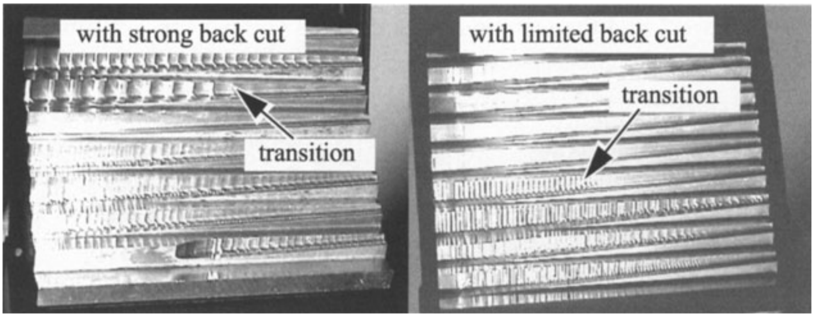


Figure -3. Stability-instability transition and influence on machined surface

In this case, we can notice defects of 2nd and 3rd order, randomly distributed or in consequence of a periodic alternation of irregular undulations, and an increase of the roughness due to changing cutting conditions. On the other hand, the case of milling with a strong back cut effect appears more problematical than the case with limited back cut. It seems actually that the back cut here provokes an increase of force transmitted to the part which result in a more immediate vibratory phenomenon with a large amplitude. We are more particularly interested in the depth of cut limits under which the machining process seems stable. Using a 3D roughness tester, we can identify this parameter and draw these measures on the theoretical layout of stability lobes (figure 4). For tests with limited back cut, the experimental interpretation is consistent with the theory of stability lobes. Generally speaking, the areas deduced from theoretical transitions have a good correspondence with the depths of cut seen during tests. Only the rail obtained at 12000 rev/min does not satisfy the layout of the lobes, which can be explained by cutting conditions not recommended by the tool manufacturer.

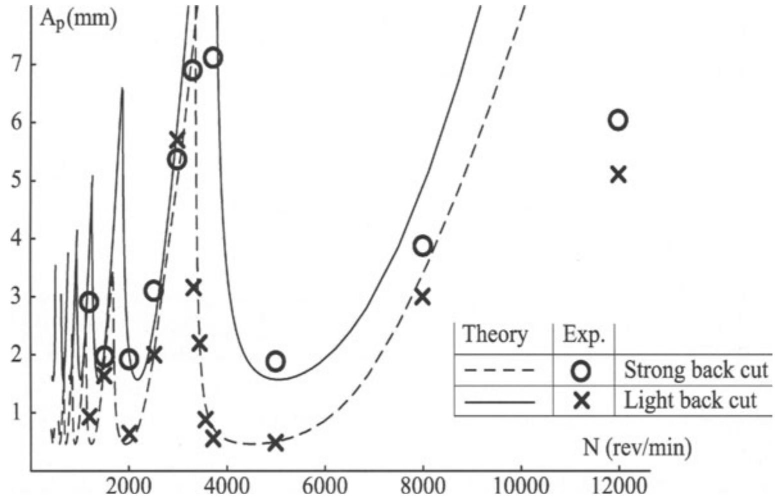


Figure -4. Theoretical and experimental cut depth limit

Tests with a strong back cut show a comparable dynamic behaviour but a modification of stability lobes parameters can still be noticed. In order to set our tests in correspondence with the layout of the lobes, it is necessary to modify the specific cutting coefficient K_s in the direction of the degree of freedom from 250 MPa to 415 MPa. The damping coefficient ζ during cutting is then around initially 0.02 instead of 0.04. This can be explained by an increase of the forces on the body of the tool as a consequence of the friction. The interpretation of a reduction of ζ is more difficult to provide. An accurate modelling of the back cut phenomenon will be required for validating the new values of these parameters. Similarly as in the case of limited back cut, the test with N = 12000 rev/min is difficult to interpret.

3.3 Comparison of theoretical and manufactured profiles

The examples of figure 5 and figure 6 shows a good consistency between the numerical simulation of the profile and its measurement on the roughness tester. The transition zone between stability and instability is identical, but the amplitude and the period of roughness defects vary from 10 % to 40 %. This imprecision is only observed in the unstable area and could be due to the simplification of the cutting and back cutting model. Globally, tests with high back cut confirm its bad influence on the surface finish, this configuration being naturally harmful.

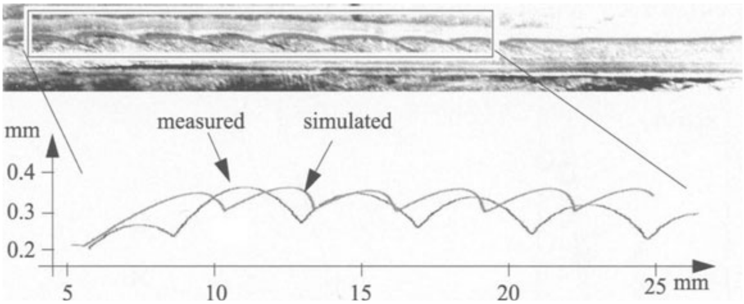


Figure -5. Simulated and measured profile for high back cut

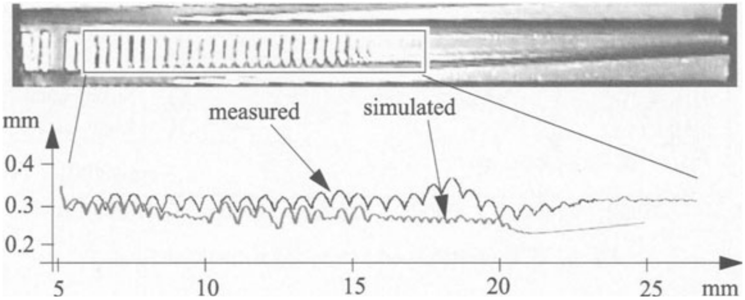


Figure -6. Simulated and measured profile for low back cut

The engaged length of the tool is much greater than the advance per tooth. It is therefore difficult to directly interpret the vibratory behaviour of the part from the manufactured profile since the generated surfaces are almost completely re-manufactured by the passage of the following teeth.

The criteria of quality of surface finish (typical R_a) hardly allows the validation of the stability of the milling process. We have chosen the occurrence of a visible jump on the profile for characterising the vibratory behaviour change. In this way, it is possible to determine rapidly the optimal cut depth on the machined rail.

4. PERSPECTIVES

The adaptation of the methodology to slim blades requires other phenomena to be taken into account. The effects of edges can no longer be neglected and it is therefore necessary to integrate the twist modes of the part in the stability lobes model. The tool, with a diameter of 16 mm, is notably adapted to end milling and is characterised by a small helix angle: the distribution of cutting forces is therefore different. On the other hand, its deformation is no longer negligible and must be taken into account in the model. The first tests on a blade with a thickness of 6 mm (figure 7) with an adequate tool shows consistent results. Further experiments are in progress to model 3D stability lobes in order to take into account the modification of the natural frequency of the part during the manufacturing process.

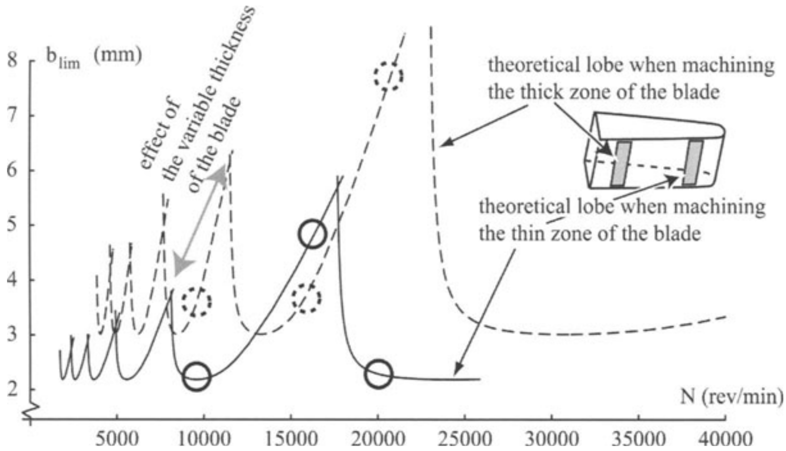


Figure -7. Typical behaviour of the 6 mm thick blade

In this, the experimental approach here based on rails machining can provide an interesting alternative to the complexity of the theoretical models due to the multiplicity of parameters. A global model using several vibration modes (part and tool) could guide an experimental procedure depending on the expected shape of the rails.

In the longer term, the integration of this methodology (numerical modelling and experimental refitting) in a CAD-CAM system could allow a better definition of the cutting conditions, on the basis of a depth of cut limited according to the geometry of the part.

5 CONCLUSION

Experimental results show a good correlation with calculations based on the theory of stability lobes in the case of rails produced by side milling on a flexible part. This technique could allow the very rapid and easy determination of the vibratory characteristics of a tool-part couple. The back cutting process is interpreted in the lobes model as an increase of the cutting forces, which provokes an earlier instability. The numerical simulation allows the quantification of the influence of several parameters and to guide more precise experimentation. The case of the blade makes obvious the limit of our initial modelling and shows the necessity to take into account new parameters. This approach will therefore have to evolve towards an integration of the methodology in CAD-CAM systems so that they will be able to determine strategy and cutting condition for an optimal manufacturing process.

REFERENCES

1. S.A. JENSEN, Y.C. SHIN, *"Stability analysis in face milling operations"*, Journal of Manufacturing Science and Engineering, N°121, 1999, pp. 600-614.
2. J. TLUSTY, F. ISMAIL, *"Basic non-linearity in machining chatter"*, Annals of the CIRP, N°30, 1981, pp. 299-304.
3. Y. ALTINTAS, E. BUDAK, *"Analytical prediction of stability lobes in milling"*, Annals of the CIRP, N°44, 1995, pp. 357-362.
4. H. FU, R. DEVOR, S. KAPPOR, *"The optimal design of tooth spacing in face milling via dynamic force model"*, Proc. of the 12th NAMRC, 1984, pp. 291-297.
5. T. DELIO, S. SMITH, J. TLUSTY, *"Use of Audio Signals chatter detection and control"*, Journal of Engineering for Industry, N°114, 1992, pp. 146-157.
6. S. SOLIMAN, F. ISMAIL, *"Chatter detection by monitoring spindle drive current"*, International Journal of Advanced Manufacturing Technology, N°13, 1997, pp. 27-34.
7. J. TLUSTY, M. ELBESTAWI, *"Analysis of transients in a adaptive control servomechanism for milling with constant force"*, Transaction of the ASME, 1976, N°76-WA/prod- 31.
8. S. SMITH, J. TLUSTY, *"Stabilizing chatter by automatic spindle speed regulation"*, Annals of the CIRP, N°41, 1992, pp. 433-436.
9. R. RADULESCU, S. KAPOOR, R. DEVOR, *"An investigation of variable spindle speed face milling for tool-work structures with complex dynamics"*, Journal of Engineering for Industry, N°119, 1987, pp. 266-280.
10. E. I. AL-REGIB, *"Machining systems stability analysis for chatter suppression and detection"*, Ph.D. Dissertation, 2000, University of Michigan (USA).
11. A. DUGAS, J.J. LEE, J.Y. HASCOËT, *"Feed rate and tracking errors simulation in high speed milling"*, Proc. of the 4th ESAFORM conf., 2001, pp. 667-670.
12. A.G. ULSOY, C.D. Jr MOTE, *"Vibration of wide band saw blades"*, Journal of Engineering for Industry, N°104, 1982, pp.71-78.
13. G. PEIGNÉ, H. PARIS, D. BRISSAUD, *"Simulation of non-linear vibrations and surface generation in flank milling : application to HSM"*, Proc. of the IDMME 2002, Clermond Ferrand, France, 2002, p.126
14. W.A. KLINE, R.E. DEVOR, J.R. LINBERG, *"The prediction of cutting forces in end milling with application to cornering cuts"*, International journal of machine tool, 1982, N°22/1, pp. 7-22.
15. D. W. WU, *"A new approach of formulating the transfer function for dynamic cutting process"*, Journal of Engineering for Industry, N°111, 1989, pp. 37-47.

ENERGY ASSESSMENT AND CUTTING TOOL BEHAVIOR IN MACHINING

R. Laheurte, P. Darnis and O. Cahuc
University of Bordeaux 1, France.

Abstract: To optimize the cutting conditions in machining, it is necessary to quantify the energy parameters involved in the process. Thus the control of a cutting process is realized if the cutting parameters, the kinematic parameters and contact actions are accurately estimated. To estimate these parameters, original metrological devices have been developed in the various laboratories involved in this study. Force and moment measurement has been carried out using a six component dynamometer [4]. Various image processing techniques allow to determine the chip/tool contact surface and the chip ejection angle. Complete and detailed energy equilibrium realized with those devices has clearly demonstrated the influence of moments on the mechanical power consumed by the cutting process. With these results, a three-dimensional thermomechanical cutting model has been developed. The originality of this model is the complete and continuous definition of the tool geometry, integrating tip ray and edge acuity characteristics. Forces and moments, chip kinematic parameters (chip orientation, tool/chip contact area) can be evaluated. The presented study shows the geometrical and kinematic parameters effects, such as the penetration rate on force and moment values in turning. A correlation is established between kinematic parameters, forces, moments and an original experimental study using an imagery interpretation method.

Key words: Turning process – Cutting process – Forces – Moments – Model.

1. INTRODUCTION

The performance improvement of modern cutting tools allows their use for finishing operations with respect to surface and dimensional qualities, in particular for circular parts. Previously only grinding processes were used in this type of operation because of the material hardness. For finishing and

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 281–290.

hard turning operations, the cutting process is carried out in the tip ray because of depth of cut.

The majority of analytical three-dimensional cutting models characterized the tool with three planes defined by the cutting face, the principal and the secondary clearance zones. Such models cannot be used for finishing operations and hard turning since it does not take into account the different tool acuities (tip ray and edge acuity, *Figure -1*).

In this study, a new energetic three-dimensional cutting model is presented. This work is based on previous results [3], [6], [7], [8], where the primary and secondary shear zones are described and analyzed. The originality of this model is a complete definition of the tool geometry.

2. TURNING CONTACT CHARACTERISTIC

2.1 Turning tool geometrical description

The established geometrical model describes the cutting edge on a mesoscopic scale. The contact between the tool and the workpiece is considered as a continuous surface. These zones (*Figure -1*) include the cutting face, the principal and secondary clearance faces. The tip ray r_ε and the edge acuity R of the tool realize the joint of these planes. This approach is in accordance with the shape of modern cutting tools, which generally display a large tip ray and allows complete definition of the workpiece/tool/chip contact.

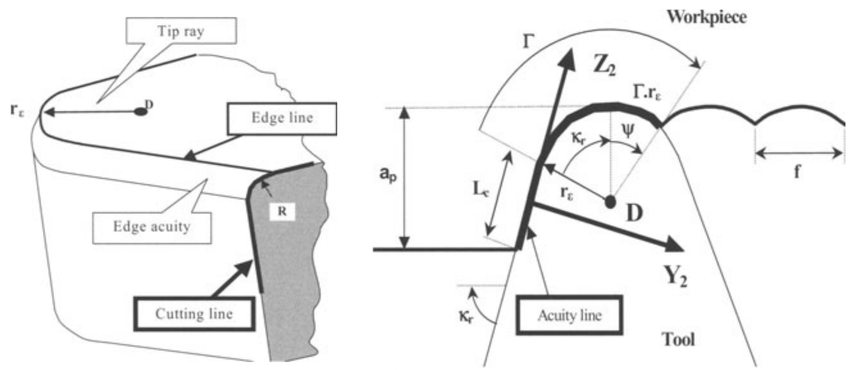


Figure -1. Geometrical description of tool and workpiece/tool contact.

2.2 Turning contact geometry description

The three-dimensional contact surface, between the tool and the workpiece, is built from a path line (or acuity line) and a sketch line (or cutting line), (*Figure -1*).

2.2.1 Acuity line definition

The workpiece/tool contact surface is partly defined by the acuity line. A straight zone (L_c *length*) and a circular zone ($r_\varepsilon.\Gamma$ *length*) compose the acuity line (*Figure -1*). Dimensions of these zones are function of tool geometrical parameters such as edge angle κ_r and tip ray r_ε and cutting kinematic parameters such as cutting depth a_p and feed rate f.

Thus, the contact length of the rectilinear part L_c is expressed by:

$$L_c = \left(a_p + r_\varepsilon(\cos(\kappa_r) - 1)\right) / \sin(\kappa_r) \tag{1}$$

The contact length in the circular zone is function of the angle Γ:

$$\Gamma = \kappa_r + \Psi \tag{2}$$

where Ψ is the complementary contact angle on the secondary clearance face (*Figure -1*):

$$\Psi = \arccos\left(\left(\sqrt{r_\varepsilon^2 - \frac{f^2}{4}}\right) \Big/ r_\varepsilon\right) \tag{3}$$

The whole contact length L_{CT} is then expressed by:

$$L_{CT} = L_c + r_\varepsilon \times \Gamma \tag{4}$$

2.2.1.1 Cutting line definition

The cutting line completes the definition of the workpiece/tool contact surface. The cutting line is bordered by three zones: the *BO* zone (cutting face), the *OJ* zone (edge acuity) and the *JK* zone (clearance face) (*Figure -2*). The length *Y* between *B* and *O* geometrically defines the cutting line and then by the length *L* between points *J* and *K*. Connection between these two lines is determined by the edge acuity *R* (*Figure -2*).

Length *Y* cannot be directly defined and will be evaluated with equilibrium considerations. The workpiece/tool contact length *L*, according to the principal clearance zone, is estimated from the hydrostatic pressure P_0, the workpiece elastic modulus E, the cutting velocity Vc, the feed velocity Vf and the edge acuity R:

$$L = \left(K \times P_O \times \left(1 + V_f\right) \times \left(-R + 1\right)\right) / \left(E \times \left(1 + V_c\right)\right) \tag{5}$$

where, K is a corrective parameter resulting from Albrecht works [1].

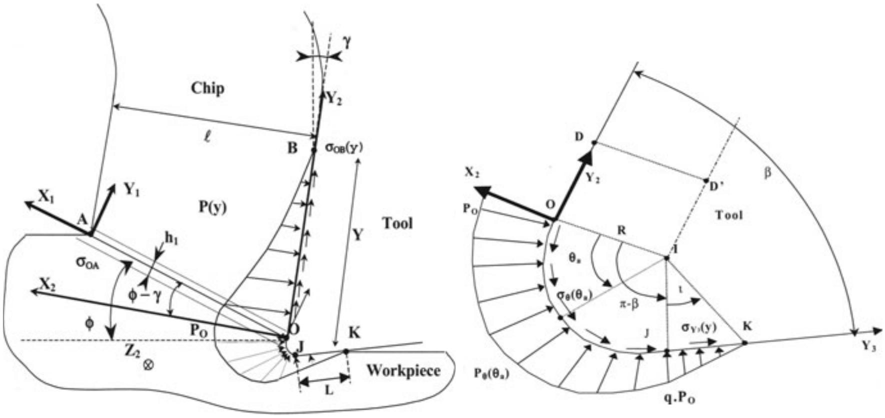


Figure -2. Description of the cutting zone, pressure and stress distribution.

3. CUTTING MODEL

The cutting model is established from equilibrium considerations. The elasto-visco-plastic behavior of the material is taken into account by a power type law. This law integrates coupling between stresses and thermal phenomena [2]. The pressure distribution in the primary shear zone and the kinematic parameters of the chip are based on previous works [2], [3], [4].

The following sections deal with the pressure and stress development in the secondary shear zone and on the workpiece/tool contact surface.

3.1 Stress and pressure distribution definition

Four parameters ($\delta.l$, Φ, Y, h_1) are identified using the geometrical description of the problem. These parameters are respectively the secondary shear zone thickness, the primary shear angle, the chip/tool contact length and the primary shear zone thickness, [2], [3], [5] (*Figure -2*).

Three zones characterize the cutting line. For each zone, a pressure and stress field is expressed. On the cutting face and relative to the secondary shear zone (*zone OB*, *Figure -2*), the pressure and stress distribution is based on Dudzinski and Molinari works [9].

$$P_{OB}(y) = P_O\left(1-\frac{y}{Y}\right)^{\eta} \; ; \; \sigma_{OB}(y') = \sigma_O\left(1-\frac{y}{Y}\right)^{\eta} \tag{6}$$

where η is a distribution parameter.

In order to establish the model, the following assumptions are made [2]: i) pressure and stress repartition has an angular decreasing evolution on the edge acuity (*OJ*, *Figure -2*). ii) Pressure and stress distribution has a linear decreasing evolution on the principal clearance face (*JK*, *Figure -2*).

According to these hypotheses, pressure and stress distribution on the edge acuity (*OJ zone*), are expressed by:

$$P_{OJ}\left(\theta_a\right) = P_0 \times \left(1 - q\left(\frac{\theta_a}{\pi - \beta}\right)\right) \; ; \; \sigma_{OJ}\left(\theta_a\right) = \sigma_O \times \left(1 - q\left(\frac{\theta_a}{\pi - \beta}\right)\right) \tag{7}$$

and on the principal clearance face (*JK zone*):

$$P_{JK}\left(y\right) = P_0 \times q \times \left(1 - \left(y/L\right)\right) \; ; \; \sigma_{JK}\left(y\right) = \sigma_O \times q \times \left(1 - \left(y/L\right)\right) \tag{8}$$

The angular repartition parameter q is defined by:

$$q = \iota/\left(\pi - \beta + \iota\right) \tag{9}$$

where β is the tool tip angle and ι is evaluated by:

$$\iota = arctan\left(L/R\right) \tag{10}$$

3.2 Chip kinematic considerations

Dependence between the chip kinematic and the machining parameters has been experimentally observed. At the cutting face, the chip orientation (Ξ) can be defined using the orientation average of the two normal contact directions, weighted by the contact lengths (*Figure -3*). The chip orientation angle (Ξ), corresponding to the chip ejection angle, is defined by:

$$\Xi = \left(\left(\Gamma.R_b\right)/L_{CT}\right)\left(\frac{1}{\Gamma}\int_0^{\Gamma} \zeta \, d\zeta\right) \tag{11}$$

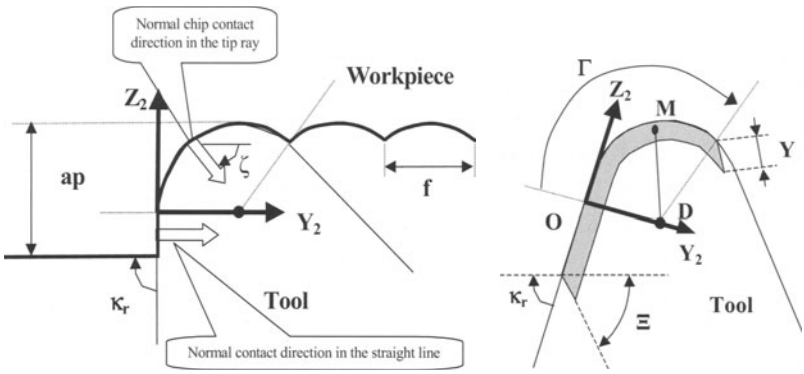


Figure -3. Chip kinematics and contact orientation on the cutting face.

3.3 Equilibrium problem resolution

The equilibrium problem resolution consists in evaluating the unknown parameters using two stages. Parameters (h_1, Y) and thermal variables are first determined using the chip equilibrium. Secondly, forces and moments are calculated out from the tool equilibrium. Then, the cutting energetic assessment can be expressed.

4. EXPERIMENTAL APPROACH

Some turning experiments have been realized to validate the proposed model. The test variables were selected in order to study the influence of the penetration rate (ap/r_ε) on the measured parameters. Force and moment measurement has been carried out using a six component dynamometer [4].

4.1 Experimental conditions

The AISI 4140 (*42CD4*) steel has been used for the experiments. Its characteristics are summarized in *Table -1*.

Table -1. AISI 4140 thermomechanical characteristics.

Thermophysic parameters	Behavior law parameters
Density: $7800 kg/m^3$	Viscosity parameter: $m=5,5.10^{-3}$
Fusion temperature: 1227 °K	Work hardening parameter: n=0,0563
Thermal conductivity: $K=4,6.10^{-2}$ °K/m	$B=-7,9.10^{-4}$

Thermophysic parameters	Behavior law parameters
Specific heat: Cp=379J/kg°K	A=1,288
Stress failure: σ_r=900MPa	
Elastic modulus: E=210GPa	
Hardness: 260 Hv	

Uncovering carbide inserts have been used. These inserts have no chip breaker (γ_c=*-6°; α=-6°*).

4.2 Experimental contact approach

To describe the orientation of the chip ejection, some video acquisitions have been carried out during the experiments. Correlation between those videos and the microscopic analyses of the tool face (*Figure -4*) is observed. The microscopic images of the tool face are obtained using an image processing method (*Figure -5*), which allows to determine the chip orientation and the contact area.

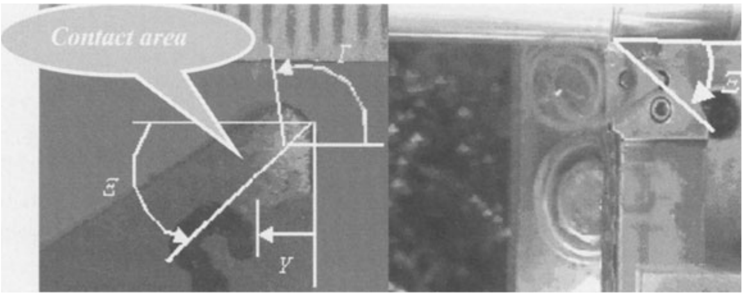


Figure -4. Geometrical contact determination.

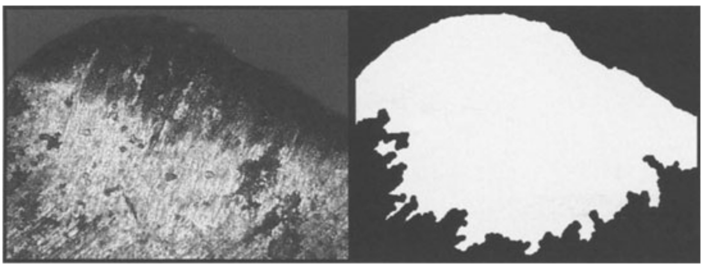


Figure -5. Contact surface interpretation by image processing method.

4.3 Results and discussion

4.3.1 Chip kinematics

It can be inferred from results (*Figure -6*) a non-linear evolution of the contact length with the penetration rate. This non-linear evolution demonstrates that the tip ray has a great influence on the cutting process, essentially for the finishing operation.

For the contact area, the curves presented *Figure -6* show a nonlinear evolution when the tool is engaged in the tip ray and linear when the penetration rate is higher than 1 (for the experimental conditions used).

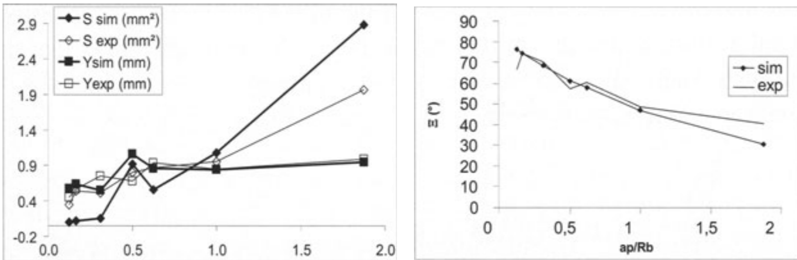


Figure -6. Contact area (*S*) evolution, Chip orientation (Ξ) evolution.

The chip orientation (as for the length of contact) is inflecting when the penetration rate exceeds 1. *Figure -6* show that a simple geometrical model of the tool active part led to a good chip/tool contact parameter prediction.

4.3.2 Cutting forces

The cutting actions are expressed in the frame of the *Figure -7*. The moments are determined at the tool tip (point *D'*).

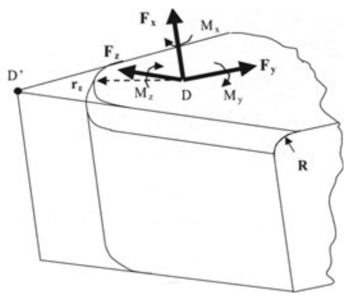


Figure -7. Force and moment framework.

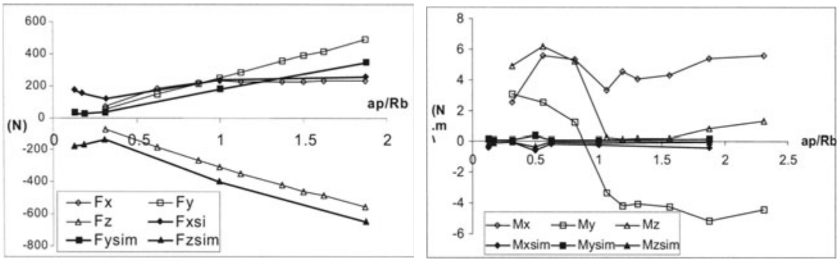


Figure -8. Forces and moments evolution at the tool tip.

The force F_x has a non-linear evolution (*Figure -8*) with the tip ray. On the other hand, the cutting force F_z and the feed force F_y are not dependent to the tip ray.

The moment analysis shows that the penetration rate has a great influence on M_y moment. The variation of M_y moment has a direct influence on the power consumed. Some visual observations show that a relationship exists between the chip rolling up and the moment vector orientation. Nevertheless, it is necessary to confirm these observations with an experimental study involving both video acquisition and the cutting action measurement. The complete description of the workpiece/tool/chip contact gives a good evaluation of the tool behavior in the tip ray zone. Consequently, the force evaluation has been improved. However, the moment evaluation at the tip tool is not correlated by experiments. It is necessary to develop a new approach that takes into account a specific behavior law which itself takes into account the extreme solicitation involved during a cutting process.

5. CONCLUSION

A 3D cutting model has been developed. It can predict the chip kinematic and its ejection orientation on the cutting face in turning. It also evaluates three forces and three moments components and therefore the total power consumed. This model takes into account the tool geometry. This approach is completely fit to the hard turning process and finishing operations. The tool/chip contact surface is evaluated with a micrographic observation coupled to an image processing method. Actually, this method is a new concept and gives good results.

The moments are influenced by the penetration rate. There is a direct relationship between the power consumed and the moment values. Similarly, the direction of these moments influences the formation of the chip.

To evaluate correctly the moments, this model must be improved with a better analysis of the mechanics using for example a micropolar theory [3].

The tool geometry model must consider all the actual tools geometrical evolutions. The new tool shapes are designed for the finishing operation such as Wipper and Honing.

REFERENCES.

1 Albrecht P. "New developments in the theory of metal cutting process. Part I : The ploughing process in metal cutting", Journal of Engineering for Industry, (ASME), 1960, Vol.82-1, pp. 348-358.

2 CAHUC O, DARNIS PH. GÉRARD A; BATTAGLIA J.L. *"Experimental and analytical balance sheet in turning applications"*, The International Journal of Advanced Manufacturing Technology, N°1028, Vol 18, N°9, pp648-656, 2001.

3 CAHUC O, TOULOUSE D, COUETARD Y, GERARD A. *"Un nouveau modèle de la zone de contact outil copeau"*, C.R. Acad. Sci., Paris, France, tome 328, Série II b, 2000, pp.1-4.

4 COUETARD Y. "Caractérisation et étalonnage des dynamomètres à six composantes pour torseurs associés à un système de forces." Thèse de doctorat de l'Université Bordeaux 1, 2000.

5 DARNIS PH, CAHUC O, COUETARD Y. *"Energy balance with mechanical actions measurement during a turning process."* Second International Seminar on Improving Machine Tool Performance. Nantes - La Baule. 2000. Session B1 p.1-12.

6 GILORMINI P. *"Contribution à la modélisation de la formation du copeau en usinage des métaux."* Thèse à l'Ecole Nationale Supérieure des Mines Paris, France, 1982.

7 MERCHANT E. *"Mechanics of the metal cutting process. I. Orthogonal cutting and a type 2 chip."* Journal of Applied Physics, May 1945.Vol. 16 p.267-275.

8 MERCHANT E. *"Mechanics of the metal cutting process. II. Plasticity conditions in orthogonal cutting."* Journal of Applied Physics, June 1945. Vol. 16. p.318-324.

9 MOUFSKI A, MOLINARI A, DUDZINSKI. *"Modelling of ortogonal cutting with a temperature dependent friction law."* J. Mech. Phys. Solids, 1998. Vol. 46. No. 10. p. 2103-2138.

A STUDY AND ANALYSIS OF MACHINE TOOL ERRORS WITH STATIC AND DYNAMIC MEASUREMENTS

M. Rahman, J. Heikkala and K. Lappalainen
Production Technology Laboratory, Department of Mechanical Engineering,, University of Oulu, Finland

Abstract: For higher productivity and product quality, it is necessary to know the accuracy of machine tools so that defective parts are not manufactured. Angular and positioning errors are quite common problems in machine tools. The origins of these errors are kinematic parameter deviations resulting from manufacturing errors, assembly errors or quasistatic errors. By static measurements such as laser interferometer, linear comparator (Heidenhain VM182), electronic inclinometers (Wyler Leveltronic) and dynamic measurements (Double Ball Bar, DBB) different kinds of error origin information of the machine tools can be obtained. It is possible to convert the measurement results of one method to another with certain degree of accuracy. This helps to justify the correctness of the measurement method and locate the error origins more precisely. This study shows a mapping between static and dynamic measurements of machine tools by calculating tool tip errors using different devices as mentioned above. For static measurements the laser interferometer, VM 182 and Leveltronic measuring systems and for dynamic measurements the DBB has been used in this study. Theoretical trace for DBB is obtained based on roll, pitch, yaws and positioning errors found by laser, VM182 and Leveltronic measurements. These traces are compared and simulated with the trace obtained by DBB.

Key words: multi-axis machine tools, accuracy, error origins, measurements

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 291–300.

1. INTRODUCTION

The positioning accuracy of a multi-axis machine tool depends on different kinds of error origins depending on the machine tool's structure, control system etc. In a typical machine tool, there are multiple error origins including geometric, static and dynamic loading, thermal, mismatches between servo-loop parameters etc. All these error origins affect positioning accuracy at the tool tip in a complex way.

Substantial work has been performed in the past on the development of error models for geometric, thermal etc. error sources for multi-axis machines. K. Kim and M. K. Kim have developed a volumetric error model based on 4*4 homogenous transformation for generalized geometric error [5]. K. F. Eman and B. T. Wu have developed error model accounts for errors due to inaccuracies in the geometry and mutual relationships of the machine's structural elements as well as errors resulting from the relative motion between these elements [3]. Kakino, Ihara and Shinohara have measured the positioning errors of multi-axis machine tools in a volumetric sense by Double Ball Bar (DBB) [4]. Real time volumetric error has been measured and compensated by Chen [1]. The theory behind the machine tool measurement for a particular measuring system can be used to justify the results of measurements of the same machine using different measuring systems. In this paper two types of machine tools measurement systems are described and compared. Positioning accuracy has been calculated for different measuring systems and a theoretical formula has been derived which can be used to justify the other measuring system.

2. MACHINE TOOLS AND MACHINING ACCURACY

Nowadays the NC machine tool is the most basic manufacturing equipment for mechanical part production. It has been widely used for various purposes such as for flexible automation, to improve machining accuracy, to reduce lead-time, to cut costs etc. The most desirable improvement is the ability to achieve high precision machining. The primary factor that influences machining accuracy is the motion accuracy of the machine tool. When there is a motion error in a machine tool, it will be transferred to the machined profile and thus increase the profile error of the machined surfaces.

Since the machine accuracy is the major source of errors, the control of machine error sources is critically important. Among the sources of machine errors, thermally induced errors and geometrically induced errors are known to be key contributors. Thermal errors come from thermal distortions of machine components due to external and internal heat sources, bearings, hydraulic systems, sunshine etc. The thermal bending of C-shaped machine frames, thermal expansion of lead screws and spindle thermal growth are typically the dominant error sources. Typical magnitude of the spindle growth and lead screw expansion could be more than one hundred micrometers as the machine warms up from a cold-start.

The geometric errors of machine tools arise from manufacturing defects, machine wear and static deflection of machine components. Geometric errors are especially significant with medium size and large size machine tools for which a highly rigid machine structure is difficult to achieve. Geometric errors are created not only during manufacturing but also from improper assembly and installation. Furthermore, environmental and operational conditions that are not under the vendor's control also affect machine performance.

2.1 Users of machine tools Inspection

Machine tool builders use the results of machine accuracy inspection for the purpose of making design improvements to increase accuracy. Such measurements also help them to optimize the commissioning parameters of the control loop whenever they influence the accuracy of a CNC machine. Machine tool users can use these measuring systems for acceptance testing and regular inspection of their machines. They can use the measurement results to improve their machines' performance in different ways.

3. MEASUREMENTS OF MACHINE TOOLS

Machine tool performance from the point of view of compliance to tolerances, surface definition, etc, is determined essentially by the dynamic and static accuracy of machine movement. For precision machining it is therefore important to measure and compensate motional deviations. Guidelines and standards for inspecting machine tools (ISO 230-2 and ISO/DIS 230-4, and VDI/DGQ Directive 3441) stipulate a number of measurement methods for determining dynamic and static deviations [2]. An example of a machine tools' measuring system has been described by Wyler in reference [6].

3.1 Static measurements

This measurement is such that position deviations are measured in the linear axes using a comparator system that produces conclusions based exclusively on the geometric accuracy of the machine. Conventional inspection and acceptance testing of machine tools has been limited essentially to static measurement of the geometrical machine structure without load. If X_{real} is the real position and X_{ideal} is the ideal position, then the positioning error is defined by the following equation:

$$E_{xn} = X_{real} - X_{ideal} \tag{1}$$

where n varies from 0 to maximum measuring points.

3.1.1 Measurement by laser interferometer

The laser interferometer uses light interference principles to precisely measure the linear displacement or velocity of a body. By laser it is possible to measure positioning, pitch and yaw error of multi-axis machine tools. Roll error may be obtained by electronic level.

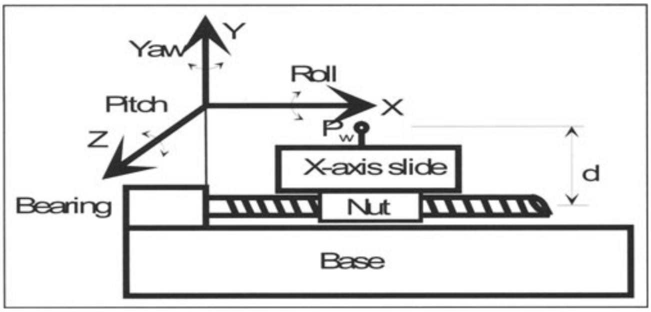


Figure -1. Example of roll, pitch and yaw (RPY) of X-axis

If the roll, pitch, yaw and positioning errors of a particular axis are known (in this case Z-axis) the tool tip position A can be predicted by the following equation:

$$A_{orealZ} =$$

$$\begin{bmatrix} 1 & 0 & 0 & Z\cdot k_{zx} \\ 0 & 1 & 0 & Z\cdot k_{zy} \\ 0 & 0 & 1 & (Z+\Delta Z) \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & k_{zy} & 0 \\ 0 & 1 & -k_{zx} & 0 \\ -k_{zy} & k_{zx} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -c_z & b_z & 0 \\ c_z & 1 & -a_z & 0 \\ -b_z & a_z & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & d \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where Z is the nominal movement, k_{zx} and k_{zy} are direction cosines of Z-axis, c_z, a_z and b_z are roll, pitch and yaw respectively. ΔZ is the positioning error.

3.1.2 VM 182 Measurements

The VM 182 linear comparator system from Heidenhain is used for acceptance testing, inspection and calibration of machine tools and measuring equipment with traverse range up to 1520 mm. Machine tool builders and distributors can determine the linear and non-linear curves as well as the reversal errors of machine axes according to ISO 230-2. In addition to determining the positioning errors, the VM 182 also measures guideway errors perpendicular to the traverse direction of the machine axes. The VM 182 consists of a steel scale with a highly accurate graduation and a scanning head that moves over the graduation without mechanical contact. For example, for Y-axis movement the following equation expresses the real position:

$$OP_{spindle} = \begin{bmatrix} 1 & 0 & 0 & \Delta X_{yx} \\ 0 & 1 & 0 & Y + \Delta Y_{yy} \\ 0 & 0 & 1 & \Delta Z_{yz} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

ΔX_{ij} is the positioning error in direction *j* during feed direction *i*. If we know all the error movements in all other directions it is possible to obtain the volumetric error for any point in the working space by the following equation:

$$E_{vol} = \begin{bmatrix} \Delta X_{yx} + \Delta X_{xx} + \Delta X_{zx} \\ \Delta Y_{yy} + \Delta Y_{xy} + \Delta Y_{zy} \\ \Delta Z_{yz} + \Delta Z_{xz} + \Delta Z_{zz} \end{bmatrix} \tag{4}$$

3.1.3 Leveltronic measuring systems

When measuring with inclinometers the electronic levels are used to measure the RPY of an axis with some interval. To learn about the volumetric errors of a machine tool we can interpret the results. In our modeling we have used a matrix to calculate the volumetric errors.

The level is moved along the machine tool axis with a suitable interval. The measured data is recorded to PC via RS232 cable for processing. An inclination to the right is positive and declination to the right is negative. To be sure about the results the measurement can be repeated with the same interval in different locations along the Y-axis. If the result is the same then it confirms that the error is pitch error of the X-axis.

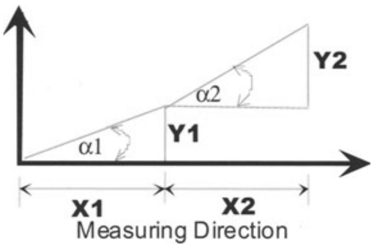


Figure -2. Principle of measurement with inclinometer

According to figure 2, the data received at points 1 and 2 are shown in the following table. Y is expressed in μm and X has been expressed in meters.

Point 1	Y_1/X_1 μm/m
Point 2	Y_2/X_2 μm/m and so on

By measuring the guideways in parallel lines the whole working surface of the machine tool can be accessed. The spindle deviation can be calculated from figures 1 and 2.

$$PW_{real1} = \begin{bmatrix} 1 & 0 & \dfrac{-y1}{x1} & \dfrac{-y1}{x1} * d \\ 0 & 1 & 0 & d \\ \dfrac{y1}{x1} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

The real locations of all other points are calculated from corresponding errors. Individual errors at each location are calculated by equation 5.

3.2 Dynamic measurements

Dynamic measurements, especially at high traverse speeds, provide information on contouring behavior that permits conclusions both on the conditions of the machine tool and on the parameter settings of the control loop consisting of CNC control, drives and position feedback systems. The accuracy of machining depends on dynamic deviations from the nominal contour and also on high acceleration in the machine tool.

3.2.1 Double Ball Bar (DBB) method

When a machine is commanded to make circular interpolation with a radius R the increase in radius is given by [6],

$$\Delta R_n = \frac{(C_{xn} \cdot X_n + C_{yn} \cdot Y_n + C_{zn} \cdot Z_n)}{R} \tag{6}$$

where $X_n = R \cdot \cos(\theta_n)$ and $Y_n = R \cdot \sin(\theta_n)$ and $C_{xn} = E_{xn} - E_{xo}$ $E_{xn} = X_{real} - X_{ideal}$ on the control point $E_{xo} = X_{real} - X_{ideal}$ on the center of circular interpolation. Values of C_{xn}, C_{yn} and C_{zn} are obtained for n points on the circular arc based on equation (6). This equation gives ΔR for all *n* points.

For all *n* points it is possible to find a positioning error that affects the contouring accuracy of the circular interpolation with dynamic error such as servo response. It affects by increasing or decreasing the radius based on equation (6). This contouring error can be compensated by arc replacement.

4. CONVERSION OF MEASUREMENTS

This section describes how to obtain or predict dynamic measurement results from static measurements. It has been mentioned in the definition of static measurement that no contouring error affects the static measurement because of single axis movement. In this discussion we do not consider the servo control system for the sake of simplicity. For exact examination the whole servo system, heat sources etc. should be considered.

4.1 Static to dynamic measurements

When we have the results of all static measurements of a multi-axis machine we can calculate the tool tip position deviations for the whole working space. For example, with the measurements of roll, pitch, yaw and positioning errors of all axes, it is possible to find the error values of that axis at any point by interpolation.

$$E_{des\_pt} = E_i + \frac{E_{i+1} - E_i}{\Delta} \times F \tag{7}$$

where E_{des_pt} is the error at the desired point and F is mantissa.

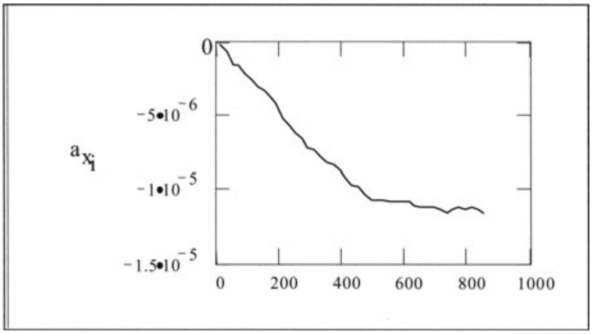


Figure -3. Roll of X-axis (in radians)

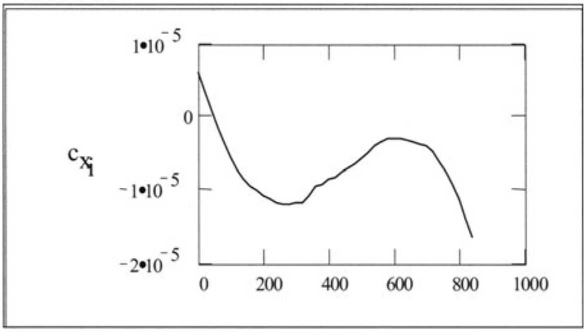


Figure -4. Pitch of X-axis

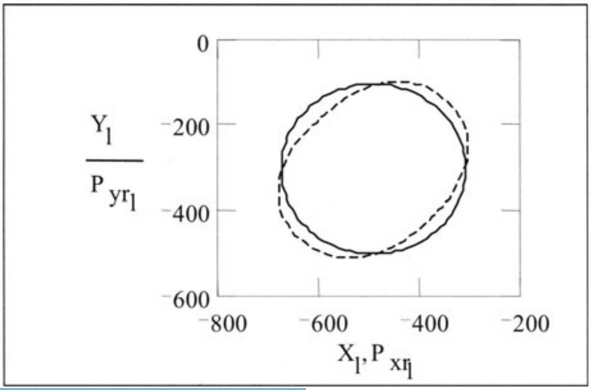


Figure -5. The DBB error trace obtained from static measurement results

4.2 Dynamic to static measurements

When the ΔR is provided from the DBB measurement data there are several error components involved in the result. The error components and error origins could be separated from each other by least-squares technique.

Let ΔR_m be the measured error and ΔR_t the theoretical volumetric error of a machine tool. The latter indicates the volumetric error that has been developed for DBB (fig. 5) [4].

$$\Delta R_t = a_1 + a_2 \cdot x + a_3 \cdot x^2 + \cdots\cdots + a_r \cdot x^{r-1} = \sum_{j=1}^{r} a_j \cdot x^{j-1} \tag{8}$$

$$v_i = \sum_{j=1}^{r} a_j x_i^{j-1} - \Delta R_m \tag{9}$$

$$S = \sum_{i=1}^{n} v_i^2 \tag{10}$$

setting $\dfrac{\delta S}{\delta a_k} = 0$ where k = 1, 2, ... r

After simplifying those equations we obtain a set of *r* linear equations, therefore we are able to find *r* number of error parameters.

$$\sum_{j=1}^{r} \left(\sum_{i=1}^{n} x_i^{j+k-2} \right) \cdot a_j = \sum_{i=1}^{n} x_i^{k-1} \cdot \Delta R_m \qquad k = 1, 2, \ldots, r \tag{11}$$

for $a_1, a_2, a_3, \cdots, a_r$

When we have all the error parameters, we are able to predict the measurement result of static measurement by an individual measuring system.

4.2.1 DBB to Leveltronic measuring system

For simplicity we only show how to obtain the Leveltronic measuring results from DBB results. From equation (10) we have individual error parameters that affect the volumetric error of the machine under investigation. Equation (10) actually gives the X component of P_{wrealX}. Now the error at any point is given by the following equation:

$$\frac{Yn}{Xn} = \frac{P_{WrealX}}{d} \tag{12}$$

where n varies from 0 to a maximum number of measuring points.

If we plot the error against the measuring axis, we obtain the trace as obtained by the Leveltronic measuring system.

CONCLUSION

This paper has described the theory of how to obtain the dynamic error profile from static measurements of multi-axis machine tools and vice versa. From these two types of measurements and comparison we are able to investigate machines' error origins and more accurately justify the measurement results. Once we know the error origins precisely, we are able to position the tool tip at the desired position. (Tool direction may not be possible to fix because of the limitation of degree of freedom of machine tools.) In conclusion the steps for measure machine tools and justify the measurement results are summarized as follows:

1. Modeling of a machine tool for error diagnosis for different measuring systems.
2. Measuring the machine tool.
3. Diagnosing the error origins.
4. Comparison of the measurement results obtained by other methods.
5. Corrective actions to improve motion accuracy.

REFERENCES

1. Chen, Jenq-Shyong, Real-time compensation of time-variant volumetric error on a machining center. The University of Michigan, 1991.

2. Heidenhain brochure, Measuring Systems for Machine Tool inspection and acceptance testing, September 1997.

3. K. F. Eman, B. T. Wu, A generalized error model for multi-axis machines, Annals of the CIRP, 36 (1), 253-256 (1987).

4. Kakino, Ihara, Shinohara, Accuracy inspection of NC machine tools by Double Ball Bar method. Hanser publishers, Munich, Germany 191(1993).

5. K. Kim and M. K. Kim, Volumetric accuracy analysis based on generalized geometric error model in multi-axis machine tools. Mech. Mach. Theory, 26(2), 207-219 (1991).

6. WYLER AG, Operator's manual for LEVELTRONIC NT, (1999)

EVALUATION OF MACHINE GEOMETRY

Presentation of an innovative method with a Laser Tracker

S. Leleu, J-M. David, G. Diolez and P. Martin*
L2MA, ENSAM, Lille, France
** LGIPM, ENSAM, Metz, France*

Abstract: The geometric error measurement techniques for machine tools have not evolved for several years. The laser tracker, a three-dimensional portable measuring system, opens the way to automatic measurements and consequently, the possibility of positioning error compensation under acceptable economic conditions. We propose strategies of measurement and treatment to improve and control the measurement uncertainty of a laser tracker used as a system of calibration within the framework of the analytical model of global compensation by grid.

Key words: Coordinate Machine, Geometrical errors, Laser Tracker, Metrology

1. INTRODUCTION

The conquest of precision is a significant condition in improving the performance of the machines and the apparatus we use every day. One of the main difficulties is to integrate phases of measurement and correction, as well as during the assembly of the machine as during its use, with acceptable time and cost conditions. We propose the use of a new tool for geometrical qualification, the laser tracker. The ease of use allows the integration of the laser tracker during all the life stages of the machine. It is however necessary to reduce the influence of its high intrinsic uncertainty to investigate ways opened up by the direct access to the three-dimensional tool's end position.

Coordinate machines, like machine tools and coordinate measuring machines (CMM), can be analyzed as systems for positioning a tool (or a probe) in a space linked to the workpiece. The objectives of any geometric

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 301–312.

control may be either the verification of the machine with respect to the specifications or the standards in force, or its calibration and correction. Systems of measurement specially adapted to these objectives must be implemented. In the first case, the machine's global capacity to fulfill a function satisfactorily must be studied, whereas in the second case, errors must be identified in order to be corrected.

In practice, two different types of intervention are open to machine users:

- Practical tests which give the user information on the machine's average level of quality by following a trajectory in the workpiece space.
- Analytical tests which measure the characteristics of each of the machine's axes individually according to its geometric model.

It is generally rather difficult to take corrective action from a practical test (e.g. : NASA, Ball Bar, ...) [1] whereas with analytical methods errors can be corrected by manual adjustment or by software compensation. The latter, developed by CMM manufacturers in the 80's, based on analytical modeling of the machine where the positioning error of the tool with respect to the workpiece is the sum of the displacement errors of the machine's constitutive axes [2], [3], [4], [5]. Expressing the errors in this form means coding a reduced volume of data in proportion to the sum of the lengths of the axes and the sampling step chosen.

However, the classic measuring techniques used in this approach remain delicate and often have to be carried out by experienced technicians. This frequently involves very long measuring campaigns.

2. THE GEOMETRIC CONTROL OF A MACHINE

The control of a machine's geometry means a model of the errors has to be installed so that they can be identified, corrected or their variation followed. As far as serial machines are concerned, we will retain the analytical model for which each axis of the machine is analyzed separately then positioned in the machine space. We will only present, in this article, the case of a machine composed of 3 axes for which there is a bijection between the Cartesian coordinates of the tool and the coordinates of the machine's axes. The analytical model can be applied to a machine with more than 3 axes if a sufficient set of coefficients is introduced.

2.1 The analytical model of a 3-axis machine

Subject to the undeformability of the machine's constitutive solids, the errors of each axis, whether translation or rotation, are described by a field of displacement of solid bodies. The displacement error retained is the difference between the position actually reached and the theoretical position aimed at. The use of the small displacement screw model to describe these position errors can, by its linearity, render the displacements independent and commutative. Thus, for each measuring step, the position error can be written by coding only 6 independent pieces of information (Figure 1) :

- 3 translation errors: 1 linear displacement error (called XTx) and 2 straightness errors (called XTy and XTz)
- 3 rotation errors: Roll (XTx), Pitch (XTy) and Yaw (XTz)

For a serial machine with 3 orthogonal linear axes, the analytical model is constituted of 18 error functions and 3 parameters of axis relative position (perpendicularities). The acquisition of the machine's errors according to the analytical approach necessitates the implementation of different measuring systems specific to each error category.

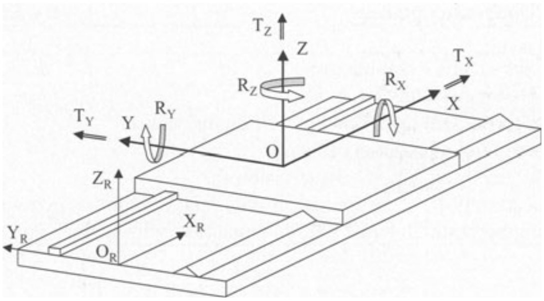


Figure -1. Modelling of the errors of a linear guide of axis X

Once the analytical model has been identified, the positioning error of the tool can be determined whatever the X, Y, Z position of the machine's axes. Error correction can be considered, either in the form of a correction matrix associated to the analytical model [6].

2.2 The measuring techniques

The main difficulty of a machine's geometric verification lies in the measuring operation which is a succession of independent measurements, problematic in the industrial environment, notably because of thermal shifts which are a source of geometric instability. We have made a list of the techniques and the associated uncertainties of measurement, which can give information to the analytical model (table 1).

The number of measurement systems and the ability level of their implementation restrict their industrial use, although the need of machine qualification is universally recognized.

Table -1. Common measuring techniques

Type of error	Measuring technique	Implementation (1)	Uncertainty (2)
3 linear displacements			
XTx, YTy, ZTz	Laser interferometry	difficult	4.10^{-6}.L
6 straightnesses			
XTy ; XTz YTx ; YTz ZTx ; ZTy	Rule + comparator (or) Alignment laser (or) Tense wire + microscope	difficult	2 µm/m
6 Horizontal axis rotations			
XRx ; YRy	Electronic levels	easy	2 µrd
XRy ; YRx ZRx ; ZRy	Electronic levels (or) Interferometer laser by double linear accuracy (or) Autocollimation (or) Double straightness	easy difficult difficult very difficult	2 µrd 3 µrd 3 µrd ~10 µrd
3 Vertical axis rotations			
XRz, YRz ZRz.	Double linear accuracy / straightness Double straightness	difficult very difficult	3 µrd ~10 µrd
Perpendicularities			
$\perp_{xy}$; $\perp_{yz}$; $\perp_{zx}$	Set square + comparator Ball bar (static/ dynamic)	difficult easy	10 µrd

(1) Easy : can be carried out by following written instruction
Difficult : needs trained experienced staff
Very difficult : needs an expert's supervision
(2) Optimum uncertainty (k=2) estimated in industrial conditions in a workshop without air conditioning varying between 18 and 23°C, on a machine with no mechanical lash.

3. TOWARDS AN INDUSTRIAL INTEGRATION OF THE MEASUREMENT OPERATIONS

During the past ten years, new systems of measurement capable of simulating tooling with integrated metrology have appeared (e.g. : Ball Bar, bi-dimensional grid, ...) thus avoiding having to make test workpieces. The great advantage of this type of tool lies in its facility to store information on a state of the machine even if the results have not been interpreted by a practical test. This is the notion of reference state of a system of production which enables an estimation at any time of whether the machine has undergone any changes since it was put into service. Trajectories which can be qualified by these instruments are limited, so their uses are reduced to global practical tests or to local analyses of servo-control errors, but these tools do not allow the exhaustive identification of an analytical model.

3-D measuring instruments like the laser tracker, mainly used in aeronautics, constitute an interesting alternative to the actual measuring instruments. Indeed the laser tracker allows the measurement of the position of a spindle's nose, whatever its position in the workspace. From information obtained from several points, one can deduce geometrical errors of the machine. This system brings about numerous advantages. Its implementation does not require long and meticulous adjustments, all geometrical errors coming from the analytical model are measured by a single device, without intermediate adjustments. This operation is automatic with a permanent information exchange between the numerical command and the laser tracker. As a consequence, total time of qualification is mainly constituted by active measurement periods, depending on the machine's dimensions and the sampling step. Implementation times and adjustments are very short. For example, qualification of a one cubic meter workspace machine requires less than a half-day for the measurement and the analysis, and so can be integrated into a follow-up policy of the production machine.

Our concern is to define a measurement methodology adapted to an objective sought in terms of uncertainty and measuring time. To that purpose, we propose, from the laser tracker's intrinsic characteristics, an implementation method of this instrument which optimizes these parameters as well as an original approach allowing error correction.

4. THE LASER TRACKER

A laser tracker is constituted of a classic interferometric laser mounted on two perpendicular rotation axes, motorized and equipped with angular encoders. It is able to follow the displacements of a measurement point materialized by the reflector center.

In this sort of device, the measurement uncertainty of a position in the volume is far from isotropic. It is composed, for the part linked to the sensor, of the uncertainty due to the interferometer and the uncertainty due to the angular encoders. The uncertainty of measurement introduced by the encoders is proportional to the distance between the position measured and the laser tracker. The measurement uncertainty due to the interferometric laser essentially depends on the environment the light beam has to cross.

Classically environmental conditions (temperature, pressure) are taken into account when measuring interferometric length. The problem is far more difficult in the planes perpendicular to the beam where thermal gradients can locally modify the beam's direction in proportions it is very difficult to quantify in the industrial environment as in the laboratory.

In order to represent the resulting measurement uncertainty, imagine a volume around the measured point. This volume is an ellipsoid each axis of which represents respectively, the uncertainty on the distance measurement, the uncertainty on the elevation and azimuth measurements. An estimate of the measurement uncertainties of a position one meter from the laser tracker is 1 micrometer following the beam direction and 5 micrometers in the perpendicular plane.

5. ACQUISITION AND CORRECTION OF THE GEOMETRICAL ERRORS

The process of geometric correction of a machine is always composed of a stage of measuring and a stage of analysis carried out through the model of correction chosen. The first method to appear was the analytical one, for at that time only unidirectional measuring systems existed and software storage possibilities were extremely limited.

The second method, born with the arrival of 3-D acquisition systems, is global correction by grid [2]. This method consists in measuring the errors Tx, Ty and Tz at the level of the tool according to a grid of the machine's work volume. This information is stored in the matrix and will subsequently permit direct correction of the programmed trajectories. Unlike the analytical correction model, the axis errors are not identified, only the sum of their influence on the position of the tool is characterized. The price to pay for this apparent ease is the considerable amount of data to be stored, now proportional to the cube of the length of the axes and the sampling step, this leads to long measurement cycles, frequently several hours long. So, it is impossible to repeat these cycles often enough to reduce the measurement uncertainty by the effect of averaging. We are directly dependent on the uncertainty of the system of measurement. Its application is limited to 3-axis machines with little variation in the length of tools. Furthermore, direct correction of trajectory presents local instability risks if corrections are not the result of a fine analysis in terms of acceleration generated.

We propose an analysis of these two approaches from the point of view of the measurement uncertainties and the ease of implementation in the industrial environment (table 2) and then we propose a mixed process combining only the advantages.

Table -2. Advantages and disadvantages of the two ways of correction

Approach 1: analytical		Approach 2: global	
Advantages	Disadvantages	Advantages	Disadvantages
Good level of uncertainty Little storage of data : 6.n.L	Implementation of several measuring systems acquisition time very long. Model needing some calculation Measurement difficult to automate	Simple modelling Automatable measuring method	Considerable data: L^n long acquisition time. Uncertainty of the correction directly linked to the measuring system
n= number of axis L = measuring step hyphen often function of the axis length			

5.1 Proposition of a new approach

We propose a third possibility (3) which makes the most of the advantages of both methods, i.e., using 3-D measuring system with a global acquisition strategy, and with the information obtained, identifying the analytical model (figure 2 and table 3). This hybrid approach allows, in comparison with the global one, a return to the analytical model of the machine by an error analysis and so to physical phenomena: straightness of slide channels, error of linear scales, etc.

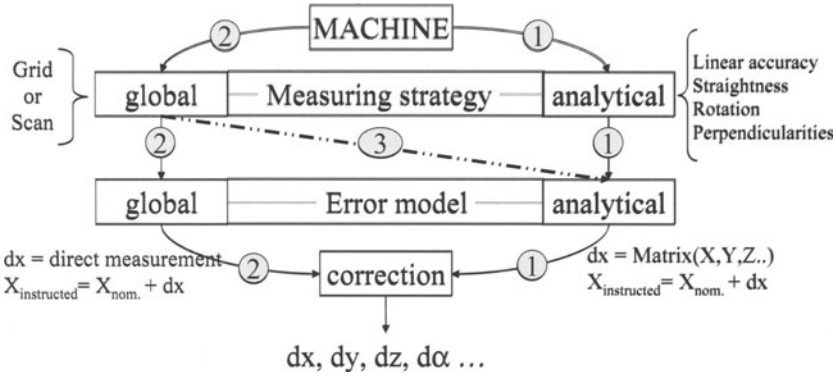


Figure -2. Three approaches for correcting errors

The analysis of a large number of three-dimensional measurements, synthesized through the analytical model, allows the influence of the measurement uncertainty of the device on the estimated characteristics to be decreased and also the minimization of the influence of possible absurd points. As a result, characteristics gain high and accelerations, generated by the corrections, are limited.

Table -3. Mixed approach for the correction of a machine tool

Approach 3 : Global measurement → partial analytical model	
Advantages	Disadvantages
Low uncertainty by effect of average Automatable measuring method Little storage of data	Acquisition of a large amount of data → Long acquisition time All the rotations cannot be determined with one position of the reflector alone

Another significant interest of this analysis strategy is that the influence of the drift during the measurement, mainly from thermal origin, is averaged. Indeed, during the measurement of a three-dimensional grid, which can take more than one hour, the temperature can vary by some degrees if the room is not air-conditioned. The error measured at one point depends on the thermal conditions during its acquisition, so the correction will also be defined for these thermal conditions. The return to the analytical model allows the definition of a correction image reflecting of the average state of the machine perturbed by the external temperature variations or from its components. The potential result of this are a correction less linked to the machine state during the measurement of trajectory errors and the generation of smaller residual errors during future use of the machine.

5.2 Experimental comparison of displacement error

To characterize the contribution of this strategy of evaluation to make a correction, we will restrict our study to the linear displacement error of a machine tool's X axis. The presented measurement is obtained from a three-dimensional grid of the machine (procedure described in the paragraph 6.1). The comparison is made between the linear displacement error XTx obtained by the linear trajectory in the center of the machine during the realization of the grid by the laser tracker and the linear displacement error obtained by the treatment of the whole grid through the analytical model. Each curve represents the deviation from traditional measurement of the central linear trajectory which was made just before the grid acquisition (figure 3).

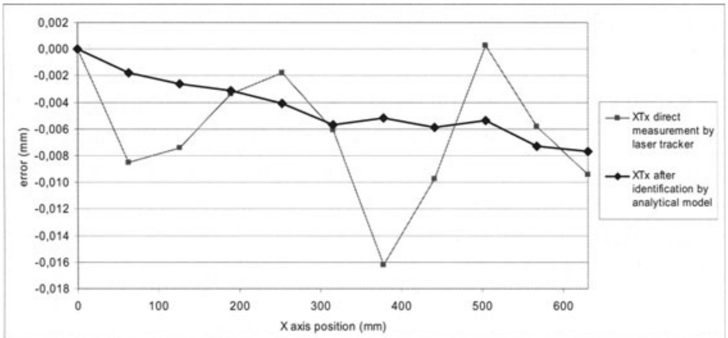


Figure -3. Grid analysis effect on linear error measurement of X axis

It clearly appears that the linear displacement error from direct acquisition is of greater significance than that resulting from the analytical model. It is the consequence, at the same time, of the decrease of measurement uncertainty and the averaging of thermal effects. The average slopes of both curves are the thermal drift signature between the grid measurement and the linear error measurement from the interferometer laser. For this new approach, there is a margin of freedom, which consists in choosing the best strategy of acquisition of points in the working volume of the machine to optimize the error identification of the analytical model. We propose the use of two strategies, which are experimentally qualified.

6. PRATICAL RESULTS

We tested two strategies of acquisition, used within the framework of campaigns of comparative studies carried out on a machine tool with a volume below one cubic meter, so as to situate the potential of this third approach compared to the classic analytical process. The first strategy of acquisition consists in a classic 3-D scan of the machine space, the second original strategy is based on the measurement of circles aligned on the axes of the machine enabling a refinement of the identification of the error model.

6.1 Scan by successive planes

We carried out 11 scans of 11 points in 11 successive planes i.e. 1331 points measured (Figure 4). For reasons related to laser tracker time integration, each point is measured statically, the machine stopped.

We compared the results obtained to conventional measuring systems. We present part of the results, i.e. the straightness error of axis Z following direction X (called ZTx).

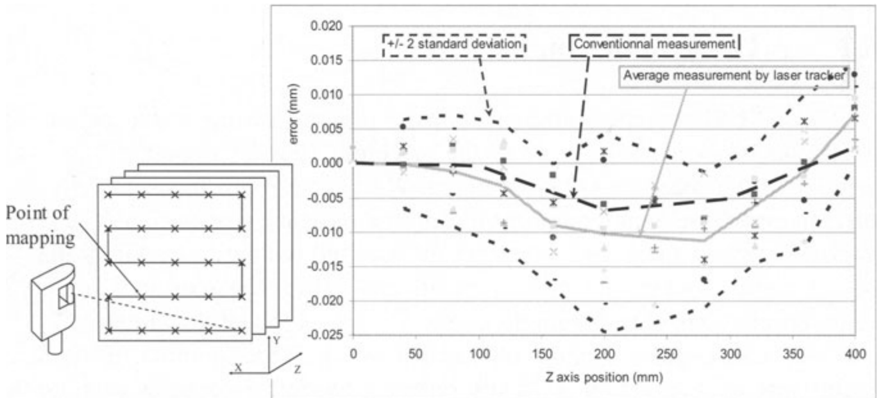


Figure -4. Scan by successive planes and straightness axis Z of the machine tool

The repeatability error with 2 standard deviation is about 5 micrometers for the 11^2, i.e. 121,.straightness measurements of the axis Z. It corresponds to the case of adverse uncertainty since the angular measurement is dominant in this error direction for this position of the laser tracker.

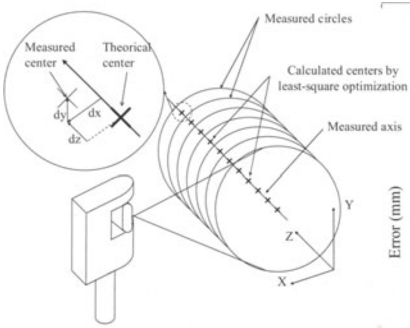
This level of repeatability is not altogether sufficient considering the errors of present machines: more or less the same over 1 meter with repeatabilities of about 1 micrometer. Improvement in the quality of measurement can be obtained by the average of the 121 measurements, which corresponds to the identification of the analytical model. We compared this result to a measurement taken with conventional systems (rule + comparator) at the center of the machine (figure 4). The cross-check is of a good level, even more so since the measurements were not taken at the same time, the test was carried out in an industrial environment (premises not air-conditioned) and the conventional measurement was made at only one place on the machine. The other five straightness errors of the machine were the subject of a similar cross-check and considerable efficacy was recorded in all the cases. The uncertainty of measurement is very clearly improved by the effect of averaging.

The method of scanning by successive planes is very interesting as it enables us to obtain a level of uncertainty comparable, and even superior on certain criteria, to conventional systems, while at the same time exploring the whole volume of the machine automatically. The laser tracker should however be placed in 3 positions so that the linear displacement errors are characterized with a level comparable to that of the classic interferometry commonly used to measure this type of error. Acquisition time is trebled of course, but the analysis of the coherence of the three measuring cycles is an excellent indicator of the machine's stability.

6.2 The cylinder method

The second scanning strategy consists in measuring a succession of circles interpolated in the plane of the machine. Thus 3 " cylinders " of axes parallel to the machine's displacements X, Y and Z are measured. In this case, the machine is in movement during the measuring operation. The laser tracker is able to measure 1 point per millisecond with an uncertainty that is not as good as when the measurements are static. The real axes of the 3 cylinders are constructed from the center of each of the circles (Figure 5).

Each circle center being obtained from a large number of points, uncertainty of its position is greatly reduced. During our studies each circle of 400 mm diameter was measured at about 1300 points. Acquisition time is considerably reduced compared to static measuring (by a factor of 8 in this case). We compared the results obtained by different techniques (Figure 6).



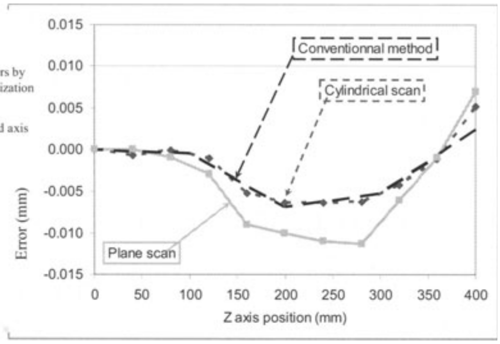


Figure -5. Cylindrical scan of Z axis Z

Figure -6. Straightness axis Z

The result obtained with the cylinder method, compared to the conventional measurement, is better than with the Cartesian grid. This seems logical because the image of the straightness obtained by the cylinder method integrates the machine's errors on a diameter of 400 mm around the average position characterized by the conventional method, while the Cartesian grid covers the whole machine. In theory, the straightness error ZTx is constant in the volume, in practice the existence of deformations due to static indeterminacy of guidance generates error variations.

As for the Cartesian scanning, linear displacement error measurement according to the angular component of the laser tracker does not have a quality comparable with an interferometric measurement. So it is necessary to envisage three measurement configurations for the laser tracker.

7. CONCLUSION

The global approach to measuring a machine tool's errors with the help of a 3-D measurement system like the laser tracker is limited by the accuracy of the angular measurement of these devices. This limitation is of about ± 5 micrometers on small volume machines (less than 1mx1mx1m).

The laser tracker is at the same time a system which carries out practical and analytical tests and one which offers the possibility of rapid investigations during assembly operations, adjustment of the geometry of the machines and for their periodic monitoring.

The implementation of a process of analysis of the machine's errors based on a strategy of global measurement and identification of the analytical error model enables a very significant improvement in the uncertainty of measurement of the errors. Campaigns of comparative measurement with the classic system of geometric qualification of the machine tools show the pertinence and the quality of the results obtained. The fundamental gain is above all to be found in the ease of implementation and the automation of the measurements which make the qualification of the machine possible in less than one day.

ACKNOWLEDGEMENTS

The authors would like to thank O. LAPIERRE, T. ROUX of SYMETRY and D. CAÑADAS, C. FIORONI, Y. PISTRE, T. RECH of RENAULT AUTOMATION COMAU for their confidence and their help.

REFERENCES

[1] G. Dessein, J.M. Redonnet, P. Lagarrigue, W. Rubio, "Machine tool trajectories Compensation by a qualification of dispersions according to machining", IDMME 98 Compiègne, 1998, pp. 759-766.

[2] P. Bourdet et A. Clément, "Controlling a complex surface with a 3 axis measuring machine", Annals of the CIRP ; Vol 25/1 ; pp. 359-361 ; Janvier 1976.

[3] R. Ramesh, M.A. Mannan, A.N. Poo, "Error compensation in machine tools – a review, Part I : geometric, cutting-force induced and fixture-dependent errors", International Journal of Machine Tools & Manufacture, Vol. 40, 2000, pp. 1235-1256.

[4] A.C. Okafor, Y.M. Ertekin, "Derivation of machine tool error models and error compensation procedure for three axes vertical machining center using rigid body kinematics", International Journal of Machine Tools & Manufacture, Vol. 40, 2000, pp. 1199-1213.

[5] A.H. Slocum, Precision Machine Design, Prentice Hall, 1992.

[6] J.M. David et T. Coorevits, "Limits of the numerical error compensation of Coordinate Measuring Machine. Definition of a new machine architecture", 6th International Metrology Congress, Lille - France, 1993.

SIMULATION OF REGENERATIVE VIBRATIONS AND SURFACE GENERATION IN FLANK MILLING
APPLICATION TO HIGH SPEED MACHINING

G. Peigné, H. Paris and D. Brissaud
Laboratoire Sols, Solides, Structures, Grenoble, France

Abstract: High speed machining is really suitable for present production constraints. However high rotational speeds inherent to high cutting speeds excite structures at high frequency. Resulting dynamic phenomena can reduce surface quality, damage material and decrease productivity. So as to avoid machining cutting conditions leading to the development of such phenomena, we developed a tool allowing the numerical computation of the machining system dynamic behaviour as well as to predict the machined surface geometry and quality. Our approach is based on the numerical computation at the macroscopic level of workpiece-tool system dynamic behaviour submitted to cutting forces. These forces are modelled and allow us to deduce the workpiece motion thanks to the Newton equation which is solved with an implicit integration scheme. The cutting nonlinearities are taken into account by the modelisation of the surface being machined. Simulation results follow the experimental trends, and a machining simulation example is given to show the main regenerative vibration effects upon the behaviour of machining system as well as upon the surface quality.

Key words: high speed milling, simulation, regenerative vibrations, dynamic cutting forces, ploughing

1. INTRODUCTION

High speed machining (HSM) satisfies present machining production constraints. As it allows high material removal rate and high surface quality, this makes this technology very competitive. However, when tool rotational

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 313–322.

frequencies increase, vibrations that may occur during cutting process may damage the spindle as well as the cutting tool and decrease surface properties.

Early researches has shown that the most important source of instability is the regenerative vibrations due to the coupling between cutting forces and structural dynamics of the tool/workpiece/machine-tool system [1]. The most outstanding result given by research on this phenomenon is, for high rotational speed, the identification of stable cutting process regions between unstable regions. These are commonly represented in a spindle speed versus depth of cut diagram where they have a lobed shape. Firstly, analytical prediction models of the lobes were made for turning operations [2], [3], latterly, models have been adapted to milling operations [4].

One of High Speed Machining's definitions is based on this vibration phenomenon, [5]. Stability lobes show that milling stability strongly depends on the ratio η of the tool teeth passing frequency upon frequency of dominant vibrating mode. Four different zones exists, depending on the spindle speed, where machining behaviours are very different. Each zone corresponds to a machining type. In low-speed machining zone, $\eta<1/10$, regenerative vibrations that may occur have short wavelengths. Thus, even with low amplitude, they generate a contact between the tool clearance face and the workpiece ; the strong resulting forces tend to reduce vibrations. This phenomenon, called machining process damping, allows high axial depth of cut. However, the longer vibration wavelengths are, the less the damping acts. In midrange machining zone, this damping no longer acts while stability lobes do not yet exist. This gives rise to a low and almost constant limit of stability. This zone ends when spindle speed is high enough, i.e. η greater than ¼, to obtain large stable regions. In the high speed machining zone, the effect of stability lobes is important, and the stability limit greatly depends on the spindle speed. The choice of ratio η equal to an integer fraction allows the highest depths of cut. Finally, in Ultra High Speed Machining zone, η over 2 or 3, the highest stability lobe is exceeded. Rotational speeds don't excite the structure very much and the machining process becomes very stable. However, at the moment, spindles cannot reach such speeds yet, and above all, high cutting speeds lead to excessive tool wear.

Therefore, the best removal material rate are obtained in HSM. However, in these, spindle speeds excite the structure close to the frequencies of weak structural modes, and vibrations may have large amplitudes even in chatter free cutting conditions. Thus the machined surface may be affected and have important form defects. Thanks to the rise of computers, time domain simulations have been developed [6], [7], allowing the resulting surface finish produced by vibrations in milling to be predicted. Furthermore, they can consider the modifications of structural behaviour during machining, the evolution of cutting conditions, like the tool penetration phase or the tool disengaging phase and real tool geometry with tooth run out or tool eccentricity.

However as time domain simulations are time consuming and they must be optimised for specific applications. In this paper a time domain simulation, in ongoing development, is presented. This simulation deals with peripheral milling operations in High Speed Machining. It is dedicated to thin walls machining application and long tool use. In section 2, the different models used are presented. Particular attention is given to the geometrical models used to describe surface machining. These models have been optimised for HSM in order to effect accurate geometrical computations with low computational cost. In section 3, a machining simulation is shown in order to see different effects of vibrations on machined surface. Finally, the paper is concluded by summarising the contributions.

2. TIME DOMAIN SIMULATION

The simulation is based on a macroscopic scale model of the dynamic behaviour of the tool, workpiece, machine-tool system submitted to cutting forces. Thanks to Newton's second equation we can derive system motions from the forces caused by these tool/workpiece interactions. An explicit time integration scheme is used

2.1 Mechanical model

A differential matrix equation is used to give the motion of The tool-workpiece-machine system :

$$\begin{aligned}[M_w]\{\ddot{X}_w\}+[C_w]\{\dot{X}_w\}+[K_w]\{X_w\}&=\{Fc(t)\}\\ [M_t]\{\ddot{X}_t\}+[C_t]\{\dot{X}_t\}+[K_t]\{X_t\}&=\{Fc(t)\}\end{aligned} \tag{1}$$

where *[$\boldsymbol{X}_w$]* and *[$\boldsymbol{X}_t$]* are respectively workpiece and tool position vectors, *[M_w]* and *[M_t]* the mass matrix, *[C_w]* and *[C_t]* the damping matrix, *[K_w]* and *[K_t]*, the stiffness matrix. However, the system must be rigid in the *Z* direction.

2.2 Cutting forces model

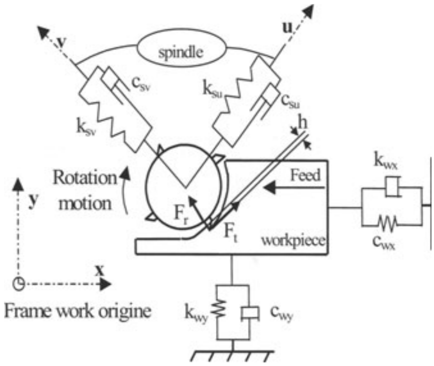


Figure -1. Mechanical model of flexible workpiece

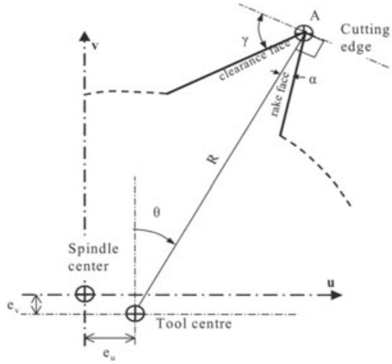


Figure -2. Tooth geometry

Cutting force *{Fc}* is expressed by its radial and tangential components *Fr* and *Ft,* figure 1. Orthogonal cutting forces model is used :

$$\begin{aligned} F_t &= Ks.a_p.h^m \\ F_r &= Kr.Ft \end{aligned} \qquad (2)$$

Where a_p is the axial depth of cut, h is the instantaneous chip thickness and *Ks, Kr, m* cutting force coefficients depending on workpiece material and tool geometry. For a non zero helix angle tool, forces are computed by slicing cutting edge in elementary orthogonal tools, which allows use of the former Single point cutting forces model. Ploughing is not modelled. Though, it is supervised computing the instantaneous clearance angle.

2.3 Geometrical models

To compute cutting forces, as well as the surface being machined, a geometrical description of the tool and of the workpiece surface is used. In order to reduce calculation complexity and therefore decrease computing time, tool and workpiece surface are "sliced". All computations are done in the workpiece referential.

Tool Geometrical model

In each slice, the tool is modelled by elementary teeth. As the slices are thin, the assumption is made that the cutting edge is straight and normal to the tool axis. Elementary tooth geometry is given in figure 2. Each tooth is defined by its rake face and clearance face, both are straight, respectively oriented by the clearance angle γ and the cutting angle α. The cutting edge, assimilated to point A, is positioned by a radius R and an angular position θ. Lastly, tool eccentricity is given by the coordinates e_u and e_v. With this decomposition, total cutting and ploughing forces are the sum of all elementary forces which are computed with the relation (2).

Cutting edge motion

Cutting edge position is computed for each time step. Therefore, the motion between two time steps must be known to generate the surface being machined and to calculate chip thickness. Step time is usually given by the dynamics of the modeled system and the integrative method used. In low speed machining simulation, cutting edge displacements between two time steps are small or quite straight and may be linearly interpolated. In high speed machining, same time increments lead to bigger displacements. Thus linear interpolation may distort computation of chip thickness and generation of machined surface. So, if we want to maintain high accuracy in geometrical computation, the time step must be reduced or the interpolation model must be changed. The model we use is based on a quadratic interpolation of cutting edge trajectory, figure 3. The following notation is used : Δ_t is the time step, instant $t=n\,\Delta_t$ is chosen as time origin and we note –i instant corresponding to $t=(n-i)\,\Delta_t$.

The interpolating curve C has a second degree polynomial equation in the referential R'=(O',x',y'), figure 4. It is built to fit the last three positions A_0, A_{-1}, A_{-2} of the cutting edge. The referential R' base (**i**',**j**',**k**') and its origin O' are built the following way :

$$\boldsymbol{i}'=\frac{\boldsymbol{r}_{A0}-\boldsymbol{r}_{A-2}}{L},\quad \boldsymbol{j}'=\boldsymbol{k}\wedge\boldsymbol{i}',\quad \boldsymbol{r}_{O'}=\boldsymbol{r}_{A0}-l\boldsymbol{i}',\ L=\left\|\boldsymbol{r}_{A-2}-\boldsymbol{r}_{A1}\right\|,\quad l=\left(\boldsymbol{r}_{A0}-\boldsymbol{r}_{A-1}\right)\cdot\boldsymbol{i}' \qquad (3)$$

The vector notation used is :$\boldsymbol{r}_A=\boldsymbol{OA}$ and $\boldsymbol{r}'_A=\boldsymbol{O'A}$. In this referential, the vectorial equation of the interpolating curve C is :

$$\boldsymbol{r}'_C(\xi)=l\cdot\xi\boldsymbol{i}'+\frac{l\cdot m}{n}(1-\xi)\left(\xi+\frac{n}{l}\right)\boldsymbol{j}' \qquad (4)$$

With $n=L-l$ and $m=\left\|\boldsymbol{r}_{A_{-1}}-\boldsymbol{r}_{A_0}\right\|$. Trajectory between time step *t-1* and time step *t* corresponds to $\xi\in[0,1]$.

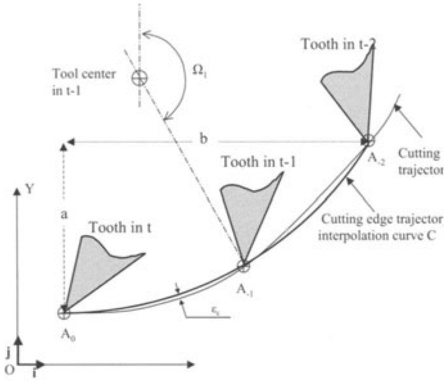


Figure -3. Cutting edge trajectory interpolation

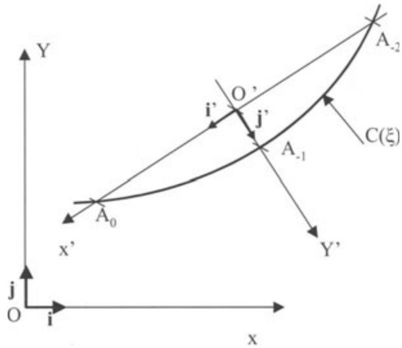


Figure -4. Cutting edge trajectory refential *R'*

Material removal process

In each slice, the surface is discretized by points. Between two consecutive vertices, the surface is linearly interpolated. To model material removal at each time step, surface points included in volume scanned by teeth rake faces are removed and new points belonging to the teeth trajectories are created. This creation process consist of three phases :

First step : For each tooth a primary identification of potentially removable surface points is made, figure 5. Vertices S_i of this set of points, *PRS,* belong to the surface scanned by a straight line going through cutting edge and tool center during the time step :

$$PRS=\{S_i\}\ ,\ \left(\mathbf{Tc_0A_0}\wedge\mathbf{Tc_0S}_i\right)\cdot\mathbf{k}\geq 0\ \ \&\ \ \left(\mathbf{Tc_{-1}A_{-1}}\wedge\mathbf{Tc_{-1}S}_i\right)\cdot\mathbf{k}<0 \qquad (5)$$

Second step : Points of *PRS* belonging to the volume scanned by the tooth rake face are determined. They form the set *RS*, figure 6. the hypothesis is made that surface and cutting edge trajectory can not have more than one intersection point. Thus, only three material removal configurations exists : case 1, the tooth is still machining during the time step ; case 2, illustrated in figure 5-7, the tooth penetrate into the workpiece ; case 3, the tooth leaves the workpiece. These cases are determined by values of chip thickness at instant t and t-1 :

$$\begin{array}{llll} \text{case 1} & h_0>0\ \&\ h_1>0, & C\cap Surf=\phi & ,\ \text{RS}=PRS \\ \text{case 2} & h_0>0\ \&\ h_1=0, & C\cap Surf=C(\xi_s) & ,\ \text{RS}=\{S_i\}_{i\in[m,m+p]} \\ \text{case 3} & h_0=0\ \&\ h_1>0, & C\cap Surf=C(\xi_s) & ,\ \text{RS}=\{S_i\}_{i\in[m+p+1,m+n]} \end{array} \qquad (6)$$

Points S_{m+p} and S_{m+p+1} are the vertices of the surface segment cut by the cutting edge trajectory. They are found by :

$$\begin{gathered} \left(\boldsymbol{r}'_{C(\xi_{m+p})}-\boldsymbol{r}'_{S_{m+p}}\right)\cdot\left(\boldsymbol{r}'_{C(\xi_{m+p+1})}-\boldsymbol{r}'_{S_{m+p+1}}\right)\le 0 \\ \xi_i=\frac{1}{l}\boldsymbol{r}'_{S_i}\cdot\boldsymbol{i}' \end{gathered} \qquad (7)$$

In order to obtain an unequivocal solution, the first couple of point satisfying the former equation are taken. The research is done following decreasing indices for case 2 and increasing indices for case 3. The intersection $C(\xi_s)$ of the cutting edge trajectory with the surface is obtained using the Newton method.

<u>Third step :</u> The surface machined during an actual time step is created. The RS set is replaced by a set *NP* of new points, figure 7 :

$$\begin{array}{ll} \text{case 1} & NP=\{E,C(\xi_k)\},\ \xi_k=1-\dfrac{k}{q}\ ,\ k\in[0,q] \\ \text{case 2} & NP=\{E,C(\xi_k)\},\ \xi_k=1-\dfrac{k}{q}(1-\xi_s)\ ,\ k\in[0,q] \\ \text{case 3} & NP=\{C(\xi_k)\},\ \xi_k=\dfrac{(q-k)}{q}\xi_s\ ,\ k\in[0,q] \end{array} \qquad (8)$$

For cases 1 and 2 the intersection E between surface and tooth rake face is created. For its computation, the rake face is assimilated to line (Tc_0, A_0), figure 6. The number q+1 of created points is chosen to keep the distance between two consecutive points lower than a maximal value Δ_s :

$$q=\mathrm{int}\left(\frac{\Delta_\theta\cdot R}{\Delta s}\cdot|\xi_e-\xi_f|\right) \qquad (9)$$

where $(\xi_e,\xi_f)=(0,1)$ for Case 1, $(\xi_s,1)$ for case 2 and $(0,\xi_s)$ for case 3 ; Δ_θ is the angular step of the spindle rotation, R the tooth radius and *int(X)* the integer part of *X*. The spatial step Δ_s is chosen equal to 1/5 of the feed rate :

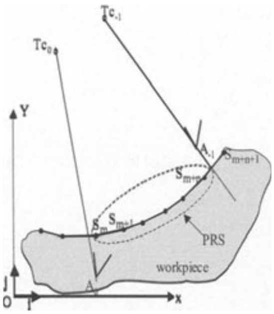


Figure -5. Material removal process, step 1

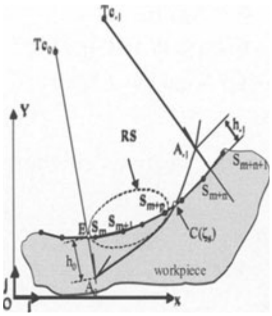


Figure -6. Material removal process, step 2

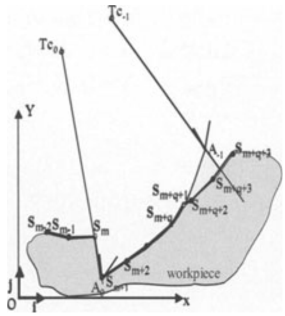


Figure -7. Material removal process, step 3

2.4 Simulation principle

Al the models have been coded in MATLAB©. The simulation input data are cutting conditions, cutting force model parameters, ploughing force model parameters, structural dynamic coefficients and workpiece and tool geometry. A second order implicit Euler integration scheme is used. For each time step, the machining process evolution is done in two phases. First, equilibrium of cutting forces and workpiece and tool motions in equation (1) is sought. During this phase, the surface is not modified. Next, when equilibrium is reached, the surface part machined during the step time is created using the material removal model. The main results are the machined surface, vibrations of both tool and work-piece and cutting forces. We get either cutting and ploughing forces.

3. SIMULATION RESULTS

In order to see regenerative vibration effects upon machined surface, simulation of a flexible aluminum workpiece machined with a rigid 20 mm diameter, 3 tooth, 35° helical end mill has been computed. The following data were introduced : cutting conditions : a_e=2mm ; a_p=18mm ; f=0.15mm/tooth ; spindle speed 3660 rev/min ; workpiece first bending mode in y direction with a natural frequency of 176 Hz, structural damping ratio of 2% and a stiffness of 2.1 10^4N/mm ; force model parameters : K_s=600 Mpa ; K_r= 0.3 ; h=0.75 ; f_{sp}=1.5 105 N/mm^3; μ_c=0.3 ; Simulation parameters : δt=4.5μs.

Figure 8 shows the machined surface. Three kinds of faults appear. First, in the z direction, while x<64mm, the surface has a marked undulation leading to a shift with respect to the theoretical position varying between 0.055mm and -0.035mm. This error is due to both helix angle and workpiece vibrations. Indeed, as a cutting edge machines line *A-B*, the tool completes 0.2 turns and the workpiece is vibrating, figure 9. Therefore, for a given point, surface shift is equal to the instantaneous workpiece deflection.

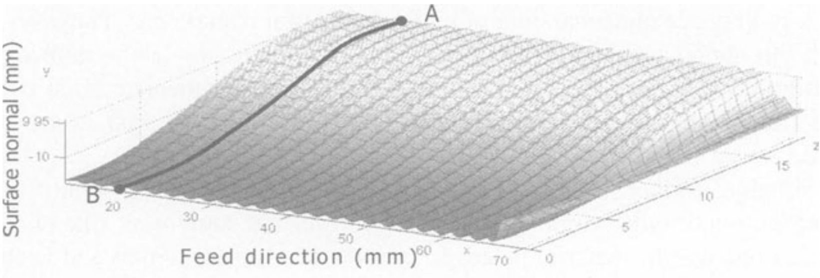


Figure -8. Form default surface

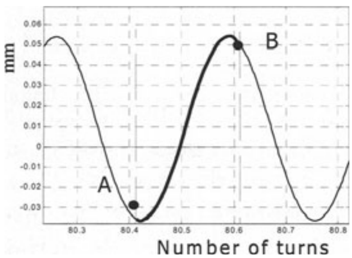


Figure -9. Workpiece displacement

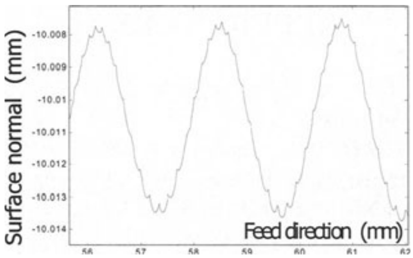


Figure -10. Surface detail

In the feed direction x, the surface has small amplitude and short wavelength undulations. Figure 10 shows a cut in the surface in the z=7mm plane. We can see both the feed marks and these undulations. The latter are due to the transient phase of vibrations occurring at the beginning of the machining. At the end, as the tool is moving out of the workpiece (x>64mm), the surface is strongly affected by cutting condition variations.

4. CONCLUSIONS

A comprehensive simulation model of flank milling operation in the time domain has been presented. Both workpiece and milling cutter mechanical behaviors are taken into account and self-excited vibrations are dealt with. Cutting forces are computed with oblique cutting mechanism and real tool geometry with tooth run out is included.

Also, new machined surface and tooth motion geometrical models optimized for High Speed Machining have been developed, allowing the control of geometrical errors. As these models permit high geometrical accuracy independently of the time step value, they enable low computational cost.

Simulation results show the same trends as experimental studies and may greatly improve understanding of cutting vibration phenomena. However, in order to obtain accurate results, validation of dynamic models in high speed cutting conditions must be achieved. to this aim, experiments carried out in the field of cutting forces show that low speed cutting models are still effective for high cutting speeds [8].

Finally, further investigations are required, in order to identify the contribution of other vibration phenomena to surface roughness, like random excitations due to material heterogeneity or the chip formation and rupture process.

5. REFERENCES

[1] TOBIAS S. A., FISHWICK W., *"A theory of Regenerative chatter "*, The Engineer-London,1958

[2] MERRIT, H.E., *"Theory of Self-Excited Machine Tool Chatter"*, Trans. ASME Journal of Engineering for Industry, vol.87, 1965, pp. 447-454.

[3] KOENIGSBERGER F., TLUSTY J., "*Machine Tool Structures-Vol.I : Stability Against Chatter*", Pergamon Press, 1967.

[4] ALTINTAS Y., BUDAK E., "*Analytical Prediction of Stability Lobes in Milling*", Annal of the CIRP Vol. 44/1, 1995, pp. 357-362.

[5] SMITH S., TLUSTY J., "*Current trends in high-speed machining*" Journal of Engineering for Industry vol. 119, 1997, pp. 664-666.

[6] MONTGOMERY, D., ALTINTAS, Y., "*Mechanism of Cutting Force and Surface generation in Dynamic Milling*", trans. ASME Journal of Engineering for Industry, vol.113, 1991, pp. 160-168.

[7] DEKELBAB K., "*Modelisation et Simulation du Comportement de l'Ensemble Piece/Outil/Machine en Usinage par Outil Coupant*", These de doctorat de l'Ecole Nationale Superieure des Arts et Matiers-Paris, 1995

[8] PARIS H., DELHEZ C., "*Modelling cutting force in high speed milling*" Proceeding of International Seminar on Inproving Machine Tool Performance, July, 3-5, 2000. La Baule, France.

Chapter 3

METHODOLOGICAL ASPECTS OF INTEGRATED DESIGN AND MANUFACTURING

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 323–324.

FEATURE SHAPE RECOGNITION TOOLS DEDICATED TO A DESIGN FOR ASSEMBLY METHODOLOGY

O. Coma*, C. Mascle* and P. Véron**
École Polytechnique de Montréal, Québec
*** École Nationale Supérieure d'Arts et Métiers, Aix-en-Provence, France*

Abstract: This paper presents an automatic feature shape recogniser tool for a Design For Assembly (DFA) methodology. For each part of a product, this feature research is performed into two steps. First, weight, dimensions and symmetries are identified and the part is classified as rotational or not. Second, form features, like steps or grooves, are extracted in order to estimate the orientation efficiency of the part. To fulfil first step requirements, an optimal bounding box, in the context of DFA, is found using the boundary representation (B-Rep) model of the part. The proposed algorithm is based on a topological exploration and on geometric properties of faces. 2D features identification step operates on a simple vertex-edge polyhedral model of the part. On every plane projection, along an axis of the bounding box, outer wire analysis serves as starting point for the 2D feature recognition. Finally, a 2D feature selection procedure, that optimises the orientation efficiency, is presented. All these algorithms are implemented in "FuzzyDFA", an assembly-oriented computer aided design software that takes advantage of a fuzzy decision support system.

Key words: Design For Assembly, Bounding box, Symmetry, Feature shape recognition.

1. INTRODUCTION

DFA is an important factor in manufacturing since it can substantially reduce the estimated 15-70% of manufacturing cost attributed to assembly. Besides the reduction of cost, DFA promises additional benefits by increasing quality and reliability, and shortening manufacturing time. In this paper, we refer the well-known Boothroyd-Dewhurt methodology [1]. In the former, the DFA evaluation process centres on establishing the cost of

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 325–336.

handling and inserting parts. The process can be applied to manual or automated assemblies. Regardless of the assembly system, parts are evaluated in terms of, ease of handling, ease of insertion and a decision as to the necessity of the part in question. The findings are then compared to synthetic data. During this evaluation process, geometric and technological data attached to every part are used. From the geometric point of view, parts are individually studied, whereas from the technological view, links between them are taken into consideration. This work, coupled to the methodology previously developed and described in [7], represents a major breakthrough from feature-based part coding technology [8], whereby generated codes were inputted for DFA evaluation. The approach herein, associated to a fuzzy decision support system, aims to intelligently automate DFA process evaluation. Herein, geometric requirements for manual and automated assembly operations are identified. First, the proposed approach to find an optimal (in the context of DFA) bounding box and symmetries of a part is presented. These geometric properties are qualified as basic DFA requirements, since the evaluation process always use them whether or not the assembly operations are automated, and operates on the B-Rep geometric model of the part. Then, a 2D features recogniser tool using the polyhedral geometric model of the part is detailed, as well as the definition of features in the context of DFA. Finally, a feature selection algorithm optimising the part orientation efficiency is presented.

2. RELATED WORKS

In Rosario [2], development and implementation of computerised algorithms to calculate overall dimensions and rotational symmetries of a part are described. A wire-frame representation scheme of the CAD part model is used. Algorithms are based on topological relationships and a palindrome search. This thesis work corresponds to the basic requirements presented in this paper. But, using a wire-frame geometric model restricts algorithms to very simple parts, due to the lack of topological information inherent in this scheme. Boothroyd-Dewhurst [1] introduced the degenerated envelope in order to find a convenient bounding box for DFA. Essentially, the degenerated envelope is a cylinder, or a rectangular prism obtained when detailed features on the part are neglected. Although this approach provides good theoretical results, it uses the relative notion of feature size and criteria to decompose features which must be found. Consequently, this bounding box research process is difficult to automate. Jiang [3] proposed a literature review for the detection of symmetries (axis and angles) in a polyhedron. The described algorithms are complex and time consuming due to the search for an axis of symmetry. Herein, a bounding box for the part is first defined, thus possible axes of symmetry are a priori known for the symmetry searches. Finally, literature abounds on feature recognition. A feature can be defined as a partial form or a part characteristic, which is considered as a unit and that, has a semantic meaning in design, process planning, manufacture, cost estimation or other engineering discipline [4]. "*Feature technology*" [5] describes advances in feature recognition as well as the context where features are of interest.

3. GEOMETRIC REQUIREMENTS FOR DFA

Analysis of mechanical assemblies always uses basic DFA requirements. Feature recognition is only to be computed for automated assembly operations. Table 1 lists requirements for a common DFA methodology:

Table -1. Part geometric characteristics searched for

	Rotational part	Other part (non-rotational)
Basic requirements	Dimensions: L, D	Dimensions: A, B, C
		Axes of symmetry X', Y' et Z'
	Weight	
	α, β symmetries	
Feature recognition	Steps, Grooves, Cavities, Steps, grooves and cavities with small dimensions	

The optimal bounding box (for DFA) of a part is the one, of minimal dimensions, which fits the best underlying geometry. Its 3D coordinate system is defined by main geometric forms (planes, cylinders, etc.). Figure 1 shows bounding boxes for a rotational part and a non-rotational one.

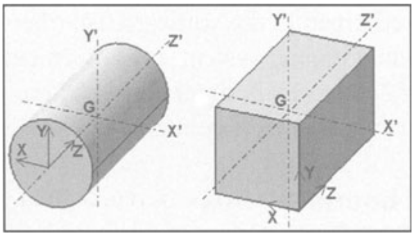


Figure -1. Bounding boxes for a rotational and a non-rotational part

The β (α) symmetry is the angle through which a part must be rotated about its axis of insertion (an axis perpendicular to its axis of insertion) in order to repeat its orientation. In figure 1, if **Z** is the insertion axis of both parts, axis **Z'** (**X'** or **Y'**) are of interest for the β (α) symmetry. In automated assembly operations, parts are oriented and distributed by bowl feeders. These bowls have mechanical selections that define, step-by-step, the orientation of a part. During every step, non-conforming parts are discarded as shown in figure 2. Accordingly, 2D features in one or more plane projections could be used to partially or completely define the part orientation. Here, only 2D steps or grooves features are of interest. If the orientation could not be defined solely by the 2D features, a 3D feature recogniser tool focused on pockets, holes and fillets should be applied to the geometry. Furthermore, DFA methodologies just focus on the type of features and hence feature decomposition is useless. The aim is to select the minimum number of features to orient the part, and the most appropriated (simple or complex) features that maximise the orientation efficiency.

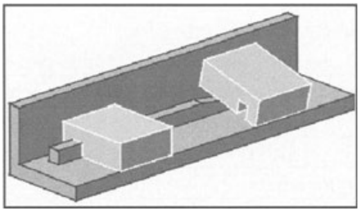


Figure -2. Orientation of parts using selections in a bowl feeder

4. SOLVING BASIC REQUIREMENTS

Although basic requirements seem easy to find, dimensions and symmetries depend on the 3D coordinate system attached to the part. When using a 3D modeller, designers usually create a part in the global 3D coordinate system predefined in the software. But there is no a priori reason to ensure that this coordinate system is convenient in measuring DFA requirements. Consequently, the optimal bounding box for each part must be first computed for further DFA evaluations.

4.1 Optimal bounding box

For a given part, the approach is based on a topological exploration of its solid master model. Geometric and topological entities are described using STEP AP203 standard where the "*manifold_solid_brep*" topological data structure $\mathbf{P}_{DS}$ (part data structure) is the shape of interest herein. The proposed algorithm explores $\mathbf{P_{DS}}$ in order to find the set of faces that characterises the best the underlying geometry. Then, the 3D coordinate system generated for that main characteristic serves as support for the optimal bounding box building. The following algorithm is illustrated with the part in figure 3:

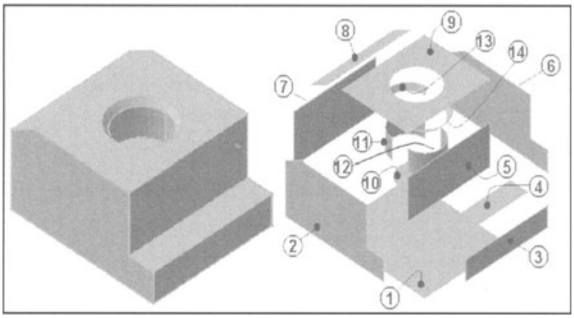


Figure -3. Face based topological decomposition of a solid

The first step consists in exploring faces of $\mathbf{P_{DS}}$. Faces are dispatched in three sets: $\mathbf{F_{PLN}}$, gathering plane faces, $\mathbf{F_{REV}}$, for rotational faces and $\mathbf{F_{OTH}}$, for other geometric face supports. Moreover, inside each of these sets, faces having common properties are gathered into sub-sets. Table 2 lists the common face properties in every set:

Table -2. First groupings of faces

Set	Type of surface	Property
F_{PLN}	Plane	Same normal
F_{REV}	Cylindrical, spherical, conical or of revolution	Same axis
	Spherical	Same centre
F_{OTH}	Of extrusion, Bézier, B-Spline or other	

But, inside $\mathbf{F_{PLN}}$, some faces must be differently treated. For example, although both end faces of a cylinder are plane, they contribute to make the part rotational. Consequently, for each plane face **f**, a test evaluates whether it must be classified as rotational or not. This test explores edges gathered in three sets (*linear*, *circular* and *other*). Then, the total length **L1** of edges in *linear* set and in *other* set is compared to the lengthiest edges **L2** in the *circular* set. If **L1>L2** then **f** is added to $\mathbf{F_{PLN}}$ else to $\mathbf{F_{REV}}$. For the solid in figure 3:

- $\mathbf{F_{PLN}}=\{\{f1,f4,f9\},\{f3,f5,f7\},\{f2,f6\},\{f8\}\}$;
- $\mathbf{F_{REV}}=\{\{f11,f12\},\{f13,f14\},\{f10\}\}$;
- $\mathbf{F_{OTH}}=\{\ \}$.

Secondly, faces in $\mathbf{F_{PLN}}$ and $\mathbf{F_{REV}}$ that have close properties are gathered into two new sets $\mathbf{GF_{PLN}}$ and $\mathbf{GF_{REV}}$ as follow (cf. table 3):

Table -3. Second groupings of faces

Set	Properties
$\mathbf{GF_{PLN}}$	Elements of $\mathbf{F_{PLN}}$ with a collinear or a perpendicular normal.
$\mathbf{GF_{REV}}$	Elements of $\mathbf{F_{REV}}$ with the same axis whether type of surface is the same or not. If the face is plane, the centre P of the lengthiest circular edge is computed. Then, this face is added to the set for which P is located on the axis.

For the solid in figure 3:

- $\mathbf{GF_{PLN}} = \{\{f1, f2, f3, f4, f5, f6, f7, f9\}\{f8\}\}$;
- $\mathbf{GF_{REV}} = \{\{f10, f11, f12, f13, f14\}\}$;
- $\mathbf{F_{OTH}} = \{\ \}$.

Thirdly, the area $\mathbf{Area\_GF_{PLN}}$ ($\mathbf{Area\_F_{OTH}}$) of the $\mathbf{GF_{PLN}}$ ($\mathbf{F_{OTH}}$) set is computed as the sum of all the face areas of this set. The area $\mathbf{Area\_GF_{REV}}$ of the $\mathbf{GF_{REV}}$ set is computed as follow:

$\mathrm{Area\_GF_{REV}} = \sup_{i=1}^{n}(area\_GF_{REVi})$ where $area\_GF_{REVi}$ is the area of the sub-set i of the $\mathbf{GF_{REV}}$ set. Then, the part is classified as rotational if $\mathbf{Area\_GF_{REV}} > (\mathbf{Area\_GF_{PLN}} + \mathbf{Area\_F_{OTH}})$. The main characteristic for the bounding box is the main characteristic of the set of maximal area. The last step consists in finding a 3D coordinate system for this characteristic and then, dimensions for this bounding box. The coordinate system is found using geometric and mass properties of the main characteristic.

4.2 Symmetries

Angle values of 180°, 120°, 90° and 72° are of interest for α and β symmetries. In the DFA methodology, parts are considered to be α or β symmetric as soon as the angle around an appropriated axis is less than 180°. But values less than 180° enable symmetry detection for a part, the transversal section of which is triangular, rectangular or pentagonal. Then, the volume $\mathbf{V}$ of the intersection solid between the part and the same part rotated around $(\mathbf{\Delta}, \mathbf{\Phi})$ is computed. If $\mathbf{V} \leq \varepsilon$, where ε is a real number close to zero, then the solid admits this symmetry.

5. 2D FEATURES RECOGNITION

When a part is progressing in a bowl feeder, 2D projection of its solid representation allows the definition of possible selections to orient that part. Thus, after computing a 2D projection, features of interest herein are those relative to the 2D bounding box of that projection. This section first presents the method to detect the outer wire of the polyhedral model projection of a part before explaining the 2D feature recognition process. In the following, the B-Rep geometric model representing the part is transformed in a corresponding polyhedral model, obtained after triangulating every of its faces according to Delaunay's method [6]. An object oriented data structure has been developed to handle a polyhedron.

5.1 Outer wire

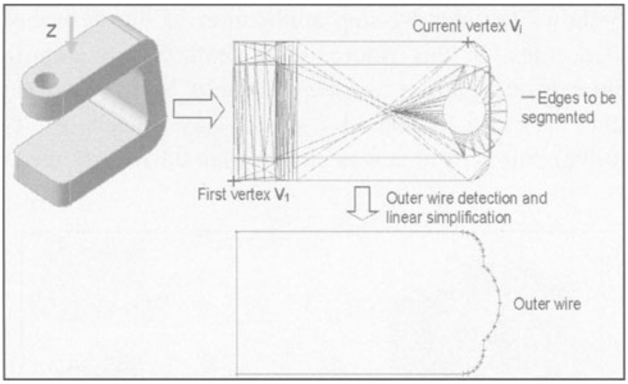


Figure -4. Projection of a polyhedral model into a plane

Figure 4 gives indication of the selected approach. A polyhedral simplification algorithm is applied (with a given precision) in order to suppress redundant vertices and edges of lengths equal to zero. Then, a first vertex $\mathbf{V_1}$ on the outer wire must be found. To do so, the 2D bounding box $\mathbf{bnd_{2D}}$ including all the vertices of the polyhedron is built. The vertex, for which (X,Y) coordinates are minimal, that lays on a bounding box border is selected as $\mathbf{V_1}$. A loop allows the selection of the next vertex to be included in the outer wire till current vertex $\mathbf{V_i}$ is at the same location as $\mathbf{V_1}$. Candidate edges are gathered in a set, each of these edges containing $\mathbf{V_i}$. Among every edge $\mathbf{E_j}$ for which the dot product, between normalised vector $\overrightarrow{V_{i-1}V_i}$ and normalised vector $\overrightarrow{E_j}$, ($\mathbf{V_i}$ being the starting vertex of $\overrightarrow{E_j}$) is minimal, that of $\mathbf{E_K}$ of minimal length is selected. Then, $\mathbf{E_K}$ is segmented if necessary. $\mathbf{V_{i+1}}$ is the vertex belonging to $\mathbf{E_K}$, which is not equal to $\mathbf{V_i}$.

5.2 2D features from projection

A plane face **F** is created using the outer wire defined above. The set of faces $\mathbf{N} = \mathbf{Fbnd_{2D}} - \mathbf{F}$ is the basis for feature detection. Only three types of 2D features are defined to fulfil DFA requirements: **Feat2d_Groove**, **Feat2d_Step** and **Feat2D_Other**. For each face $\mathbf{f_{Ni}}$ belonging to **N**, the type of the related feature is found according to the bounding box $\mathbf{bnd_{2D}}$ and area of the face **F** following three rules:

1. a feature with 2 vertices on a border of $\mathbf{bnd_{2D}}$ is of type **Feat2d_Groove**;
2. a feature with 3 vertices on a border of $\mathbf{bnd_{2D}}$ is of type **Feat2d_Step**;
3. a feature type of which is either **Feat2d_Groove** or **Feat2d_Step** and area of which is less than 0.01 times area of face **F** becomes of type **Feat2d_Other**.

Figure 5 shows the step-by-step application of the described procedure and associated rules. In this figure, three features are identified. Feature number 1 and 2 are recognised like steps (Feat2d_Step) whereas number 3 is of type other (Feat2d_Other). It would have been of type groove (Feat2d_Groove) unless its area was greater than 0.01 times the area of **F**.

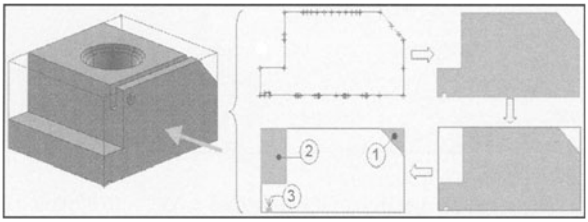


Figure -5. Projection of a polyhedral model into a plane

6. DEFINING THE ORIENTATION

This section aims to define an optimal part orientation towards DFA methodology. The following approach selects features, which maximise the objective function $\mathbf{F_{obj}}$=OE/FC, where **OE** is the orientation efficiency and **FC** is the relative cost feeder. First, in each 2D projection of the part, the orientation capability of each feature is determined. Then, among the combinations of features defining an orientation solution for the part, the one of maximal $\mathbf{F_{obj}}$ value is selected.

6.1 Orientation of a projection

In order to define the part orientation, the proposed algorithm evaluates the inner symmetries and searches for symmetric features for each 2D feature in a 2D projection. Five symmetry attributes (**symX**, **symY**, **has_symX**, **has_symY**, **has_symZ**) are therefore added to the 2D feature data structure, as well as the feature usefulness. Figure 6 illustrates the meaning of these symmetries for 2D features.

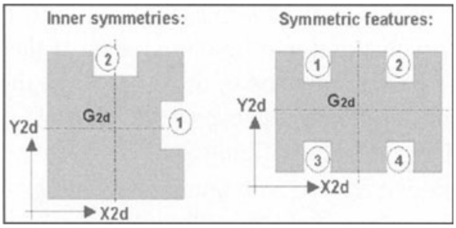


Figure -6. Symmetries used to define the projection orientation

For the 2D projection on the left in figure 6: **1** has X2d inner symmetry (**symX**), **2** has Y2d inner symmetry (**symY**). For the 2D projection on the right in figure 6, attributes **has_symX**, **has_symY** and **has_symZ** are true for features **1**, **2**, **3** and **4**. Symmetry **symZ** is not of interest for 2D features. According to the given definitions, no 2D feature can be symZ symmetry. Only features of type "hole" or "cavity" could have been symZ symmetry but they are not used in this approach due to their insignificance in the DFA methodology. Furthermore, only "hole" or "cavity" features could have been both symX and symY inner symmetries. Table 4 and figure 7 give indication on $\mathbf{F2d}_{\mathbf{ik}}$ capabilities, the $\mathbf{k}^{th}$ feature of the projection **i**, where $\mathbf{i} \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$.

Table -4. Orientation capability for a 2D feature

Case number	Number of inner symmetries	Number of symmetric features	Properties
1	0	0	Orients the part
2	0	1	This feature removes two rotation
	1	0	axes for the part
3	0, 1	2	This feature and one of these symmetric features orient the part
4	0, 1	3	This feature is useless
	1	1	
5	2	0, 1, 2, or 3	Impossible

Figure -7. Part samples for cases n°1, 2, 3 and 4 in table 4 from the left to the right

6.2 Orientation of a part

In each of the three 2D projections, the best DFA feature (case n°1 in table 4), or combination of features (case n°3 in table 4), is selected, i.e. the feature that, in decreasing order of importance, defines the part orientation, has the maximal $\mathbf{F_{obj}}$ value and is of maximal area. If there are one or more such features in the part then, the one of maximal $\mathbf{F_{obj}}$ value is selected as the only main feature for defining the part orientation. But, such features might not exist in a part. The example in figure 8 shows such a part, for which two projections have only one feature with one inner symmetry each.

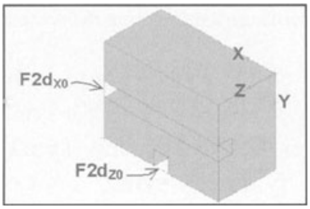


Figure -8. Sample part for which orientation is defined by two projections

The solution consists in using features in more than one projection. From a technological point of view, this approach means that, during the guidance in translation of the part, a second direction of translation would allow the complete definition of the part orientation. From the algorithmic point of view, this consists in searching for a combination of features i.e. case n°2 in table 4. If $\mathbf{F2d_{ik}}$ is such a feature, a 3D characteristic axis, named $\mathbf{Axis_{3D}}$, is defined as the axis around which the part can be 180°-rotated so the face representing the projection **i** stays the same before and after this rotation. Table 5 provides the $\mathbf{F2d_{ik}.Axis_{3D}}$ value, which is one of the 3D coordinate system of the optimal bounding box, corresponding to a symmetry axis (inner symmetry and/or symmetric feature) of $\mathbf{F2d_{ik}}$.

Table -5. Connection between 2D and 3D coordinate systems

Projection name	2D symmetry axis	3D axis
	X2d	Y
ProjX	Y2d	Z
	G2d	X
	X2d	X
ProjY	Y2d	Z
	G2d	Y
	X2d	X
ProjZ	Y2d	Y
	G2d	Z

For the part in figure 8, ($\mathbf{F2d_{X0}.Axis_{3D}}$ = **Z'**) ≠ ($\mathbf{F2d_{Z0}.Axis_{3D}}$ = **Y'**) therefore, by selecting these two features, the part cannot be 180°-rotated around any of the three axes (**X'**, **Y'**, **Z'**), and its orientation is defined without ambiguity.

6.3 Further geometric problems

Sections 6.1 and 6.2 provide guidance for selecting an optimal DFA 2D features combination. But, they also point at a 3D feature recogniser usefulness. The current implementation of "FuzzyDFA", an assembly-oriented CAD software, which uses a fuzzy decision support system (FDSS) and the Boothroyd-Dewhurst methodology, already takes advantage of the above-described algorithms and, in the near future, of a 3D feature recogniser. The former aims to find 3D notches, holes, pockets, grooves and fillets (non visible in 2D projection). Recognition algorithms are based on pre-existing methods [5]: concavity of edges (material angle between two faces), concavity of underlying parametric curves of edges, inner loops of faces and local topology analysis. Figure 9 summarises 3D features of interest in the context of DFA.

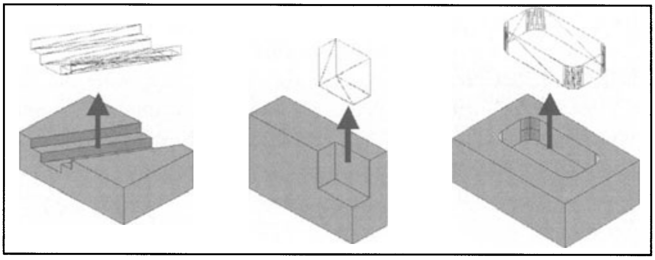


Figure -9. Examples of 3D groove, notch and pocket to be recognised

"FuzzyDFA" was built in Visual C++ and Open CASCADE [9], which is a powerful 3D modelling kernel that consists of reusable C++ object libraries available as Open Source. This application enables the creation of 3D assemblies, as well as the DFA definition of each part, assuming that its geometry has already been modelled in the STEP AP203 standard. Moreover, DFA evaluation can be computed, either on a time/cost basis or on a merit scale, for manual and automated operations.

7. CONCLUSION

In this paper, most of the geometric DFA requirements are pointed out. Weight, dimensions and symmetries are sufficient characteristics for computation in manual operations. For automated assemblies, we focused on the usage of bowl feeders, which were investigated in depth by Boothroyd-Dewhurst. These devices operate on 2D features seen in projection of a part. Herein, we detailed how to find these features and described an original approach for defining the orientation of a part. The use of two geometric models (B-Rep and polyhedral) allows algorithmic effectiveness, taking advantages of both models while avoiding their inconveniences. Every geometric characteristic serves as input for the fuzzy decision support system used in "FuzzyDFA". Therefore, evaluation of parts may be performed early in the design process, even if the design of the part is not detailed or remains uncertain, thanks to the use of fuzzy logic. Now, a 3D recogniser would complete geometric requirements for most DFA methodologies and this is being implemented.

REFERENCES

[1] G. BOOTHROYD, P. DEWHURST, W. KNIGHT, *"Product design for manufacture and assembly"*, Dekker, ISBN 0824791762, 1994.

[2] M. ROSARIO, *"Automatic geometric part features calculation for design for assembly analysis"*, PhD. Thesis, Univ. of Rhode Island, 1988.

[3] X.-Y. JIANG and H. BUNKE, *"A simple and efficient algorithm for determining the symmetries of polyhedra"*, Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing, vol. 54, no. 1, January 1992, pp. 91-95.

[4] A. FEBRANSYAH, *"A Feature-based Approach to Automating High-Level Process Planning"*, PhD. Thesis, North Carolina State Univ., 2001.

[5] S. PARRY-BARWICK, A. BOWYER, *"Feature technology"*, University of Bath, Technical Report, January 1993.

[6] M. BERN, D. EPPSTEIN, *"Mesh Generation and Optimal Triangulation"*, Computing in Euclidean Geometry, Ding-Zhu Du and Frank Hwang (editors), World Scientific, Singapore, 1992, pp. 23-90.

[7] O. COMA, C. MASCLE, P. VERON, S. MOISAN, *"Application de la logique floue à une méthode de conception pour l'assemblage"*, 4th Int. Industrial Engineering Conference, Aix-Marseille, 2001, pp. 109-118.

[8] AMES A. L., *"Production ready feature recognition based automatic group technology part coding"*, Proceedings of the first symposium on Solid Modelling foundations and CAD/CAM applications", 1991, pp. 161-169.

[9] OPENCASCADE, http://www.opencascade.com

AN A PRIORI ADAPTIVE 3D ADVANCING FRONT MESH GENERATOR INTEGRATED TO SOLID MODELING

V. François and J-C. Cuillière

Laboratoire Interdisciplinaire de Recherche en Imagerie et Calcul Scientifique. Département de Génie Mécanique, Université du Québec à Trois Rivières, Canada

Abstract: We have developed the nodal pre-optimization concept in the context of our work on finite element analysis and CAD/CAM integration. This concept consists in deriving from all engineering data (geometric features, discretization error, boundary conditions, materials and physical constants) a nodal density map for automatic mesh generation. Nevertheless, this concept can only be applied if a mesh generator is able to satisfy precisely the density map mentioned above. This is the reason why we have developed a 3D automatic mesh generator based on advancing front techniques, featuring a strong ability to respect an imposed variable density map. After a brief description of the process, we introduce new parameters related to the density map respect. A set of applications of the concept to industrial mechanical parts are also presented in the paper.

Key words: CAD/CAM, solid model, automatic meshing, density map, adaptive mesh.

1. INTRODUCTION

The recent and rapid development of computer systems has allowed significant improvements in the resolution of mechanical problems by numerical analysis. Among these, the finite element method (FEM) has led to a surge in demand. This method which, ten years ago, was only present on large computer systems, is now available on all personal computers. The availability of this method has shown a dramatic increase and is expected to

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 337–343.

rise in the upcoming years. This democratisation of the method brings about more and more sophisticated demands from users. A lot of questions remain unanswered. Among these, a certain phase gives rise to concern : data entry under digital simulation software by finite element method. Although this phase is not technically complicated, it is lengthy and fastidious. As a consequence, the productivity of such product has slowed down. The problem originates from the fact that the finite element software does not provide performing modeling tools. Thus, these performing tools are part of the CAD domain. The idea of combining the two domains was immediate. This link already exists in many business design solutions but it is static and lacks intelligence. This link basically consists in transforming the CAD domain into a valid meshing for the finite element method. Efforts have been made to improve this reality for several years. We have introduced the nodal density pre-optimisation concept which is summarised in the following section. This concept involves the necessity to resolve a multitude of problems pertaining to automatic mesh generation. This paper demonstrates 3D automatic meshing methods of variable nodal density.

2. THE PRE-OPTIMISATION CONCEPT OF NODAL DENSITY

The transition phase between the CAD and FEM domains is performed by defining a unique model[1] based on two distinct models to which the bilateral links are added. To realise this model, several tools have been developed in order to obtain an integrated mesh generator[2]. The pre-optimisation concept of nodal density is one of them. The nodal density pre-optimisation concept has been constantly evolving for several years[1,2,3]. Its aim is to transform any type of data in terms of size map (figure 1).

The calculated size map must further be recognised by the mesh generator to obtain a final meshing of the pre-optimised domain. Data taken into account in this concept are of different kings :

- shape features of the geometric model[1,4]
- the discretization error of the geometric model inherent in the finite element method[3,5]
- boundary conditions and efforts applied to the model
- the material and physical constants of the finite element problem

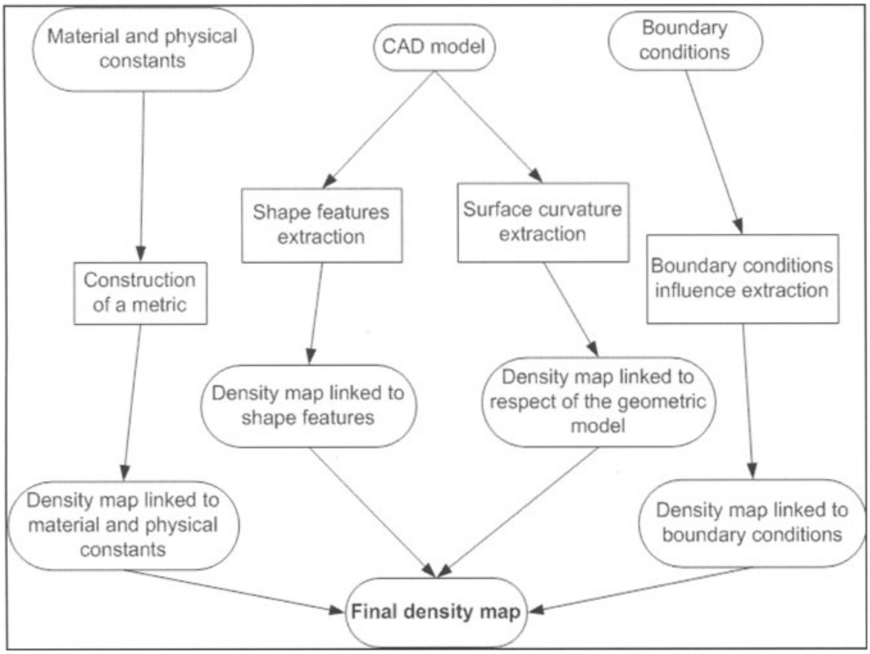


Figure -1. The pre-optimisation concept of nodal density[3]

The two last aspects may lead to the definition of a tensorial size map (anisotropic material, fluid flow). We are now speaking of metric. The concrete application of this concept as stated, involves mesh generators, with the ability to respect as closely as possible, the size map or the metric calculated. Several open-ended problems are however, inherent in the implementation of this concept. As an example, all mesh generation levels must be developed in order to take the new constraints into account. This is why, the different meshing methods have been adapted :

– Optimal mesh edge with respect to (conforming to)a size map[6]
– Surface mesh with respect to a size map[5,7]

Other methods are in the process of adaptation :

– Volume mesh with respect to a size map
– Surface mesh with respect to a metric
– Volume mesh with respect to a metric

The goal of this paper consists in explaining the problem of a given volume mesh with respect to a size map. After reviewing various automatic meshing techniques, we will demonstrate the modifications carried out in order to reach our goal.

3. THE CLASSIC VOLUME MESH

The different mesh generation methods fall into the two following categories :

– The so-called 'Delaunay" method[8,9,10,11]
– The advancing front techniques[3,12,13,14,14,15,16,17]

Our ultimate choice was the advancing front techniques in order to achieve other objectives which are not exposed here[3]. The following is a simple description of a 3D front mesh basic algorithm :

1. initialisation of a front on the surface triangular mesh of the volume skin,
2. classification of the front,
3. selection of the first front element called candidate element ,
4. calculation of the ideal location of the node called candidate node,
5. searching for existing nodes close to the candidate node,
6. classification of the selected nodes ,
7. creation of a valid tetrahedral element with the first node that allows it,
8. front update,
9. if the front is not empty return to 3.

This algorithm presents several aspects which do not have a theoretical solution but rather several trial-and-error-solutions which are sometimes antagonistic : steps 3, 4 et 6 are specific to each 3D front mesh.

The base of our mesh respects this algorithm and the trial-and-error-solutions used derive from those of Golgolab[14].

First of all, a classification rule of the front must be established. Golgolab classifies the front in a descending order of the irregularities weighted by the dimensions. Secondly, the calculation of a node's ideal location must be established. The Golgolab solution consists in considering exclusively the candidate element to find the co-ordinates of the candidate node : it is located at 1/3 of the candidate element's perimeter length on the normal line passing through its barycenter. Thirdly, a classification of the candidate nodes for the creation of a valid tetrahedron must be implemented. Golgolab defines several interesting concepts to differentiate the nodes surrounding the candidate element (figure 2). Thus an adjacent node is a node connected to the candidate element via another triangle, a linked node is a node connected to the candidate element via a meshing edge and an neighbouring node is a node surrounding the candidate element which is neither adjacent nor linked. The classification of the nodes to form a new tetrahedron is as follows : adjacent node, linked node, neighbouring node, ideal node.

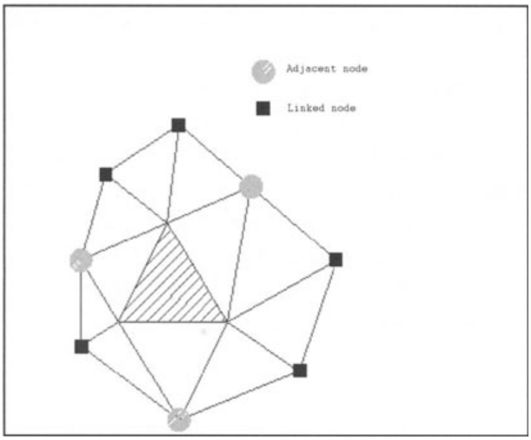


Figure -2. The node classement criterion[14]

The algorithm thus defined practically never converges. There are some situations where the configuration of the front, does not allow for the construction of a valid tetrahedron. It is often the case with the last mesh iterations. It is then necessary to anticipate destruction of the mesh in order to allow evolution of the front. This tetrahedron destruction mechanism must be accompanied with a mean allowing to avoid a tetrahedron construction destruction infinite loop. To succeed in recognising the size map, other constraints must be added. The Golgolab solution, however seems generated by a compromise between convergence of the algorithm and the number of constraints imposed. The whole difficulty in considering the density imposed consists in adding several constraints in order to promote conformance to the size map without however altering the convergence of the 3D front mesh.

4. THE VOLUME MESH OF VARIABLE NODAL DENSITY

By studying the Golgolab algorithm previously described, we notice that the control central axis of the process is searching for convergence. The notion of conformity with the prescribed nodal density is less significant. In fact the recognition of the size map can only be realised in the ideal point location calculation phase. With the node classification indicator used, this case holds the lowest priority. In other words, a node is created, if and only if, there is no other solution. To reach our goal, the prior use of an ideally positioned node must be increased. The clue to the meshing software development is therefore to increase this priority while maintaining the ease of convergence developed by Golgolab. For this, the notions of adjacent, linked and neighbouring nodes are redefined in order to reduce the potential number of nodes available during the classification phase of the selected nodes. In addition, a new parameter is defined : The degree of a node represents the number of connections of the same type as that of a node with the candidate element. Thus, Golgolab's notions are refined as follows :

- An ADJACENT3 node is a 3-degree adjacent node. It is connected to the candidate element by three triangles,
- An ADJACENT2 node is a 2-degree adjacent node. It is connected to the candidate element by two triangles,
- An ADJACENT1 node is a 1-degree adjacent node. It is connected to the candidate element by one triangle. Also the minimum angle between the candidate element and the connection triangle must be lower than $4\pi/9$
- A LINKED3 node is a 3-degree linked node. It is connected to the candidate element by three edges,
- A LINKED2 is a 2-degree linked node. It is connected to the candidate element by two edges. Also the minimal angle between the candidate element and the connection edges must be lower than $4\pi/9$,
- A LINKED1 is a 1-degree linked node. It is connected to the candidate element by an edge. Also the angle between the candidate element and the connection edge must be lower than $4\pi/9$,
- A GENERAL node is a node in an ideal location (ideally positioned node),
- All remaining nodes are not taken into account. The notion of adjacent node is particularly abandoned.

By following these rules, the number of nodes selected is significantly reduced in relation to the Golgolab solution. The angular restriction imposed is crucial. The classification order is as follows : ADJACENT3, ADJACENT2, ADJACENT1, LINKED3, LINKED2, LINKED1, GENERAL. The calculation of the candidate node's location is also refined. The edge lengths of the tetrahedron to be formed are calculated in two ways:

- the first value d_1 indicates the quality aspect of the mesh and equals the average length of the three candidate triangle edges,
- the second value d_2 indicates the conformity with the prescribed size map and equals the average of the size map value at the three nodes of the candidate triangle.

The length determined d is thus :

$$d = (1-\alpha)d_1 + \alpha d_2 \tag{1}$$

α is a user parameter having a 0.65. default value.

The maturing phase of the front is not easy to perform. Our 2D experiment[18] was not sufficient. In fact, in 2D, the number of possible configurations during the meeting of the front is limited to one. In 3D, performing this count seems difficult. Another solution is recommended :

an automatic update of the front. The problem is to find the neighbouring front elements of a prescribed element. A general procedure automatically transforms a list of triangular elements into an oriented front where all the neighbourhood node relationships are established. As a result, the constituted front allows to determine immediately the nature of a node in relation to a candidate element. Thus, for each selected and classified node, a corresponding tetrahedron construction operation exists.

According to Golgolab[14] and as Rassineux[17] confirms, a tetrahedron destruction procedure is necessary to obtain a definite convergence. In fact, the algorithm previously described practically never converges. This destruction phase must be inserted without neglecting a priori respect to a prescribed size map. Thus, the ultimate choice is to destroy all tetrahedrons preventing the formation of a tetrahedron from the candidate element and a GENERAL node. To avoid these destruction construction infinite loops of a same tetrahedron, a destroyed tetrahedron cannot be reconstructed. A front management by layers is performed in order to facilitate closure of the front. Each new front element created during the process is placed in a queue. When the front is empty, it is reinitialised from the queue entries. The process is completed when the front and the queues are empty. The front closure is performed at the heart of the solid. It is the location where there are less geometric constraints. The front can close itself more easily.

Finally, a candidate element may not bear a solution even with the active destruction procedure.

It is notably the case where a boundary triangle prevents the formation of a mesh. The solution determined is to put the element concerned in the queue described previously. This problem can further be solved in two ways

– the queue element can be absorbed by the treatment of another front element
– the queue element may again become a candidate element. To perform this treatment, the equation (1) is modified to obtain

$$d = \frac{(1-\alpha)d_1 + \alpha d_2}{\text{number of times that this element was a candidate element}} \quad (2)$$

5. RESULTS

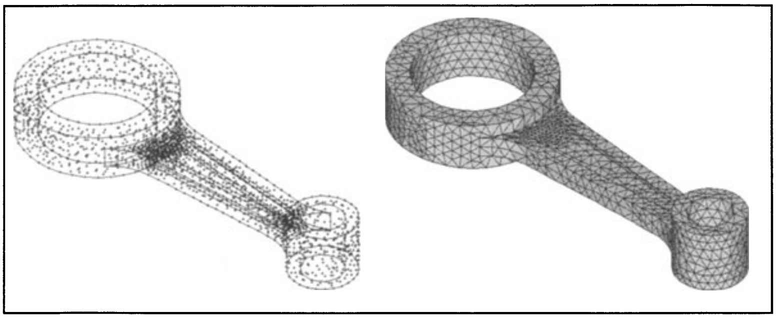


Figure -3. Mesh pre-optimization of a crank arm by controlling the discretization error[3]

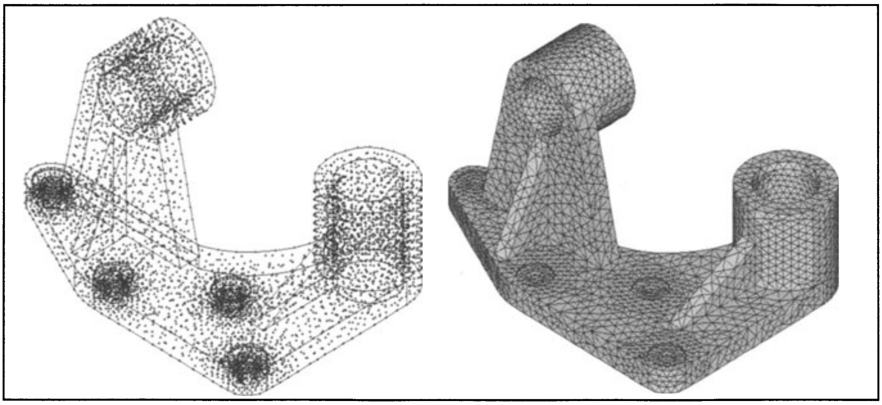


Figure -4. Mesh pre-optimization of a lever bracket by controlling the discretization error[3]

Table -1. Results of the two sample parts

Parameters	Crank arm	Lever bracket
Number of nodes	2232	7128
Number of tetrahedrons	7959	29268
Minimum quality	0.15	0.1
Maximum quality	0.99	0.99
Mean quality	0.61	0.61
Theoretical number of tetrahedrons	8422	31732
Error	5.5%	7.7%
CPU (Pentium III 650 MHz) (s)	36	188
CPU old method	92	272

Examples of results on two mechanical components are shown in figures 3 and 4. In fact, these results complete the former results concerning surface meshing with respect to a size map[2,3,5]. The size map used is the one developed in the discretization error context. Statistics are gathered in table 1.

The quality indicator used is the sphere indicator marked in the tetrahedron and the theoretical number de tetrahedrons is calculated according to the method developed by Cuillière[7].

The CPU time is also recorded and compared to the CPU time of the method used previously[3,17]. The latter method does not respect the prescribed size map.

6. CONCLUSION

The results presented, together with the other results that we have, demonstrate the efficiency of the developed meshing method. The compliance error of the size map is always lower than .. This result is better than the one obtained for surfaces[2]. This can be explained by the fact that geometric constraints are less significant in a solid than on a surface. In addition, the time required to obtain a final mesh network with respect to a size map is much shorter than with the tool that we used in the past.

This work allows to conclude the nodal density pre-optimisation process based on the geometric discretization error. In addition, it paves the way to other solutions. Anisotropy can be produced by adapting this mesh. To do so, we need to introduce a new length calculation formula. Our present work moves towards this direction.

ACKNOWLEDGEMENTS

This research was conducted as part of a project supported by reseach funding from the *Université du Québec à Trois Rivières (UQTR)* and the *Natural Sciences and Engineering ResearchCouncil of Canada (NSERC).*

REFERENCES

1. J. C. CUILLIÈRE. Pré-optimisation de maillages automatiques tridimensionnels pour les méthodes numériques application à l'ingénierie simultanée. Thèse de doctorat, Institut National Polytechnique de Lorraine, 1993.

2. V. FRANÇOIS, J. C. CUILLIÈRE AND M. GUEURY. Méthodes de remaillage automatique appliquées à la modification de modèle. Revue Européenne des éléments finis, 8(4) pp. 433–454, 1999.
3. V. FRANÇOIS. Méthodes de maillage et de remaillage automatiques appliquées à la modification de modèle dans le contexte de l'ingénierie simultanée. Thèse de doctorat, Université Henri Poincaré, Nancy 1, Novembre 1998.
4. J. C. CUILLIÈRE AND R. MARANZANA. Automatic and a priori refinement of threedimensional meshes based on recognition techniques. Advances in Engineering Software, 30 pp. 139–149, 1999.
5. V. FRANÇOIS AND J. C. CUILLIÈRE. Automatic mesh pre-optimization based on the geometric discretion error. Advances in Engineering Software, 31 pp. 763–774, 2000.
6. J. C. CUILLIÈRE. A direct method for the automatic discretization of 3D parametric curves. Computer Aided Design, 29 pp. 639–647, 1997.
7. J. C. CUILLIÈRE. An adaptive method for the automatic triangulation of 3D parametric surfaces. Computer Aided Design, 30 pp. 139–149, 1998.
8. J. C. CAVENDISH, D. A. FIELD AND W. H. FREY. An approch to automatic threedimentional finite element mesh generation. International Journal for Numerical Methods in Engineering, 21 pp. 329–347, 1985.
9. K. KONDO. Three dimentional finite element meshing by incremental node insertion. International Journal for Numerical Methods in Engineering, 39 pp. 3519–3535, 1996.
10. R. W. LEWIS, Y. ZHENG AND D. T. GETHIN. Three dimentional unstructured mesh generation : Part 3. volume meshes. Computer Methods in Applied Mechanics and Engineering, 134 pp. 285–310, 1996.
11. P. L. GEORGE AND H. BOUROUCHAKI. Triangulation de Delaunay et maillage : applications aux éléments finis. Hermes, 1997.
12. R. LOHNER AND P. PARIKH. Generation of three-dimentional grids by the advancing front method. International Journal for Numerical Methods in fluids, 26 pp. 2135–2159, 1988.
13. J. PERRAIRE, J. PEIRO, F. FORMAGGIA, K. MORGAN AND O. C. ZIENKIEWICZ. Finite element euler computations in three dimensions. International Journal for Numerical Methods in Engineering, 26 pp. 2135–2159, 1988.
14. A. GOLGOLAB. Mailleur 3D automatique pour des géométries complexes. Rapport technique, INRIA, 1989.
15. S. W. CHAE AND K. J. BATHE. On automatic mesh construction and mesh refinement in finite element analysis. Computers and Structures, 32 pp. 911–936, 1989.
16. P. MOLLER AND P. HANSBO. On advancing front mesh generation in three dimensions. International Journal for Numerical Methods in Engineering, 38 pp. 3551–3569, 1995.
17. A. RASSINEUX. Maillage automatique tridimensionnel par une méthode frontale pour la méthode des éléments finis. Thèse de doctorat, Université Henri Poincaré, Nancy 1, Janvier 1995.
18. V. FRANCOIS, J. C. CUILLIÈRE AND M. GUEURY. Automatic meshing and remeshing in the simultaneous engineering context. Research in Engineering Design, 11 pp. 55–66, 1999.

A COLLABORATIVE TOOL FOR THE VISUALIZATION OF SCIENTIFIC SIMULATION RESULTS

D. Lovinfosse, L. Fine+, J-C. Léon, X. Rogovitz* and T. Truong-Trung*
Laboratory 3S, Grenoble, France
+Geonuméric, Grenoble, France
**EADS Centre Commun de Recherches, Antenne de Toulouse, France*

Abstract: Numerical simulations of mechanical phenomena using the Finite Element approach are extensively used in industry. On one hand, the FE models manipulated are evolving toward a high complexity in terms of size (number on FE elements in the model) and phenomena addressed (non linear behaviour laws, time dependent simulations, ...). On the other hand, design methodologies are evolving towards computer supported collaborative work. The paper proposes a concept of a visualization model for the analysis and collaborative work around FE simulations results. The model presented takes into account the configurations and the equipment characterizing the use of these results and is based on a decimation technique to reduce the size of the FE model. To compress the model even further, it is covered with textures generated from the FE solution. In order to generate this visualization model in a transparent manner for the analyst, automatic partitioning methods have been set up. The compactness of the model is addressed using a multi-resolution approach on user-defined threshold values to extract the most significant part of the simulation results.

Key words: collaborative tools, visualization, scientific results, multi-resolution models, textures

1. INTRODUCTION

In the context of scientific visualization of simulation datasets, it is important to provide tools that fit into the design flow. This remark applies particularly to the field of simulation of mechanical phenomena using a

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 347–358.

Finite Element (FE) technique and contributes to the integration of FE techniques into the design process. Up to now, two distinct needs have been identified in this context:

- The engineer needs specific approaches to obtain models adapted to the visualization capabilities of workstations so enabling him (resp. her) to get the best possible understanding of the phenomenon simulated. Currently, the FE models are becoming larger and larger and their visualization strictly relies on the straightforward visualization of these models. Even with high-end workstations, these models become difficult to visualize and don't allow the engineer to analyze and deeply understand a physical phenomenon because there is no possibility of animation when transient or modal analyses are performed. As a result, a model dedicated to the visualization phase has appeared as critical to produce a compressed model that still contains the significant information of the simulation while it is significantly compressed. Such a model becomes also a basis for co-located collaborative work during digital mock-up reviews to provide explanations to non specialists,
- The engineer needs to collaborate with other engineers of the same skills during the design process through distant synchronous or asynchronous work. At present, such an activity is not possible when FE models are large because they can't be sent through the mail and they are not compatible with synchronous collaborative tools. Here again, the compression of such FE models becomes critical to exchange data over an intranet/internet network.

As a result, the objective is to present some approaches that lead to compression of simulation datasets through the use of control parameters that are compatible with the analysis activity performed by an engineer. The compression process should be transparent for the user and hence, automatically produce a visualization model that can be effectively used for the analysis task.

Our method relies on two decimation algorithms. The first one is dedicated to the compression of the geometry and is based on a unique scalar parameter related to the chordal distance between the initial and the simplified model. The second one is dedicated to the scientific data. It performs data compression according to thresholds set by the user after the FE numerical computations and prior to any visualization of the results using a histogram of the scientific values computed.

Different ways of blending together the compression of the geometry with the compression of the data attached to the geometry are presented. Each way gives an idea of the compression rate that can be obtained while providing as much information as possible to contribute to the collaborative work. We propose two approaches that contribute to the integration of Finite Element (FE) techniques into the design process:

- Compression based on a blend of geometrical and numerical criteria to illustrate how control parameters can be set up from a user point of view.
- Compression with priority to the geometry and use of textured models to illustrate higher order compression.

These approaches are illustrated on real world examples of simulations. Decimation approaches have been largely developed over the past years [1, 2, 3, 4, 5, 6] to set up mainly two categories of operators, i.e. vertex removal or edge collapse. These criteria and operators have produced different categories of decimators capable of restoring the initial shape of an object that has been arbitrarily triangulated. Bounding the deviation between the initial and decimated triangulations has appeared critical for many applications in the context of mechanical engineering and design. To this end, the approach by Véron and Léon [4, 5] uses a concept of error zone attached to each vertex of the initial triangulation to model a discrete envelope of variable distance around the object.

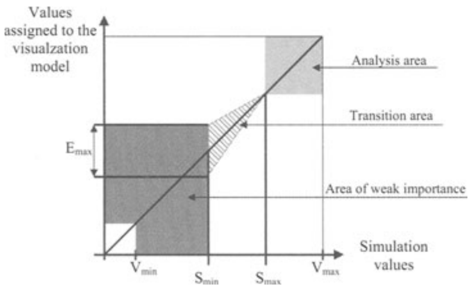


Figure -1. Threshold values of simulation data used to control the compression process.

All these approaches are based on a similar concept of preserving, as much as possible, the shape of the initial object. However, combining textured models and decimation techniques is currently emerging [7].

The next section will present the control parameters proposed for the compression phase. Section 3 will describe how textures can be used to further increase the data compression. Results and example will be described at section 4.

2. CONTROL PARAMETERS FOR THE SCIENTIFIC DATA PREPARATION

Because the purpose of the visualization model can be considered as the basis of the analysis process of a physical phenomenon, the engineer is interested in preserving roughly the shape of the mesh used for the FE computation and then, concentrating on the values of the numerical simulation process and the areas where critical values are located. Several approaches, initiated by EADS CCR (Centre Commun de Recherches), have already been addressed to generate compressed models [8, 9].

In the context of mechanical engineering, this principle is set up using conceptually different areas. These areas are identified automatically from the numerical simulation values assigned at faces of the FE mesh.

Prior to any visualization, the engineer can access the histogram of simulation values using the file output of the computation phase. According to Figure 1, the values obtained are V, $V \in [V_{\min}, V_{\max}]$. Then, the engineer can specify a first threshold value $S_{\max}$, $S_{\max} < V_{\max}$ such that in the interval $]S_{\max}, V_{\max}]$ the simulation values should not be altered by the decimation process to enable a rigorous analysis of the results. This interval defines the "area of interest" for the engineer and no compression should take place there. Therefore, this area is not modified from the simulation point of view. In addition, the engineer can specify a second threshold value $S_{\min}$, $S_{\min} < S_{\max}$, to define new areas where the decimation process will take place as a multi-criteria process. At present, the approximation criterion for the simulation data located around a candidate vertex is similar to an energy preserving criterion [9]. Other data compression can be performed using wavelets approaches [10, 11, 12, 13] and form current directions for future developments. Within the transition area depicted in Figure 1, the simulation values attached to the visualization model must be kept into the interval $[V - E(V), V + E(V)]$, $V \in [S_{\min}, S_{\max}]$ where the function $E(V)$ is linear within the transition area. Hence, the transition area is characterized by the deviation over the simulation results that the engineer can accept in the interval $[S_{\min}, S_{\max}]$. Such a behavior is described by two parameters $E_{\min}$ and $E_{\max}$ (Figure 1).

As a result, the visualization model in the context of collaborative work is generated through a multi-criteria decimation approach. The multiple criteria involve the above parameters for the scientific criterion and a distance parameter ε_g. The decimation process is applied as follows to select the areas where the compression will take place:

- Vertices are selected using a geometric criterion to ensure the preservation of the object shape,
- The contour of the vertex removed is replaced by a new set of faces with new simulation data attached to these faces,

- The geometry restoration criterion is checked. If the new triangulation lies within the error zones of radius ε_g the remeshing scheme is validated otherwise this vertex is not removed,
- The approximation criterion for the simulation data is checked. If this approximation lies in the interval $\left[V_{init} - E(V_{init}), V_{init} + E(V_{init})\right]$, the candidate vertex is effectively removed otherwise it is not.

Finally, a third area is considered, which designates a region of low interest for the engineer, i.e. areas of low stress levels, … In this area $V \in \left[V_{\min}, S_{\min}\right[$, the decimation process is applied with the geometric deviation criterion ε_g only and the simulation data are propagated without any further constraint using the energy preserving criterion [9]. The geometric decimation process is set up to handle non-manifold models as required by the FE models currently set up for simulation purposes [14, 15].

As a result, the control parameters set up by the engineer are: ε_g, $S_{\min}$, $S_{\max}$, $E_{\max}$ and eventually $E_{\min}$ if a non-zero value is wanted. All these parameters are strictly based on the results of the simulation process and easily set up from the histogram of the results. Hence, the compression is adaptative over the triangulation and driven by the simulation results. The ε_g parameter is also easily set by the engineer since it characterizes the chordal deviation between the initial mesh and the final triangulation.

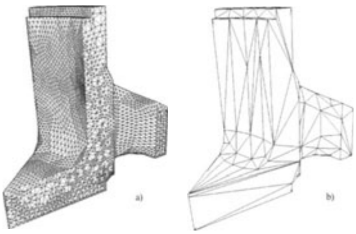
To illustrate the effect of the multi-criteria approach set up in the transition area, an example is shown in Figure 2 where the object is decimated using an approximation of the simulation data at the transition area. The data originate from the modal analysis of a car body. No modification of the simulation data is performed for high values (rather dark colors around the location of the spare wheel in Figure 2). Only the geometric criterion is active in the low values areas (grey and dark grey colors mostly over the front half of the car body in Figure 2), which are not relevant for the engineer because the displacement amplitudes there are fairly small. Figure 2 clearly shows the effect of the multi-criteria decimation process over plane and curved areas.

Figure -2. FE model (left) (238 945 faces) and visualization model (right) (62 282 faces) of a Clio car body structure produced by a modal analysis (Courtesy Renault). The car body is seen from underneath. The geometric deviation is: ε_g =1cm.

3. TEXTURED MODELS FOR COLLABORATION

The previous section has described the parameters and principle of generation of the visualization model used as dialog object for collaborative work purposes. However, the models obtained do not reach a sufficiently high level of compression to enable collaborative activities using current desktop PCs because the model size (number of faces) cannot be dealt with by their level of graphic performance. To achieve higher compression rates and better performances, texture mapping seems a promising solution.

The solution currently set up using textures decouples the geometric compression, i.e. the polyhedron decimation process, from the approximation of the scientific data. Thus, the polyhedron visualized is obtained through the ε_g parameter only to preserve as much as possible the shape of the object. Figure 3 illustrates the result of the decimation process using a simple simulation model. This simplified model will be the basis of the visualization model used to display the simulation results.



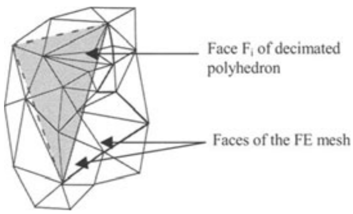
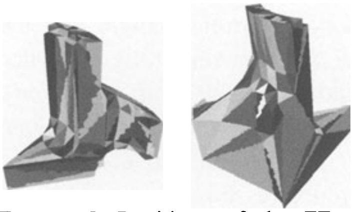


Figure -3. Decimation of a polyhedron using the distance parameter only. a) Initial polyhedron (15996 faces), b) decimated polyhedron (234 faces).
Figure -4. Projection operation of the faces of the FE mesh onto a face F_i of the decimated polyhedron.



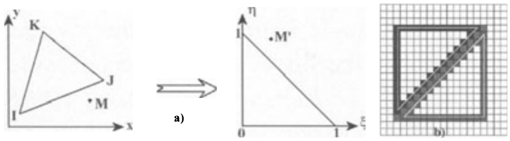


Figure -5. Partitions of the FE mesh onto a face F_i of the decimated polyhedron.
Figure -6. a) Mapping between a partition of the initial FE mesh and a texel of the visualization model. b) Illustration of the discrete representation of two adjacent texels.

The principle of the textured model is to colour each face F_i of the decimated model with a picture representing the distribution of simulation data over the area covered by F_i. To this end, it is necessary to create a relationship between the FE triangulation and the decimated one. This relationship helps defining where the triangles of the FE mesh should be located on the decimated triangulation. This relationship also forms the basic data required to generate the texture of the decimated polyhedron.

Currently, the relationship discussed before is set up using a projection operator from the FE mesh onto the decimated triangulation (see Figure 4). The projection sorts the faces of the FE mesh according to the following categories:

– a face of the FE mesh lies entirely inside the face F_i,
– a face of the FE mesh lies partially inside the face F_i,
– a face of the FE mesh lies entirely outside the face F_i.

As a result of the projection operation, the subset S_i of faces of the FE mesh that either overlap the face F_i or lie entirely inside this face can be identified. Applying this projection operation for each face F_i produces an atlas of subsets S_i that entirely covers the surface of the FE mesh. Due to the fact that a subset S_i contains faces which partly project onto F_i, these faces belong to two adjacent partitions of the atlas. Hence, the partitions partly overlap each other. Figure 5 shows the atlas of partitions obtained from the example part of Figure 3 where each partition S_i is represented by a different colour on the initial FE mesh.

Finally, the projection operation is used to construct the intersection between the faces of the FE mesh, which are partially included in a subset S_i. Thus, the faces F_i of the decimated polyhedron form an intermediate parametric domain where the positions of the faces of a subset S_i can be located to generate the texture required for the visualization. These faces are arbitrarily shaped and located in 3D whereas a texture requires a specific parametric space depending on the mapping technique chosen.

Here, the mapping technique chosen is a 'vertex coordinate mapping' technique, which is well suited for arbitrarily shaped polyhedrons. Using such a mapping, the texture space is reduced to a unit triangle (see Figure

6a). The mapping between the coordinates of polyhedron triangle and the reference triangle is linear as is the reverse mapping between this reference triangle and the displayed textured triangle handled by the graphics library used for the visualization purposes. Hence, the transformation performed does not generate visual distortions. The direct and inverse linear transformations between the points *M* and *M'* are expressed by:

$$x_m = \xi . x_i + \eta . x_j + (1 - \xi - \eta) . x_k , \quad y_m = \xi . y_i + \eta . y_j + (1 - \xi - \eta) . y_k$$

$$\xi_{m'} = \frac{\left((y_m - y_k) - (x_m - x_k) . \left(\dfrac{y_j - y_k}{x_j - x_k} \right) \right)}{\left((y_i - y_k) - (x_i - x_k) . \left(\dfrac{y_j - y_k}{x_j - x_k} \right) \right)}, \quad \eta_{m'} = \frac{\left((y_m - y_k) - (x_m - x_k) . \left(\dfrac{y_i - y_k}{x_i - x_k} \right) \right)}{\left((y_j - y_k) - (x_j - x_k) . \left(\dfrac{y_i - y_k}{x_i - x_k} \right) \right)} .$$

where the *I*, *J*, *K* vertices of the polyhedron triangle are expressed as 2D in-plane coordinates according to Figure 6a. Such linear transformations enable the preservation of a relationship between the displayed model and the initial FE mesh. Thus, it is possible to map any pixel of the displayed model onto the initial FE model, e.g. when the user wants to get access to the effective scientific value at a selected location or wants to interrogate the FE model to obtain element characteristics at that location.

Texels (Figure 6b) are discrete representations of reference triangles shown on Figure 6a and hence of the mapping between them and the 3D polyhedron triangles F_i. Similarly, they incorporate with a discrete representation the mapping of the subset S_i to effectively construct the texture colouring the faces F_i. The resolution, in terms of pixels, and the layout of the texels are important issues contributing to the compression of the model [16].

The configuration currently set up is based on texels of constant resolution, i.e. 100x100 pixels whose layout is shown on Figure 6b. Line segments representing the boundaries of the initial FE mesh are discretized using a Bresenham like algorithm. For visual purposes, graphics treatments are added, like anti-aliasing between adjacent textels, to provide an efficient restoration of the initial FE model. JPEG compression of the textures generated has been applied.

Finally, the coupling between the control parameters described in Figure 1 and the model texturing are merged together to further increase the model compression while preserving the meaningful information areas. The principle is to generate a mixed model incorporating both coloured areas and textured ones, the geometric model being compressed independently using the principle described in Figure 3. Presently, only one threshold, $S_{\max}$, is used to distinguish the textured areas from the coloured ones. If a face F_i contains scientific values greater than $S_{\max}$ it is considered as an area of interest and hence, it is textured. Conversely, if the largest scientific value on a face F_i is smaller than $S_{\max}$ this face is colored for it is considered as an area of low importance. The color assigned to this face is based on an area-weighted averaging process of the scientific values spread over this face.

4. RESULTS

Based on the approach described previously, Figure 7 shows an example where the decimation and texture mapping procedures are applied to generate a 3D textured model incorporating the scientific data displayed in Figure 7a. Figures 7a, b, c represents the initial model as well as details of the initial polyhedron. Figures 7d, e depicts the decimated polyhedron and the fully textured model, which can be compared to the initial model. Depending on the texel size, which is uniform in the current approach, a first level of compression can be obtained from the initial model in terms of geometric model size and scientific data compression. However, the model provided still entirely preserves the scientific data attached to the initial model and is more suitable for use on desktop PCs to perform collaborative tasks.

Figures 7f, g illustrates the effect of incorporating user-defined threshold values to further compress the model for collaborative tasks. All the models presented are based on the decimated polyhedron of Figure 7d, which is the most compact geometric model that fits within a given ε_g value specified by the user. Figures 7f and g enables the distinction between the colored part of the model and the textured one depending on threshold value $S_{\max 1}$ set by the user. A face F_i of the decimated model is replaced by coloured ones if the maximum scientific value attached to this face is smaller than $S_{\max 1}$. The color assigned to F_i is an approximation of the values distributed over this face. This face lies in an area outside the area of interest of the user according to the graph of Figure 1. Depending on the size of the texels, the compression of the initial model reaches 89% while preserving the most representative information both from a geometric and scientific point of view.

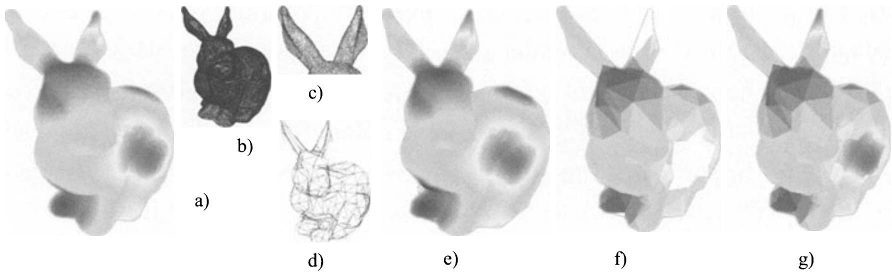


Figure -7. Comparison of a coloured model and a textured one: a) Initial polyhedron with the scientific data attached to the faces (69473 faces). File size 4552 Ko. b) Representation of the initial polyhedron. c) a detail of the initial polyhedron. d) decimated polyhedron with a constant distance parameter (979 faces). File size 221 Ko. e) Entirely textured model using the decimated model to give the same feeling to the user. Texture file size is 2514 Ko and corresponds to a texture size of 16463x303 pixels. ε_g = 1,3%.

Mixed models incorporating textured areas and coloured areas depending on the user threshold specified. All the models are based on the same decimated model in figure b. f) Coloured part of the model displayed on g) using a second user threshold $S_{\max 2}$ to generate coloured faces. File size containing the polyhedron and the coloured faces 179 Ko. Size of the texture file 228 Ko with a texture size of 909x404 pixels. g) Full mixed model based on the decimated polyhedron and a user threshold $S_{\max 2}$. Threshold levels for scientific data are 230 to 250 for e) to define the textured areas. The minimum scientific value is set to 100.

However, the simplified model is still adequate for collaborative work purposes and can provide access to the initial FE results using the relationships between the texture model and the initial triangulation. Although, the transition area mentioned on Figure 1 has not been implemented yet using the mixed approach, the visual transition between the colored area and the textured area appears satisfactory.

5. CONCLUSION

The paper presents the implementation of the visualization aspects of a collaborative tool to analyse and disseminate FE simulations during a design process. This framework is a basis for future enhancements of the model compression performed to incorporate texture compression approaches to improve the quality of the graphic model displayed and evolve toward a multi-resolution model with multiple levels of textures and a progressive behaviour.

Currently, the multi-criteria decimation approach as well as the mixed textured model allow the evaluation of the compression level that can be achieved to produce a model compact enough to be used on small desktop computers. Thus, the compressed models can effectively serve as a basis for collaborative work during the design process to disseminate 3D simulation results among engineers and to enable the use of annotation facilities and the synchronous evaluation of simulations between remote sites. Such models extend the use of models to 3D simulation reports where engineers can access the whole 3D model and simulation data outside the simulation software environment.

Further perspectives include the evaluation of such models within design configurations to address the definition of the user interface and the integration of collaborative functions specific to the simulation context.

The authors would like to thank EADS CCR staff members for their fruitful collaboration over the past four years and their contribution to the approach described as well as in the opening of new research fields.

REFERENCES

1. H. Hoppe, Progressive Meshes, Computer Graphics, SIGGRAPH'96, pp 99-108, 1996.
2. R. Klein, G. Liebich and W. Straber. Mesh reduction with error control, IEEE Visualization'96 Proceedings, pp 311-318, 1996.
3. A. Guéziec. Surface simplification inside a tolerance volume, Technical report RC 20440, IBM Research Division, T J Watson Research Center, New York, USA, Mat, 1996.
4. P. Véron, J-C Léon, Static polyhedron simplification using error measurements, CAD, 29(4), pp 287-298, 1997.
5. P. Véron, J-C. Léon. Shape Preserving Polyhedral Simplification with Bounded Error, Computers & Graphics, 22(5), pp 565-585, 1998.
6. E. Puppo, R. Scopigno. Simplification, LOD and Multiresolution – Principles and Applications, Eurographics'97 Tutorial Notes PS97 TN4, Eurographics association, 1997.
7. P. V. Sander, J.Snyder, S. J. Gortier, H. Hoppe. Texture mapping progressive Meshes, Computer Graphics Proceedings, ACM SIGGRAPH, 2001.
8. J-C. Léon, L. Fine, X. Rogovitz, L. Graux. A Design Review environment for Complex Products, CSCWD99 Conference, University of Compiègne, Compiègne, 28-30 September, 1999.
9. L. Fine, J-C. Léon, X. Rogovitz. Viewer de déformées modales, MICAD 2000 Conference, Paris, March, 2000.
10. P. Schröder, W. Sweldens. Spherical Wavelets. Efficiently Representing Functions on the Sphere, Computer Graphics Proceedings, ACM SIGGRAPH, pp. 161-172, Aug. 1995.
11. G.P. Bonneau. Multiresolution Analysis on Irregular Surface Meshes, IEEE Transactions on Visualization and Computer Graphics, 4(4), pp. 365-378, 1998.
12. G.P. Bonneau, A. Gerussi. Level of Detail Visualization of Scalar Data Sets on Irregular Surface Meshes, IEEE Visualization'98 Proceedings, pp. 73-77, Oct. 1998.
13. G.P. Bonneau, A. Gerussi. Analysis of Scalar Datasets on Multi-Resolution Geometric Models, in Curve and Surface Fitting: Saint-Malo 1999, A. Cohen, C.

Rabut, and L.L. Schumaker (eds.), Vanderbilt University Press., Nashville, pp. 209-218, 2000.

14. L. Fine, L. Rémondini, J-C. Léon. Automated Generation of FEA models through idealization operators, Int. Journal For Num. Methods In Eng., 49(1-2), pp 83-108, 2000.
15. P. Véron, J-C. Léon. Using polyhedral models to automatically sketch idealized geometry for structural analysis, to appear in Engineering with computers, 2001.
16. M. Soucy, G. Godin, M. Rioux. A Texture-Mapping Approach for the Compression of Colored 3D Triangulations, The Visual Computer, 12, pp 503-514, 1996.

SYSTEMS ENGINEERING USED IN PRODUCTS AND MANUFACTURING SYSTEMS DEVELOPMENT

Case of Automotive Industry

E. Lardeur, J-C. Bocquet and C. Auzet*
Ecole Centrale Paris, France
**PSA Peugeot Citroën, France*

Abstract: Systems Engineering (SE) was born in 1960's at NASA to solve issues due to the increasing complexity of products developed in the space industry. Nowadays, the automotive industry is adapting and introducing SE. In addition to early applications, car production basically provides a complex and mass product, which is also characterized by high diversity. This specificity obliges us to review the role of manufacturing systems in the SE methodology described in international standards. This article has three major aims: the first aim is to define SE's span versus other industry's design methodologies, comparing concepts and scopes; the second aim is to give an approach of SE applied to manufacturing systems, especially requirement engineering and architectural design; the third aim is to highlight benefits expected from a level-defined application of SE.

Key words: Systems Engineering, design, design methodology, automotive industry, manufacturing systems.

1. INTRODUCTION

The automotive industry is now confronted with the increasing mechanic and electronic complexity of its products. The issue is to develop new vehicles with a high level of innovation, in a very competitive context, illustrated mainly by shorter development cycle time-to-market.

These reasons explain why Systems Engineering (SE) has been adopted in the automotive industry.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 359–368.

The source of SE description in this context is early applications that emerged via complex equipment and sub-systems suppliers in the aeronautic and aerospace industry: such as in the commercial aircraft field. “Automotive SE Application Framework” [1] has now appeared. We speculate on the idea that criteria of volume and variety may justify the introduction of Manufacturing Systems Engineering in the automotive framework.

This study proposes a first bridge between SE and other design methods, despite the origin of SE, surely based on existing theories. SE isn’t usually mentioned in other references. They are not openly dedicated to the SE community. Few articles from the SE community deal with existing design theories [2]. The area of interest of the Annual Symposiums of the International Council on Systems Engineering (INCOSE) shows similarities with the areas of interest of the International Conference on Engineering Design (ICED). The main difference seems to be that SE focuses more on Industry than on Research.

Section 1 links SE methodologies and existing design theories. Section 2 gives therefore an introduction of SE, implementing the design methods integration point of view. Section 3 deals with the theories we have developed to introduce Manufacturing System Engineering in our framework. As a conclusion, the last section indicates benefits expected from the application of SE and introduces studies to come.

2. SE METHODOLOGY

2.1 Introducing SE

SE is now deployed as a whole methodology, used to develop and manufacture large-scale systems. In our industrial context, systems are the Products and the related Manufacturing Systems. SE can also be considered as a way to deploy the best practices around the company engineering description.

SE describes a maximalist view of all the processes performed to develop a complex system, considering its entire life-cycle and systems implied as a consequence. SE is based on a structural breakdown of the system into sub-systems and includes the project management.

The whole description [3] should include enterprise processes, project processes and technical processes performed by a given organization, and agreement processes between two organizations, as shown in Figure 1.

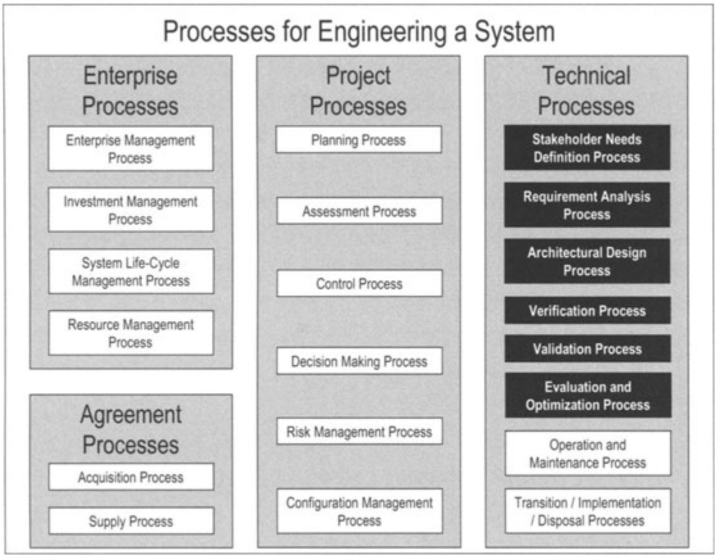


Figure -1. The processes in the system life cycle

The reason why we call the SE description by processes as maximalist is that all the tasks required to develop a system are entirely described in the processes, with a sufficient level of abstraction as to be well understood by all of the engineers from any dept.

The aim of this paper is to focus on design processes and processes that support the design, as shown in Figure 1 (see processes colored in black): the tasks included in the three design processes of SE are systematically supported by tasks described in the Verification and Validation Process (V & V) and in the Evaluation and Optimization Process (E & O).

2.2 The Methodological Integration

Considering common standpoints of design processes, studies on methodological integration only deal with essential phases of design: the "four C": collect, create, construct and product, for Cavallucci [4]; and a minimalist design process for Martin [5].

Both these works operate with a survey of existing methods, extracting value–added for the design process and give the ideal combination of methods to maximize the efficiency of engineers.

Considering now that our company framework provides a detailed description of the logical order of design tasks, we must revise the early hypotheses of methodological integration.

The following table (Table 1) gives an idea of potential contributions to SE processes.

Table - 1. Potential Contributions of Design Methods to Design Processes in SE

Design Methods	Needs Definition	Requirements Analysis	Functional Design	Physical Design	V & V	E & O
Value Analysis	*	*	**	*		**
TRIZ (ARIZ)		*	**	*		*
QFD	**	**		*		*
Axiomatic Design		**	**	**		*
Robust Design	*	*		*	**	*
FMEC Analysis				**		
Prel. Risk Analysis		*	*			
Fault & Event Trees			*	*		
...						

Three types of marks identify where a design method : brings an imperative contribution (**), a desirable contribution (*) or doesn't bring any contribution (empty boxes). The interest of this approach is not restricted to the production of an intelligent toolbox of design methods. As described in quoted references, direct benefits can be extracted considering focal points and coverage of methods. Identifying exactly which tasks call a method, how, why, etc. also gives their consistency to SE Design Processes and is a way to increase efficiency of development.

"Feeding" SE with existing design methods is not a one-way view, both will enhance mutually.

2.3 Mutual enhancement of SE and existing Design Methods, example of Axiomatic Design (AD)

The three design processes of SE shown Figure 1 give birth to four domains: Needs, Requirements, Functional domain and Physical domain.

As done in Axiomatic Design (AD), mapping and zigzagging can be done among all theses domains.

- Mapping: our implementation of SE currently advocates the use of matrices to ensure traceability of engineering's data and have a visual control of the coupling between them.
- Zigzagging: our approach affects first the functional and the physical decomposition of the systems, where we can find sub-systems to which our processes are applied recursively.

We face here too many specificities and disparities of knowledge and practices to describe exactly how to operate the decompositions, although SE provides the best framework to manage decomposition. Furthermore, the decomposition is restricted to the systems: for instance Requirements aren't yet treated with explicit decomposition, as it can be done with AD. Very schematic descriptions of each method brings a good idea of the similarities, as shown in Figure 2.

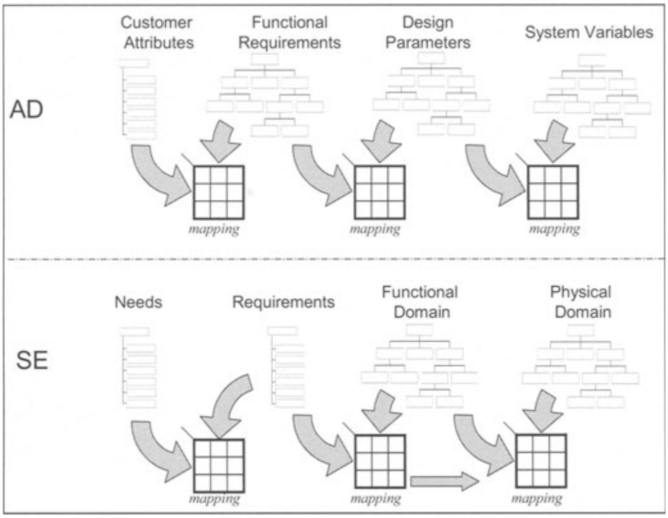


Figure -2. Very schematic comparison of ED and SE

SE can therefore partially take advantage of AD: Two major benefits can be respectively found exploring the introduction in SE of AD uses of mapping and zigzagging. The first benefit comes from the greater use of AD's design matrix analysis and the second by formalizing and improving zigzagging in each discipline.

On the other hand, AD's actual efforts in hierarchical decomposition [6] can be supported by SE concepts, including Technical Processes, as for instance the Implementation Process.

3. INCLUDING MANUFACTURING SYSTEMS ENGINEERING IN SE, A WAY TO IMPROVE PROJECTS EFFICIENCY

3.1 Describing methodological concepts

As shown in Figure 1, processes for engineering a system are divided in four types. We have identified these four types by their ability to be applied distinctly or commonly to the Products and/or the Production Systems:

- Technical and Agreement processes provide a generic description and can be distinctly applied to Product and Production Systems.
- Enterprise and Project processes bring coordination in the development of Products and Production Systems needed to produce them.

It also seems necessary to characterize in the whole technical tasks those which belong or contribute to Products, Productions Systems or both. The four Knowledge Areas of the PMBOK Guide [7] and the description of the Project Management Context are directly concerned with establishing a link between Product Systems and Production Systems.

Table -2. Management Links between Products and Production Systems development

Knowledge Area	Link between Products and Production Systems
Project Coordination (in Project Management Context)	Multi-projects and Development Team Coordination through Enterprise Organization
Project Integration Management	Project Plan Development
Project Time Management	Activity Sequencing, Schedule Development and Control
Project Communication Management	Information Distribution
Project Risk Management	Risk Identification and Risk Quantification

Of course, links are more complicated: for instance Risk Quantification, where risk concerns simultaneously Products and Production Systems, can't be restricted to an additive or multiplicative link.

The coordination engineering data between developers is also technically very complex and must be described in SE with the best practices known in Concurrent Engineering, to ensure wide Information Distribution, in levels of Communication, Co-Ordination and Co-Operation [8].

A common framework, provided by SE, where processes concerning Manufacturing Systems Development as well as Product Development are described, gives us a new approach to describe data exchange between engineering tasks. To give an illustration of our approach, we present in this paper an elementary example.

3.2 Manufacturing System Engineering, theoretical example of production of staple remover (abstract)

This example shows SE principles and only concerns the first level of decomposition, in spite of SE's interest to drive design of all levels.

Our staple remover design gave the followings results and choices :

- product physical decomposition, and up-graded staple remover physical decomposition (see Figure 3).
- choices of followings materials and processes (based on feasibility and cost analysis studies, called by E & O and V & V technical support

processes of the product design): (make) cutting and stamping stainless steel, extruding plastic, riveting plastic grip and metallic tooth, riveting parts into final product; (buy) springs, rivets (two sizes), plastic bars, steel rolls.

– choices of followings principles : same plastic grips for each L and U-parts and same riveting.

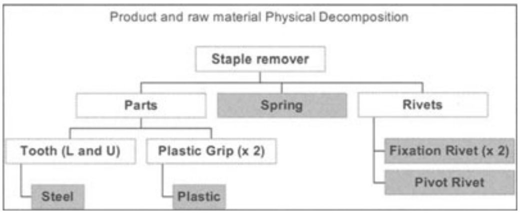


Figure -3. Up-graded Staple Remover Product Decomposition

This decomposition constitutes a source of requirements, complemented by others stakeholder's needs and including product design choices. Allocations of requirements to functions can consequently be done, and some constraints can directly affect components. Products which are physically present in the manufacturing system are taken with requirements and results of functional analysis to construct the logical architecture of the production system, introduced in Figure 4.

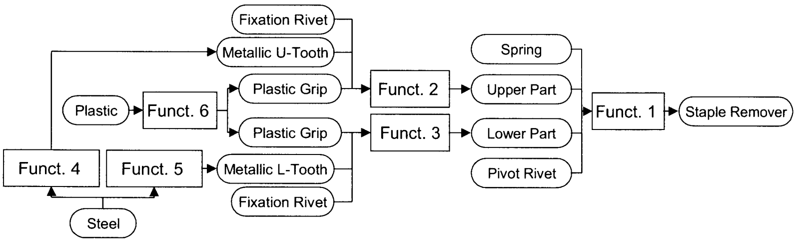


Figure -4. Logical Architecture, with macro-functions

Functions are initially determined by requirements, and logical decomposition is driven with possibility of allocation of a function to a feasible organic solution. The matrix of Functions and components of the manufacturing system is given in Figure 5.

	Components Conveyors	Positioning Machine	Pivot Riveting Machine	Rivet Conveyors	Flexible Fixation Riveting Machine	Metal Conveyors	Flexible Stamp Cell	Plastic Conveyors	Extrusion Machine	Functions	Sectors
Rivet Teeth			x							Funct. 1	Assembly
Supply Pivot Rivets	x										
Place Teeth and Spring		x									
Supply U-Teeth	x										
Supply L-Teeth	x										
Supply Springs	x										
Fix Grip to Metal Part					x					Funct. 2 & 3	Sub-Assembly
Supply Fixation Rivets				x							
Stamp Steel (2 types)							x			Funct. 4 & 5	Metal
Cut Steel							x				
Unroll Steel							x				
Supply Steel Rolls						x					
Extrude Plastic									x	Funct. 6	Plastic
Supply Plastic								x			

Figure -5. Allocation Matrix of Functions and Components

Considering the first level of decomposition will lead us to the flow chart of Figure 6.

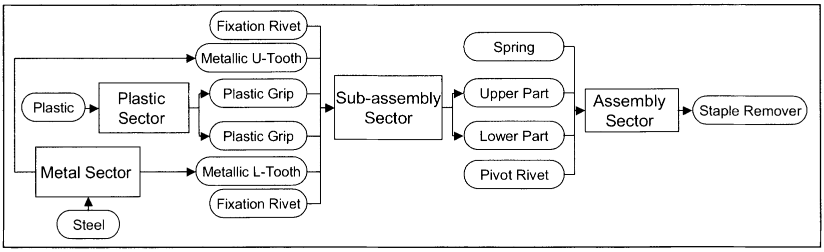


Figure -6. Physical Architecture, with sub-systems (sectors)

E & O and V& V processes can be performed at each level of design to ensure efficiency. Technical processes of SE can also be done (they've started with all valid information of upper level) to sub-systems, including emergent properties of design and maintaining the traceability of engineering data.

4. CONCLUSION ON SE'S PROFITS AND PERSPECTIVES

SE gives a structured framework to manage and justify the existence of engineering data and artifacts.

It allows developers:

– to define precisely what methods must be used or what activity must be carried out and why it is necessary.
– to assess qualitatively and quantitatively all the realizations as outputs of each process.
– to dissociate logical and physical domains through an allocation mechanism.
– to control the introduction or the suppression of systems and sub-systems in a reutilization context, considering respectively needs, requirements, functional and physical equivalence.

As a maximalist view of development, we have to talk about a level-defined application. As in many methodologies, a difference must be pointed out between what is defined, deployed or applied. A way to describe this issue is the triptych described in Figure 7. A way to solve this issue is to introduce an evaluation loop in this triptych, as shown in the same figure.

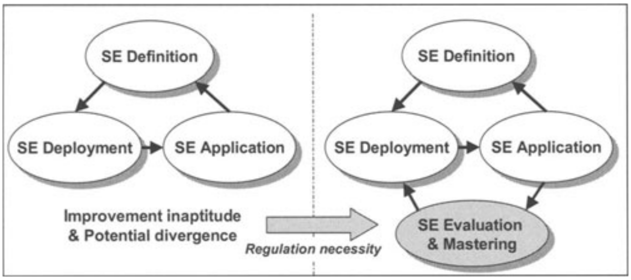


Figure -7. Method Deployment Issue

This question of evaluation to regulate SE realization in the company has already been treated in [9]. We have now to focus with further work on giving the awaited profits while identifying what can be provided by application of SE.

This paper opens up new areas for further studies to come:

– Taking the methodological history of the company into account, based on [4] and [5] conclusions.
– Giving enlightenment to the coordination of data between product and production systems development, based on [8].
– Implementing an AD contribution, based on [6].

This first work and following studies may open a debate about SE and existing Design Theories. It may open the door of SE to the Mechanical Engineering Community, and initiate improvements of both sides.

REFERENCES

1. Loureiro G., Leaney P. G., and Hodgson M. "A Systems Engineering Framework for Integrated Automotive Development" in Ninth Annual International Symposium of the International Council on Systems Engineering, 1999, Brighton, England, 6-11 June 1999.
2. Preston White K. Jr., "Systems Design Engineering" in The Journal of The International Council on Systems Engineering, 1998, Vol. 1-4, p. 285-302.
3. ISO15288, "Life Cycle Management & System Life Cycle Processes", ISO 15288 CD2, 2000, ISO/IEC.
4. Cavallucci D. and Lutz P. "Intuitive Design Method (IDM), a new approach on design methods integration" in First International Conference on Axiomatic Design, 2000, Cambridge, MA USA, ICAD2000, June 21-23, 2000.
5. Martin C., "Process design modeling through methodological integration", thesis from the Laboratoire Productique - Logistique, 2001, Ecole Centrale PARIS, PARIS. 148 pages.
6. Cochran D. S., et al., "The Application of Axiomatic Design and Lean Management Principles in the Scope of Production System Segmentation" in The International Journal of Production Research, 2000, Vol. 38 (6), p. 1377-1396.
7. PMBOK, "A Guide to the Project Management Body of Knowledge", PMBOK Guide, 1996, Project Management Institute, Newtown Square, Pennsylvania USA.
8. Tichkiewitch S. and Brissaud D. "Co-Ordination Between Product and Process Definitions in a Concurrent Engineering Environment" in The 50th General Assembly of CIRP, 2000, Sydney, Australia, received January 7, 2000.
9. Lardeur E., Bocquet J.-C., and Auzet C. "Etude sur l'évaluation de la performance globale des activités de conception des systèmes complexes et de leurs systèmes de production" in Deuxième Conférence Annuelle de l'Association Française d'Ingénierie Système, 2001, Toulouse, FR, AFIS, June 26-28, 2001.

PERFORMANCES IN ENGINEERING CHANGES MANAGEMENT

A. Riviere*, C. DaCunha** and M. Tollenaere**
**EADS Centre Commun de Recherche, Centre de Toulouse, France*
*** Laboratoire GILCO, ENSGI / INPG, Grenoble, France*

Abstract: Engineering Changes (EC) implementation in product definition and at all stages of the product life cycle cannot be avoided when struggling for innovation and maintaining the product in accordance with the contractually agreed specifications. These "micro" design actions usually provide large profits on both product and processes and allow a flexibility that Concurrent Engineering requires. If EC management looks a rather simple activity for organizations offering small products on the market, it gets more complex in industries such as aeronautics or automotive industry where the whole supply chain can be impacted. In such industries, this key process remains difficult to control as few performance indicators exist. This lack of measurement does not enable any corrective action to be undertaken. In this paper, an overview of EC management for complex products is presented from the experience we have gained in the aeronautics and automotive industries. First, potential causes and consequences are presented and a generic process for EC management is proposed. Then, the importance of measuring the performances of such a process is highlighted in order to identify any room for improvement. Afterwards, our approach to design and implement a measurement plan for EC processes is detailed. A significant set of indicators and measurement is proposed and discussed according to the specific process we have focused on.

Key words: Configuration Management, Engineering Change Management, Process Performances Measurement

1. INTRODUCTION

Over the past twenty years, the industries dealing with complex products have deeply changed. The market place is asking for more and more customized items. So, they have to deal with many more variants within the

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 369–378.

same family. This lead to the introduction of Concurrent Engineering methods and techniques enhancing collaborative work practices (Prasad, 1996).

Therefore, many more engineering changes are introduced, particularly during the development phases. It has been observed in the automotive and aeronautic industries that such changes could have an impact on a company's competitive advantage. As a matter of fact, they influence product performances, often delay activities and impact development costs. This is sometimes due to a lack efficiency and responsiveness of the Engineering Change management process. Even if this process is now clearly described in quality procedures, the use of indicators and measurements revealing its performances are not implemented. Therefore, serious optimization of this activity is possible.

In this paper, we, first present the potential causes and consequences of Engineering Changes from the survey we undertook in the automotive and aeronautic industries. Then, a generic process for the treatment of such changes will be proposed. Based on it, a set of indicators will be established as well as a method to implement them.

2. WHAT IS AN ENGINEERING CHANGE ?

An EC can be considered as "an alteration in the approved configuration of a product related item" (US Military Standard 480B, 1988). An item can be a document or a physical component (virtual or real) of the product structure. The approved configurations of these items move along the product life cycle depending on formal or informal configurations reviews.

Wright (Wright, 1997) defines an EC as "[… a modification to a component of a product, after that product has entered production]". Like most of the authors on this subject, Wright tackles the problem of EC from a production and inventory point of view. But generally speaking, the work undertaken in this field is dedicated to production-inventory communities or design communities. In this paper, we consider engineering changes from a business perspective so we do not address a particular phase of the product life cycle.

2.1 Causes and consequences of engineering changes

Firstly, let's consider the primary causes of an EC. A primary cause can be defined as the fact which triggers an EC. We have to bare in mind that causes of engineering changes are often combined as most Engineering Change Requests (ECR) are not treated individually. However, it is still possible to classify their causes. Some work has been done on the categorization of engineering changes (Hsu, 1996) but only for military aircraft programs. We try to complete it with the causes of changes we identified in the automotive and aeronautic industries.

a) Changes in needs and requirements : customer needs evolves with the identification of new product attributes and operational modes depending on the way they want to use it, new offers introduced to the market by competitors, suppliers they want to privilege, new technologies introduction (Wenzel et al., 1997), etc. Requirements can also be reconsidered due to problems or misunderstandings in capturing them right first time.
b) Program or project interactions : within an organization, programs and projects can be tightly linked. An EC studied and implemented in a program can lead to an upgrade of an earlier program or can become the standard for a future program.
c) Need to fix deficiencies : during development and production phases, not expecting changes due to mistakes sounds rather unrealistic even nowadays with TQM and collaborative work methods.
d) Technological changes
e) Legislation changes : authorities are important actors in the transportation industry where security considerations are significant.
f) Changes in project scheduling : due to customers wishes or internal delays, project schedules can be re-examined in order to meet previously agreed specifications

Unlike causes, it seems rather difficult to predict the possible consequences of an EC. But, most of the time, engineering changes are considered as problems rather than opportunities for improvement (Ring et al., 1998). As matter of fact, with a near-term view, engineering changes cost money and delay product life cycle tasks. But changes can also improve product quality, save money in a the long term and preserve product position against competitors. Moreover, all possible consequences should be taken into consideration when initiating an EC. Even if it is difficult to quantify the potential consequences of an EC, it is possible to qualify them :

a) Near Term cost Impacts : whatever the change is, it will have an impact on costs. Forecasting the cost of an EC will be among the factors to take in consideration to decide whether or not a change must be studied further and implemented.
b) Impacts on program schedule : EC can introduce delays in some activities. Impacted activities must be re-scheduled with respect to their related milestones.
c) Impacts on product performances with respect to expectations : Naturally, changes can increase product performances. Some changes

affecting the manufacturing or support systems can indirectly affect product performances. It highlights the process deficiencies to cope with product related changes. The third category encompass changes affecting product documentation. They do not have any incidence on product performances.

d) Primary impacts on suppliers, sub-contractors and work partners.
e) Impacts on other programs or projects : it can happen that an EC relative to a particular program can impact other programs. This likelihood must be investigated prior to any change study.
f) Additional changes resulting from the same issue : evolutions applied to a specific function, element or sub-system of a product can be propagated to other items through their physical interfaces or links between product structures (Clarkson et al., 2001).
g) Life cycle phases impacted : the gravity of the consequences of an EC depends of the life cycle phase where it occurs. The "Rules of Ten" states a change detected in the development phase in ten times less expensive than if detected later.

Like causes, it is fairly rare for a change to produce only one consequence. In an Extended Enterprise and for the development of complex products, the first challenge is to clearly point out the causes motivating an EC and to evaluate precisely the potential consequences of that change.

2.2 Different kinds of Engineering Changes

It is important to characterise an EC in order to decide whether or not it should be "opened" and studied further. It also helps to adapt the treatment process to the nature of the change. First, we could imagine that this categorisation is based on the classification of causes and consequences. But it is indirectly used. In fact, the concept of interchangeability is used to define the consequences of an EC on a product. This characterisation is further refined by predicting the impacts of that change on the activities performed and the amount of work to be done.

Interchangeability is defined by Watts (Watts, 2000) as the ability of an entity to be used in place of another to fulfil the same requirements, without modification,. We distinguish engineering changes having an impact on the interchangeability of the components from those which do not.

An EC which affects one or more of the following product characteristics will be considered as a ***modification*** : failure of the functional or dimensional interchangeability between the new part and the part, impacted by the EC, or an EC affecting an item material, safety, behavior and reliability.

If an EC does not match the modifications characteristics, it will then be considered as an ***amendment***. An amendment is an EC which has no repercussions on the use of the amended product by users.

When an EC only affects the presentation of information and has no repercussions on the component definition itself, it will be consider has a ***correction***.

As the evaluation of items interchangeability is tightly coupled with EC impacts studies, it seems rather difficult to state if a change is a modification rather than an amendment when it is initiated. Nevertheless, it is possible because the actor initiating a product evolution usually has an idea of a potential solution. Actors knowledge and experience in change management and product design are important for the evaluation of possible repercussions of a product evolution.

So, to characterize an EC, its potential effects on the product are first examined using the concept of interchangeability This categorization is further refined by predicting the possible effects on activities or product life cycle phases. A critical modification will affect product functions, therefore the specification process has to reconsidered. A major modification will have repercussions on the product definition process, the manufacturing process and the support processes (maintenance, documentation, etc.). Minor modification will only affect the manufacturing process. This characterization of engineering changes helps in defining the systems and life cycle phases impacted as well as the people to involve in the process.

3. A GENERIC PROCESS FOR EC MANAGEMENT

From the surveys we have conducted, it appears that, most of the time, an EC process can be broken down into three major stages. Each of them can be split in activities (Cf. Figure-1). The generic process we propose must be adapted a to the company's business and position in the supply chain as well as the complexity of the product they design or manufacture.

In practice, some of the stages described are difficult to identify as most of the activities are organized in a concurrent way and sometimes performed in an informal manner. However, at the first stage, requests for change (or Engineering Change Requests- ECR) are collected with the necessary information to process the treatment further.

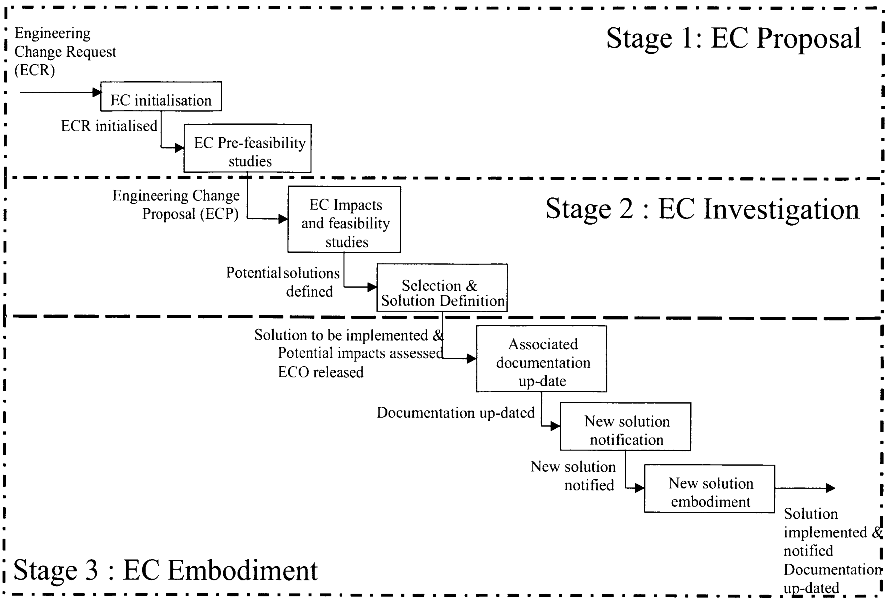


Figure -1. A generic Engineering Change Process

After initialization, a pre-feasibility study is performed in order to present the requested change and its potential impacts to an Engineering Change Committee. This group, composed of the different departments representatives, assess the risks and costs of the requested change and decide if it is necessary to "open" it so a solution can be studied. If it is opened, they schedule the tasks to be performed, notify the actors who will be involve in the process and affect a change to the solution. If not, the request is cancelled. An accepted request becomes an Engineering Change Proposal (ECP). At the second stage, all trades affected by the EC think of a set of potential solutions taking their consequences on other systems and activities into account. Then, from this set of solutions the best one is chosen. This selection is usually done by the EC Committee is based on the costs, schedules, product performances, activities and systems impacted and the date of implementation. The chosen solution is then completely defined in order to be implemented. At the third stage, the solution is physically implemented, the documentation (drawings, calculation files, maintenance guides, etc.) is updated and people concerned by this change are informed. This simple process becomes more difficult to manage in complex organizations dealing with complex products. But, from the experiences we have collected in such businesses, we have identified potential room for improvement.

4. IMPROVING THE EC PROCESS

In complex systems design, engineering changes are the rule and it is unwarranted to think they could be avoided to improve the organization's efficiency. This efficiency can only be gained if the methods, techniques and process implemented in the organizations aim at serving competitive strategies for engineering changes management. Fricke (Fricke et al., 2000) define five possible strategies :

1. **Prevention :** Avoid changes by a better in depth analysis before design and by the implementation of quality dedicated methods and techniques during the development phases.
2. **Front-Loading :** Trigger changes as early as possible by speeding up their detection and feed back loops.
3. **Effectiveness :** Assess whether or not changes are necessary and beneficial.
4. **Efficiency :** Strive to get an optimal use of resources such as time and money when treating a change.
5. **Learning :** Increase efficiency and effectiveness of the process by continuous learning from previously performed changes.

The achievement of these different strategies is not only linked to the improvement of the EC process. Continuous improvements made in design methods and in the management of the design process can also lead to better performances of the EC management. As a matter of fact, the prevention strategy depends on the methods implemented to predict and detect potential engineering changes. So, the requirements capture process, quality design methods, efficient collaborative work, responsive complaints and feed back treatments are among the activities subject to continuous improvement. The learning strategy also depends of the means implemented to capture, store, organize, share and reuse change related knowledge of previously performed changes. As most of indicators related to these two strategies are not only dependent of the EC process, we do not take them into account. Therefore, we propose a set of indicators, suitable for the EC process and dedicated to front-loading, efficiency and effectiveness measurement.

5. PROPOSITION OF A SET OF INDICATORS FOR THE ENGINEERING CHANGE PROCESS.

In our context, indicators aim at measuring the performances or variables of a process in order to control their discrepancies with some targeted objectives and to plot their evolutions in time. It is difficult to define some targeted objectives as engineering changes may be different from one to another. However, this process is not continuous and ECR can be treated individually. So, it is possible to check if the process is under control by making comparisons between the actual results and those recorded from previously performed changes. A distinction has to be made between measurements which can be established while the process is running and those that can be established once the change is implemented and "closed". However, after analysis, they are utilized to set corrective actions and/or to communication purposes. For the three stages of the EC process we described (Cf. Figure-2), we provide, in the tables below (Cf.Figures-2, 3, and 4) a set of indicators for the strategies we have highlighted.

	EFFECTIVENESS	EFFICIENCY	FRONT-LOADING
Stage 1	1. ESTIMATED EFFECTIVENESS Estimated product performances efficiency / Estimated product lead time Estimated product performances efficiency / Actual product performances Estimated cost efficiency / Processing costs Estimated product performances efficiency / Estimated number of activities impacted	1. Throughput time 2. Lead time / Scheduled lead time 3. Number of ECR related to the same problem 4. Costs of waiting time until the decision to proceed is taken 5. Urgency 6. Number of elements impacted by ECR 7. Planned effectivity of a solution	1. Product life cycle phase where the change is detected

Figure -2. Indicators dedicated to stage 1

During the first stage, it is only possible to estimate the effectiveness of processing an EC. The decision to raise a ECR is based on this estimation. Indicators related to efficiency can be grouped in two categories, those which measure the performances of the process and those which help the right decisions to manage the process. Monitoring the product life cycle phase where a change is triggered helps in defining if some changes should have been processed earlier on.

At the second stage, efficiency measurements are based on lead times, proposed solution, resources involved, and related to the "best solution" selection process. At this stage and further, the indicators related to the "front-loading" strategy aim at assessing if the detection of the change and its implementation occur within the same product life cycle phase. The effectiveness is calculated using information associated to the proposed solutions. It is a refinement of the estimated effectiveness indicators.

	EFFECTIVENESS	EFFICIENCY	FRONT-LOADING
Stage 3		1. Throughput time 2. $\frac{\text{Lead time}}{\text{Scheduled lead time}}$ 3. Number of documents impacted 4. Number of times a document has been impacted 5. Number of persons to inform of this new solution 6. Time spent on physically implementing the solution	1. Product life cycle phase where the EC is closed

Figure -3. Indicators dedicated to stage 3

At the third stage, it is no longer useful to introduce indicators associated to effectiveness as the solution is implemented and an in-depth analysis can be undertaken over the whole process. Efficiency indicators are based on the documentation impacted by the change and on the persons to be informed of the introduction of a new solution.

	EFFECTIVENESS	EFFICIENCY	FRONT-LOADING
Whole Process Analysis	1. REAL EFFECTIVENESS $\frac{\text{New product performances efficiency}}{\text{Process lead time}}$ $\frac{\text{New product performances efficiency}}{\text{Old product performances}}$ $\frac{\text{New cost efficiency}}{\text{Total processing costs}}$ $\frac{\text{Real product performances efficiency}}{\text{Real number of activities impacted}}$	1. Number of ECR in a period 2. Number of ECR proposed/ rejected 3. Real cost of the solution/ calculated cost 4. Calculated cost/ estimated cost of the solution 5. Real solution's effectivity/ planned effectivity 6. Planned effectivity/ estimated effectivity 7. Responsiveness of the notification process	1. Detection and implementation in the same product life cycle phase 2. Could the change have been detected earlier? When? 3. Could the solution have been implented earlier? When? 4. Cost of late detection and implementation

Figure -4. Indicators dedicated to whole process analysis

Most of the indicators suggested for the analysis are built from those used at other stages. They reveal the overall performances of the EC process. The results of indicators dedicated to effectiveness measurement show if it was beneficial or not to process a given ECR. That kind of results should be reused to compare with new and similar change request. The efficiency of the process is measured trough the calculation of the difference between real results and estimations. This reveals if the methods used to assess performances, costs and development lead times of a new solution should be improved or not. The "front-loading" indicators show if the ECR was raised at a propitious moment and if it could have been submitted earlier in the product life cycle. Therefore, design methods and detection technique could have to be examined.

6. CONCLUSION

In this paper, we strove to gain a better understanding of the Engineering Change phenomenon and of its associated management process in complex products development. First, we identified the main causes of engineering changes as well as their primary consequences (disregarding propagation occurrences). Then a generic management process was proposed based on the observations we made in the automotive and aeronautic industries. During these surveys, we noticed that few indicators have been implemented in this process so improvements of related activities were almost impossible to undertake. So, a set of indicators has been proposed .

7. REFERENCES

(Prasad, 1996) : B.PRASAD, *"Concurrent Engineering Fundamentals : Integrated Product Development*", International series in industrial and systems engineering, Vol. 2, Prenctice-Hall, 1996.

(US Military Standard 480B, 1988) : US Military Standard 480B, *"Configuration control – Engineering changes, deviations, and waivers"*, July 1988, section 3.1.30.

(Wright, 1997) : I.C. WRIGHT, *"A review of research into engineering change management : Implications for product design"*, Design Studies, Vol.18, 1997, pp. 33-42.

(Hsu, 1996) : T.C. HSU, *"Causes and impacts of class I engineering changes : an engineering study based on three defense aircraft acquisition programs"*, MSc in Aeronautics and Astronautics, Massachusetts Institute of Technology, 1996.

(Ring et al., 1998) : J. RING, E. FRICKE, *"Rapid evolutions of all your systems – Problem or opportunity",* Proceedings of the IEEE 17^{th} Digital Avionics Systems System Conference , Seattle, 1998.

(Calrkson et al., 2001) : J. CLARCKSON, C.SIMONS, C. ECKERT, *"Change prediction for product redesign"*, Proceedings of ICED 01, Glasgow, 2001.

(Fricke et al., 2000) : E. FRICKE, B. GEBHARD, H. NEGELE, E. IGENBERGS, *"Coping with changes – Causes, Findings and Strategies"*, Journal on Systems Engineering, Vol.3, N°4, 2000.

(Watts, 2000) : F. WATTS, *"Engineering Documentation Control Handbook – Configuration Management in Industry"*, William Andrew, ISBN 0-8155-1446, 2000.

PROPOSAL OF A NEW DESIGN APPROACH INTEGRATING THE CONCEPT OF THE WORKING SITUATION

R. Hasan*, A. Bernard** and J. Ciccotelli***
** Centre de Recherche en Automatique de Nancy, Université de Nancy, France*
*** Institut de Recherche en Communication et Cybernétique de Nantes, France*
**** Institut National de Recherche et de Sécurité, Nancy, France*

Abstract: The design process is a crucial phase in the life cycle of an industrial system. It requires the co-operation of different specialists with the skills and knowledge vital to the creation of objects satisfying the needs of the future users. In the case of industrial systems, one of these needs is the safety of those who act on the system during its life. In this paper, we propose a new design approach aimed at correctly integrating all the factors linked to the real contexts of use of the industrial system. The design method is based on a conceptual data model that has already been presented, namely the generic model of the working situation, and therefore only the essential elements will be reviewed to allow a better understanding of the fundamentals of our work.

Key words: Industrial system, working situation, design

1. INTRODUCTION

In traditional design methods, the work of the different specialists overlaps and communication between them is not always easy. This is particularly true when safety is taken into account during the design of socio-technical systems. Some of the work carried out on the integration of safety at the design stage is presented in [1] [2]. In practice, however, safety is very often taken into consideration at the end of the development cycle and even during the installation of the system on site [3]. Our research effort is focussed on the integration of the information of the different specialists intervening during the life cycle of the system (mechanical, electrical,

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 379–390.

materials, cost, technical information, safety, etc.) through an original approach based on the concept of the « working situation ». In this respect, we have proposed a definition [4] to integrate this into a model of the system [5] and propose a design approach that allows the different phases of the life of the system to be taken into account. Indeed, the working situation presents the contexts of use of the system and the team working on this system.

This paper outlines a new design approach aimed at correctly integrating all the factors linked to the real contexts of use of a production system. This method is based on a conceptual data model, namely the generic model of the working situation proposed in [6] [7]. In particular, taking safety into account systematically during the design process allows proper respect of the requirements of the standards without creating contradictions (e.g. safety – productivity) linked to this integration [8] [9].

2. WORK METHODOLOGY

The aim of our research work is to propose an innovative approach enabling human safety parameters to be taken into consideration as early as possible in the design process. Firstly, we had to overcome two main difficulties. The first concerned the characterisation of the safety parameters. For this problem, the solution was partly contained in the results of the work of our partners in this project (cf. Acknowledgements). The second was how to integrate these parameters into the design process in a simple and natural way without complicating the work of the designer and without increasing his or her workload. This means carrying out this integration at the best moment at every stage of the design process. The methodology described below was employed to achieve these objectives [10] [11] :

1- observation of and interviews with the design teams,

2- analysis of documents in the partner enterprise: Much of our early work focussed on modelling the operation of the industrial system by including the safety point of view, and therefore the levels of human-system interactions. This work was developed from an analysis of the documents supplied by the partner enterprise (the design and integrator of this system) to its customers concerning utilisation, adjustment and maintenance. We also carried out a functional analysis of the design process applied by the design office (DO) in this enterprise [11]. The results of our field analysis [3] [4] modelled by means of the Structured Analysis Design Technique (SADT) method [12] [13] have already been presented.

3- formalisation and modelling.

On the basis of the analyses conducted, and for reasons of validation and verification, we decided to create data models using the UML modelling method [6]. *UML is a language designed to view, specify, construct and document the artefacts of systems with a high software content* [14]. The fundamentals of this modelling are described in the following paragraph.

3. MODELLING

Modelling is the process of creating a logical image of the data and/or the processing, and allows four objectives to be achieved :

- models help us to visualise a system in its working situation, either as it is or the way we would like it to be ;
- they clarify the structure and/or the behaviour of the system and of the team working on this system ;
- they provide a framework to guide the construction of a system by taking into account the contexts of use ;
- finally, they allow documentation of the decisions taken during the design process and capitalisation of the knowledge acquired with the passage of time.

The domain examined relies on knowledge relative to different technical fields and concerns the life cycle of industrial systems in which numerous human actors are involved, from design through to exploitation. Thus, the concepts of « multi-views », « multi-actors », and multi-technology products » imply the definition of data models allowing a complete representation of the designed system and the dialogue between the actors. This must also take into account the constraints inherent to each speciality and to each technology at every level of the product life cycle. In this respect, the two extreme approaches, namely the anthropocentred and technocentred approach, are incapable of judiciously and simultaneously taking into consideration the two interacting components: operator and production system.

Our work, positioned between these two approaches, develops the concept of globally taking stock of the working situation, where the production system and operator co-operate to achieve a given performance in accomplishing a well-determined mission. To do this, it is essential to propose a new design approach that has recourse to the concept mentioned earlier, i.e. the working situation. According to Fadier [1], the analysis of real working situations (on the site of the user company) reveals, from the safety point of view, divergences from the designed working situations. In addition, safety is often taken into account at the end of the development cycle, even during installation of the equipment, both to respect legislative requirements and to satisfy normative provisions, which in our case are those contained in draft European Standard Pr EN 1010 [15]. These arguments gave us an additional motive to take into consideration in the model all the parameters linked to the context of the use of the production system through coherent and innovative models.

Thus, the conceptual model, the essential elements of which are reviewed in what follows, is meant to construct a comprehensible and simplified model of a real working situation of a system on its site of operation. This model served as the basis of a computer model, the main purpose of which was to provide coherent help to designers to simultaneously take into account technical and human aspects during the design of a production system [16]. To do this, we studied the design process of the production system and its process of integration, and then created a model to represent the entire « socio-technical system » by employing the concept of the « working situation » to allow the subsequent development of a predictive risk model. Particular attention was paid to the « human-production system » relationship.

4. THE WORKING SITUATION CONCEPT

Following our analyses of the real working situation carried out at the premises of the customers of our industrial partner, we observed that the co-existence of operator and technical system was an absolute necessity. As presented earlier in paragraph 2, the working situation is based on the existence of operators and a production system carrying out a particular mission. Our problem was to try to identify all the elements influencing the Human-System interaction from the point of view of the designer. In fact, we considered from one angle the technical provisions of the production system as the system itself, i.e. the machines, the information systems, the tools and the instruments, and, from another angle, the physical conditions linked to the system or to environmental conditions such as noise, vibration, heat, cold, etc. A third angle was examined encompassing the intervention modes of the operators, the objective of the work, the nature of the tasks carried out by the operators, aspects linked to the physical activity and the psychological characteristics (knowledge, experience, qualification) of the operator to respect the instructions imposed by legal standards. The interactions between the working team and the production system are presented in figure 1. In contrast, we did not take into account aspects linked either to the socio-economic situation of the enterprise designing the system or, in the user enterprise, to the temporal and spatial organisation of the work or the type of work (isolated, team, production line, autonomy, etc.). The model of the situation is presented in [7] [17]. Improvements and enhancements were made to the model and are presented in [6] along with possible scenarios for using this model. Figure 2 presents a global view of the model of the working situation in UML format; the attributes of the different classes have not been detailed.

4.1 The working situation in terms of time scale

The model can be instantiated by a set of working situations. Each instance of the situation presents a situation at a given time t. These working situations characterise a system during its operational lifecycle and its context of use. In fact, the working situation can be represented as a series of images over time. This concept represents a set of tasks, activities and intervention modes in a period of time. This period of time is determined by a change to one of the elements of the situation. For example, we are no longer in the same instance of the working situation when the working team is changed, when the work system is changed or is no longer in the same mode of operation or when the working team changes intervention mode.

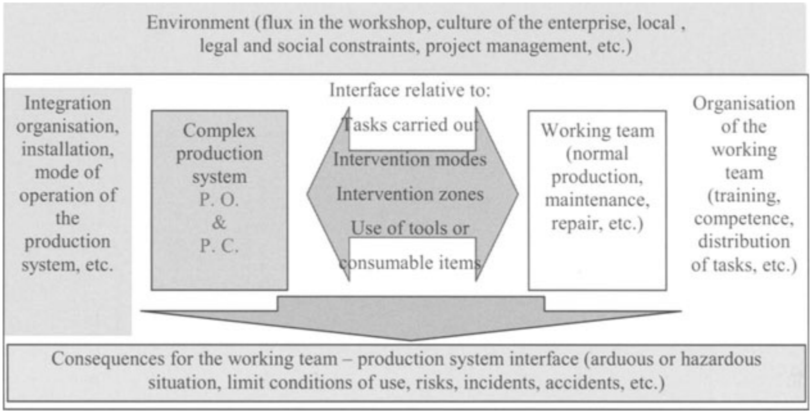


Figure –1. The working team – production system interactions

In contrast, changing other elements does not change the instance of the situation itself. It does however change the value or the image of this situation. Put another way, at time t1 all the elements of the working situation have values (operator X carrying out task Y). This set of values presents an image of the working situation. At time t2 elements of the situation take other values (operator X has finished task Y and is carrying out task Z), which presents another image of the same working situation. In fact, our model presents the working situation as a sequence of events (figure 3).

More generally speaking, this representation can be extended to the entire life cycle of the product and takes into consideration all the specialists linked to the design and exploitation of the production system.

In terms of design, employing the concept of the working situation consists in imagining the contexts of use of the system and therefore the design of working situations (definition of the system and its operating modes, definition of the tasks and intervention modes, definition of the context of this intervention (number of people, tools and consumable items, qualifications, etc.)). Within the context of this work, the aspects linked to the modelling of real working situations were focussed more on the system exploitation phases and less on installation, start-up and dismantling.

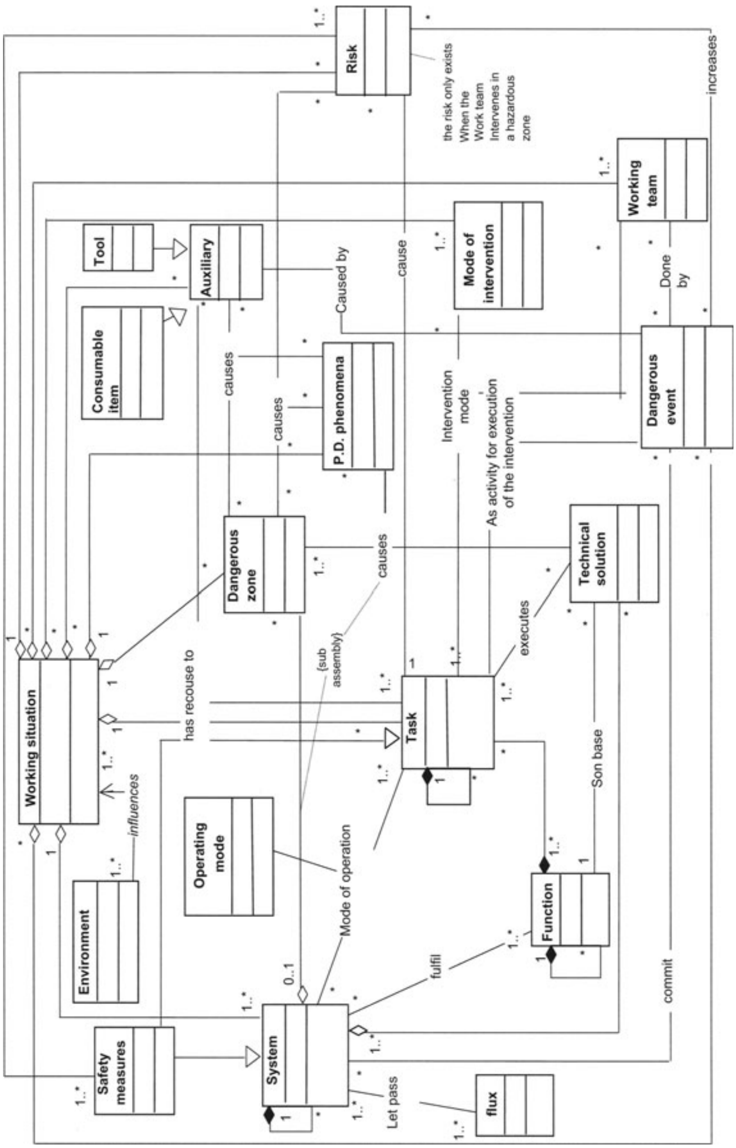


Figure –2. A global view of the generic model of the working situation

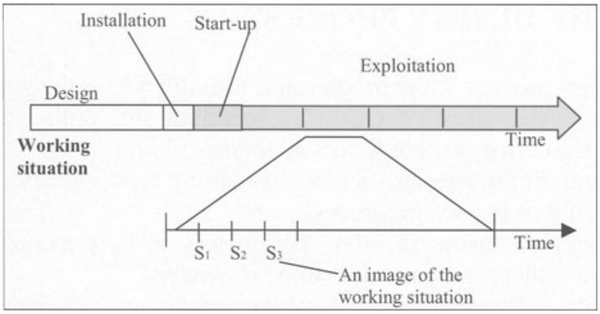


Figure –3. Temporal aspects of the Working situation concept

4.2 The dynamic of the « working situation »

The dynamic of the utilisation of the system in the field is taken into account in the « Working situation » and « Task » concepts by their attributes and their sequencing and composition operators. Indeed, the sequencing operator determines the follow-up of composite tasks and their order of execution. The duration of a task is identified by the duration attribute. Tasks can be broken down into composite tasks by the composition operator to determine the hierarchy of this breakdown. Hence, a task is only considered executed if all the elementary tasks have been carried out. Tasks, in the working situation, are accomplished by the system and/or the working team. The system ensures the technical tasks are completed and the working team carries out the manual tasks and the surveillance and verification tasks. The activities necessary for the execution of a task are presented within the scope of the Task concept since:

- a utilisation task can be broken down into several elementary tasks that can only be done in one way. In other words, only one activity is necessary to carry out the global utilisation task,
- during the design phase, the designers are not in a position to assess the activities of the working team. They can merely imagine the possible activities that may be undertaken to accomplish the task.

Thus, we propose specifying and breaking down the task to a sufficiently detailed level at which the operator cannot be excluded from the operation required.

Generally speaking, during the design process, designers lay down the tasks of the system and those of the working team. In reality however, during the utilisation of the system, operators can, for various reasons and on the basis of their own reasoning, accomplish the task in a different way. Our approach, which allows capitalisation of experience feedback through the representation of real working situations, therefore has an additional advantage.

5. THE DESIGN PROCESS

The design process is the progression through which the system acquires the properties necessary to satisfy the objectives set. Within the context of this project, we define our objectives as focussing on :

- integrating all the specialists linked to taking working situation concepts into account in the design process,
- integrating and assessing safety parameters as early as possible without considering them as constraints to be respected,
- integrating experience feedback systematically and efficiently,
- facilitating communication between designers of different professions and, if possible, making this permanent throughout the design process. This allows for a circulation or flow of information that is enriched as the development cycle of the system progresses,
- avoiding costly modifications due to late interventions,
- facilitating the task of the project leader by follow-up related to completion of the designers' tasks,
- allowing the earliest possible intervention of one specialist when a decision is being made by another designer about a problem concerning the speciality of the former,
- supplying the necessary information for taking a decision linked to the development of the system and/or the management of its design process,

The aim is to keep all those involved in the project informed, in particular the project leader, of the state of technical advancement of the tasks entrusted to the members of the project with a view to deciding as quickly as possible the approach to follow and the following steps.

This paper offers a new design approach, the objective of which is to fully integrate all the factors linked to the real contexts of use of the production system. This approach is based on the use of the generic model of the working situation proposed. Thus, we propose a new way of working in a team in permanent communication through the use of an information technology environment and a data structure based on the model outlined earlier. This makes available the means to integrate safety-related parameters into the design process at the right moment. In no way does it prejudge the methods used to calculate a parameter, estimate a risk or determine a danger zone. For example, to transform the functions into functional specifications we do not impose the manner in which these specifications can be obtained. The same applies to the transfer of functions to the technical solutions ensuring these functions. System parameter dimensioning can also be carried out in several ways. The potential existence of risks or hazardous phenomena linked to the technical solution is made known to the designer, but to the best of our knowledge, at the present time, the estimation and assessment of risks would remain his or her responsibility. Finally, the proposed approach does not concern the definition of the manufacturing process plans of the systems or of their components.

Having broadly described the structure and the objectives of conceptual data modelling, we now go on to look at the dynamic aspects linked to the construction and utilisation of the model. In our approach, we take into account three levels of the dynamic point of view of this model :

- the dynamic of the working situation itself on the site of utilisation. This concerns the potential of the model to present the dynamic of the real working situation by the « Task » concept. This dynamic is the subject of modelling allowing the clear presentation of the sequencing and the composition of the tasks carried out in the field. In the longer term, this model should allow the analysis of the divergence between the tasks laid down and those actually carried out (coming from experience feedback) in order to indicate to the designer factors linked to this divergence,
- the project-level dynamic, this level concerns the person carrying out his or her design task to fulfil a requested function. It presents the dynamic of the use of the model and the potential passages between the different classes as well as the « reaction » of the model once a choice has been made (propagation and control of coherence). For example, if a technical solution is chosen, the model points out the potentially hazardous phenomena linked to this solution and the necessity to determine the risks caused by such a choice,
- the design process dynamic. This dynamic concerns the progression and the advancement of and the communication between the project members. It allows a representation of the collaboration between the project members, and the project leader can thus follow the progress of the project members and intervene efficiently as early as possible.

These different levels of model dynamic apply to all the specialists involved in the life cycle of the system.

From an experimental point of view, our contribution focuses on the proposal of a tool, method and approach that, while promoting the integration of safety into the design process, responds to the other traditional criteria of improving the productivity and the reliability of the production system. This requires taking into account various factors linked both to the system and to the working team and the relationship between them. We have structured the model so as to :

- allow the definition of the system in its working situation at the outset of design by extracting from the contractual specification all the information relative to the specification of the system. In fact, the specification documents normally contain specifications unique to the system and do not take into account the working situation in which the system must operate during its life time. This leads to the designer determining the specifications linked to the environment, the working team and all the elements characterising the working situations,
- enrich this initial description as the design process progresses to end up with a complete specification of the working situations of the system,
- retain the history of the design and the life cycle of each system designed to allow future use or simply consultation. In our case, the design history

is represented in the form of the state of advancement of the achievement of objectives (functions) and of the justified taking of decisions throughout the development cycle of the system. To do this, we adopted the basic concepts of the SAGEP method (Design process management decision-making system) proposed by Ouazzani [18],

- allow the systematic capitalisation of experience feedback to facilitate the integration of the information collected at the site of the user into subsequent versions of the system,
- identify the contradictions resulting from the integration of safety with a view to eliminating them (cf. paragraph 7).

Following the analyses of the design process carried out at the premises of our industrial partner, and to achieve the previously mentioned objectives, we are currently working on modelling the dynamic of the design process by taking into account the generic model of the working situation.

6. RESOLUTION OF CONTRADICTIONS

Taking safety into consideration at both design and operating levels highlights conflict management taking in technical, economic and human aspects: for example, the choice of an efficient but costly technical solution in terms of design, immediate benefit or potential risk in terms of exploitation. These conflicts are generally resolved by compromises. The TRIZ method [19] proposes several resolution principles to eliminate these conflicts. On the basis of these principles, an analysis of the safety standards and our field studies, we propose a structuring of potential solutions to assist the designer in his or her task and to take safety into account as early as possible. Modification of the geometry for better operator accessibility (normal operation, maintenance, adjustment, etc.) responds to the objective of the separation of characteristics [8]. We have already presented our analyses concerning a safety problem on a complex automated production system. Our research is now focussed on Boundary Conditions Tolerated by Use (BCTU) [20]. In [9], we studied the interest of structuring the data provided by our model as an aid to using TRIZ in taking safety into account during design. On the one hand, we defined at which level and at what time TRIZ must respond to the needs of the designer, and on the other, we sought to define the role that TRIZ can play when making technical decisions.

7. APPLICATION

The model presented earlier has been entered into a database and has undergone development in terms of methods and of human-machine interfaces specific to each profession. The application has been coupled to a project follow-up tool. This information-technology tool development has recourse to the generic model of the working situation and to the industrial requirements of the enterprise (respect of current information technology systems and processes) [16]. Indeed, the database can be linked to another functional base (e.g. technical data management, documentary database, etc.). Thus, for example, when selecting a technical solution, the designer can ascertain and identify the risks likely to be engendered by the solutions retained and, should they exist, potential alternatives. At the time of writing, the tool developed is simply a model and not an industrial product directly usable by the enterprise. Its aim is primarily demonstrative. It must firstly allow full-scale testing of the proposed method by the enterprise. Secondly, this model offers a support for the application and demonstration of our work.

8. CONCLUSIONS AND OUTLOOK

As this paper has tried to show, the aim of our research is to take into account the « socio-technical system behaviour » point of view to prevent the risks linked to its use. This has involved representing all the elements in a model adapted to such a system, and has also included specifying three levels of models relative to the production system and its design process (design – integration – utilisation) to allow the longer-term construction of a predictive risk model.

In the near future, we would also like to contribute to the construction of a methodological and technical support system to provide assistance to designers to assess the appropriateness, in terms of risks, of the choices made to satisfy the functions requested. This area constitutes a strategic point in the field of industrial system design [21] [22].

ACKNOWLEDGEMENTS

This work was carried out within the framework of a project undertaken by the « Integration of Risk Prevention at the Design Stage Group » of the French National Research and Safety Institute (INRS), and was jointly financed by CNRS as part of the Production Systems Programme (PROSPER).

REFERENCES

[1] E. FADIER, « *L'intégration des facteurs humains à la conception* », Phoebus, la revue de la sûreté de fonctionnement – Special number, p. 59 – 78, 1998.

[2] J. CICCOTELLI, « *Sécurité intégrée à la conception des systèmes de production* », PRIMECA Université d'Automne, Nancy, 20 - 22 octobre, 1999.

[3] R. HASAN, J. CICCOTELLI, A. BERNARD, P. MARTIN, « *Representation and evaluation of risks during the design phase of a complex system* », Proc. of ESREL 2000, Foresight and Precaution, Cottam, Harvey, Pape & Tate (eds), Rotterdam, Netherlands, volume 1, p. 141-147, ISBN 90-5809-1414, 2000.

[4] R. HASAN, A. BERNARD, J. CICCOTELLI, P. MARTIN, « *Intégration de la sécurité dès la conception de systèmes de Production : modélisation de l'interaction homme-machine dans le fonctionnement du produit lors du processus de conception* », Proc. of the 3rd International Conference on Integrated Design and Manufacturing in Mechanical Engineering, Mascle, Fortin & Pegna (eds), Presses Int. Polytechnique, Montréal, Canada, *2° trim. 2000*, ISBN 2-553-00803-1.

[5] A. BERNARD, « *Modèles et approches pour la conception et la production intégrées* », Revue APII - JESA, Vol. 34 - 2-3/2000, pp 163-194, N°ISBN : 2-7462-01445.

[6] R. HASAN, A. BERNARD, M. LOMBARD, « *Proposition d'un modèle de système pour l'intégration de la notion de situation de travail dans le processus de conception* », The third French-speaking conference of MOdelling and SIMulation (MOSIM'01), Troyes, 25-27 avril 2001, ISBN 1-56555-212-1.

[7] R. HASAN, A. BERNARD, J. CICCOTELLI, P. MARTIN, « *Integrating safety into the design process: elements and concepts relative to the working situation* », Safety Sciences, Special number, "Safety by design", accepted, http://www.sciencedirect.com/science, available online 20 June 2002. 2002.

[8] R. HASAN, P. MARTIN, J. CICCOTELLI, A. BERNARD, « *Contribution à la résolution de problème de conflit dans la prise en compte de la sécurité dans le processus de conception : analyse d'une ligne de production automatisée* », 4th International congress on Guality and reliability (Qualita 2001), 22-23 mars 2001.

[9] R. HASAN, P. MARTIN, A. BERNARD, « *L'intégration de l'approche TRIZ dans le processus d'utilisation du modèle générique de la situation de travail de : cas d'une ligne d'imprimerie* », 4th International Industrial Engineering Conference, Aix-Marseille – France, 12-15 juin, 2001.

[10] R. HASAN, A. BERNARD, J. CICCOTELLI, P. MARTIN, «*Etude bibliographique* », First report interns within the framework of the project PROSPER-GIPC, INRS, décembre 1998.

[11] R. HASAN, A. BERNARD, J. CICCOTELLI, P. MARTIN, « *Prise en compte de la sécurité dans le processus de conception de systèmes de production* », 2nd report interns within the framework of the project PROSPER-GIPC, INRS, novembre 1999.

[12] I.G.L. Technology, « *SADT un langage pour communiquer* », Editions EYROLLES, 1989.

[13] P. JAULENT, « *Génie logiciel : les méthodes SADT, SA, E-A, SA-RT, SYS-P-O, OOD, HOOD,...* », Editions ARMAND COLIN, 1994.

[14] G. BOOCH, J. RUMBAUGH, I. JACOBSON, « The Unified Modeling Language User Guide », The Addison-Wesley Object Technology Series, Addiso – Wesley, ISBN 0-201-57168-4, 1998.

[15] Pr EN 1010, European Standard Definitive Project. Safety requirements for the design and construction of printing and paper converting machines - Part 1 : common requirements. European Committee for Standardization ,Brussels, July 1997.

[16] F. CHERRIER, « *Prise en compte de la sécurité en phases de conception par l'intégration conjointe de l'homme et de la machine dans une Situation de travail* », studies end report, Ecole Centrale Nantes, septembre 2001.

[17] R. HASAN, A. BERNARD, J. CICCOTELLI, P. MARTIN, « *Prise en compte de la sécurité dans le processus de conception de systèmes de production* », 3rd report interns within the framework of the project PROSPER-GIPC, INRS, novembre 2000.

[18] A. OUAZZANI-TOUHAMI, « *Présentation dynamique du processus de conception : une perspectives de capitalisation des historiques de conception, la méthode SAGEP* », Ph.D. Report, Ecole Centrale des Arts et Manufactures (Ecole Centrale de Paris), Laboratoire Productique logistique, 1999.

[19] D. CAVALLUCCI. « *Contribution à la conception de nouveaux systèmes mécaniques par intégration méthodologique »,* Ph.D. Report, Université Louis Pasteur, Strasbourg, decembre 1999.

[20] A. Didelot, « *Contribution à l'identification et au contrôle des risques dans le processus deconception* », Ph.D Report, Institut National Polytechnique de Lorraine, octobre, 2001.

[21] E. FADIER, J. CICCOTELLI, « *Safety integration in industrial system design: general overview* », 9th IFAC Symposium on Information Control in Manufacturing, Nancy, FRANCE, p 233-239, 1998.

[22] A. BERNARD, E. FADIER, « *Contribution à la prise en compte des aspects liés à l'utilisation des équipements de travail, pour la prévention des risques associés, dès leur conception* », CPI'2001, Fès, 24-27- octobre 2001

TOWARDS IDENTIFICATION AND FORMALISATION OF KNOWLEDGE IN COMPUTATIONAL MECHANICS

Y. Baizet, F. Pourroy and R. Allera
Laboratoire 3S, Grenoble, France

Abstract: Scaling up computational mechanical behaviour from small scale to the industrial level within a company gives rises to new problems : lack of back tracing, lost knowledge, repeated mistakes. The work shown in this article aims at identifying, structuring and formalizing the information and knowledge on simulation, to make them capitalisable and reusable by other participants and on other projects. The Renault Company is an active partner in this research through one of its computational mechanics departments and helped us to test and confront our ideas with real industrial simulation processes. The outcome is presented in two main steps. In the first one, we show how the need for structure in simulation processes leads to specification of new links between the different kinds of capitalised information. In the second one, we precisely describe the concept of Instructional Case, whose target is to manage knowledge of a generic nature with a high potential for reuse.

Key words: Capitalization, Knowledge, Information, Structuring, Computational mechanics, Simulation

1. INTRODUCTION

Mastering numerical simulation is essential for the competitiveness of industrial companies. On the one hand, it contributes towards the quality of marketed products, and on the other hand it helps to largely decrease the product development time by using fewer but better physical prototypes. It is in all cases a central element of the digital mock-up, the problematic of

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 391–400.

which does not simply amount to the sharing and integrating heterogeneous 3D geometric data.

However, the increase in numerical simulations in product development processes brings problems of back tracing and consistency of the simulations made during the project. The current tendency of companies to shift the easiest calculations from engineers to technicians after they have been trained to simulation adds to the problem. In other respects, co-operation between the "World of Simulation ", the "World of Design " and the "World of Testing" is often very poor. Capitalisation and reuse of calculation models or simulation processes are difficult to achieve within a project or within the different projects of a company, and the same is true for identifying the knowledge generated by the simulations. The tools created to assist engineering tasks are unfortunately of marginal use. Developing and selling cars and lorries, Renault has widely increased the use of computational mechanics in some departments over the last few years, and is now confronted with these general problems which raise the difficult questions of back tracing, capitalisation and reuse of the knowledge within the company.

Research work on this subject is numerous and topical. **1** tackles the question of company memory and draws up a comparative evaluation amoung the different methods of knowledge capitalisation which are known (REX, MKSM, CYGMA, MEREX, etc.). Unlike to these methods which are based on models of formalisation of knowledge, **2** proposes a model which aims at formalising more the context of use than the knowledge itself. The object of the system is then to connect the proper people at the proper time. Closer to the profession of computational mechanics, the work presented in **3** or **4**, for example, presents systems of help for modelling which are interesting but which really help only in very precise problems and thus are difficult to generalize to the whole knowledge of simulation.

The approach presented in this paper is based on the work made at Laboratoire 3S these last few years, the general objects of which - back tracing, capitalisation and reuse - have given a capitalisation model specific for simulation of mechanics called SG3C **5**. The application of these concepts to the structure and objectives of Renault led to a revision of the initial model.

2. SPECIFICATIONS OF A CAPITALIZATION TOOL

Capitalising on knowledge is a big job and the chances of achieving it are strongly determined by the precise definition of the objectives to reach and the means to engage in it **6**.

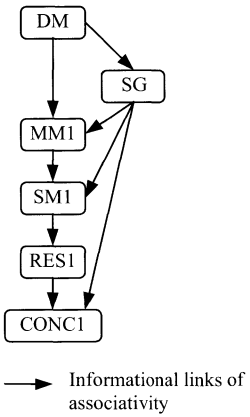
The domain on which our work is focused is computational mechanics, the participants of which are either experts (engineers in calculation) or non experts (technicians). These participants also have different status, for example permanent personnel of the company or external subcontractors. Lastly, they have different functions as some of them have a management role in teams of different sizes. This brings many differences which have much influence on the system of capitalisation to be built up.

The company where our observations are made has an organisation which splits the domain of computational mechanics in two entities. One is the operational calculation, which is made within the design department and is used to develop or validate a part or an assembly of parts in a project. The other one is the advanced calculation which is used to develop simulation methods to be supplied to operational calculation. This separation allows to distinguish on the one hand the knowledge linked to the product or project, which is generated by the operational calculation and, on the other hand, that related to the profession, generated by advanced calculation and used later in operational calculation. Nevertheless, the participants of the operational calculation can also generate professionally related knowledge, as they can modify methods of calculation when in use (changing the domain of validity, finding out better hypotheses, etc.).

The main target of the capitalisation project refers only to mechanics domain related knowledge, which means all the knowledge generated by the participants of the advanced calculation plus the mechanics domain relating part of the one generated by the participants of the operational calculation. Actually, mechanics domain related knowledge and project related knowledge are very different and the methods to be developed to capitalise them are also different. It would be too long and complex to create a structure to capitalise both types of knowledge at the same time, and we limit the work to the mechanics domain relating one : only methodologies, methods and remarks from use of methods will be capitalised, to the exclusion of the product related knowledge which is set aside for the moment.

3. GIVING SG3C METHODOLOGY EXPERIENCE OF AN INDUSTRIAL SITUATION

The SG3C methodology **5** has been tested within the advanced calculation environment of Renault and it appeared that this knowledge capitalisation system proposed concepts of industrial interest which however required a great deal of adaptation to the real demands of the company.



<u>Design Model (DM) :</u> geometrical and technological model of the part or product to study.

<u>Simulation Goal (SG) :</u> specifies the analysis to be done (conditions of use, sollicitations, wanted outputs).

<u>Méchanical Model (MM) :</u> represents the whole of hypotheses and the idealised geometrical representation of the mechanical behaviour to be studies (independently from the method or calculation tool to be used).

<u>Simulation Model (SM) :</u> associates the method and calculation tool to give a computer ready representation nf the part or product.

<u>Résults (RES)</u> : set of outputs necessary to make the analysis (graphic outputs, tables, values, …).

<u>Conclusion (CONC) :</u> made by the analyst on the results and according to the initial goal of simulation (SG).

Figure -1. Influence type links between the entities of a cycle

Therefore, work has been undertaken to further develop data structuring in order to model all of the calculations more faithfully. Thus, developments have been made to redefine links between entities and make the Instructional Case concept richer and more accurate.

3.1 Definition of the links

To describe the calculation processes and move through the capitalisation system, **5** defines a cycle of simulation as an elementary process to evaluate mechanical behaviour, starting from the definition of a single model, up to the analysis of a set of results related to this model. **5** also identifies six elementary entities in a simulation cycle (see figure 1) and defines links between them. These links express a informational associativity between two entities of the cycle, which means that they convey an information requirement from one entity to another which is assumed to contain the required information. It is therefore possible to model the influence of the data contained in an entity on the building of another entity and for that reason they are called **Influence** type links. The active role of this type of link is to assist in the implementation of data in a cycle of numerical simulation by defining a logical process to build the simulation steps.

Actually, a calculation does not mean only one cycle, but is often made of successive cycles which express the sequence and the convergence of a calculation process. Not all cycles are independent, each of them being a developmental step down from the previous one, to obtain a model which satisfies the initial objectives. The modelling of this dependence relation between the cycles of a whole project is reached with an **Evolution** type link which links on cycle to next, as shown in Figure -2.

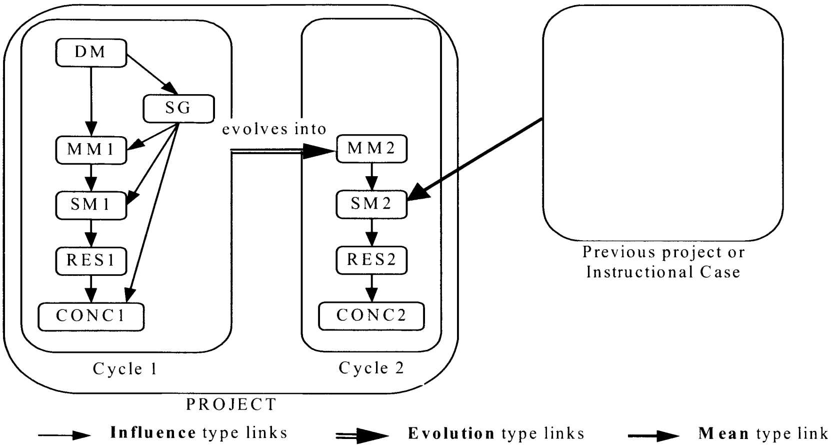


Figure -2. Links of dependence

In this figure, it is visible that the first mechanical model is inappropriate to the problem, which required an evolution towards a second model, more appropriate to carry out the objectives. Because this evolution is located at the level of the mechanical model, only the modified entities are formalised, the other ones being kept implicitly as detailed in a previous cycle.

For certain cases, building an entity in a cycle requires using a knowledge from a different project, this knowledge having been capitalised alone or through a Instructional Case (see §3.2). It is then necessary to represent the mean which permitted to create this entity, which means giving prominence to the specific knowledge or experience used. But speaking of knowledge does not limit one to the technical information contained in one or two entities. It has to take into account the context and specific environment of this knowledge, in order to have a global vision of it in order to manipulate and exploit it in the best way. For that reason, one piece of knowledge refers to at least one complete calculation cycle, but can also refer to several cycles. This representation is naturally made with a specific link, the **Mean** type link, which links the entity which uses the knowledge to the whole set of the exploited knowledge, as shown in Figure -2.

3.2 The concept of Instructional Case

Troussier presents in **7** the concept of Instructional Case, the target of which is to capitalise a particular type of knowledge. Let us see what is behind this notion and what is the desired use.

3.2.1 Definition of the concept of Instructional Case

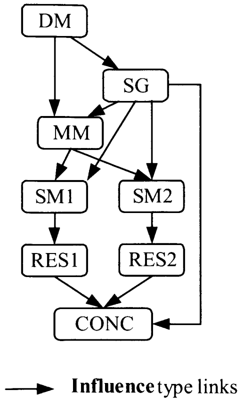
Defining knowledge is a difficult thing. **8** proposes to regard knowledge as a personal reality, as a relation between an individual and an object of knowledge. Data, information and learning are then considered as models independent of persons. From these objects of knowledge nature, the personal relation between the individual and these objects gives rise to several types of knowledge : generic knowledge, local knowledge and knowledge in the action.

The Instructional Case becomes identified with the notion of local knowledge, as it gives prominence to a field in which a contextualised knowledge having a local significance has been observed. The definition of this concept of Instructional Case comes from the need - and wish - to formalise the information which refers to calculation of a generic nature, which means reusable in a different context. The progress of a project often needs to create and process much information of various types and levels. Most of them have a signification in the current project only, and thus cannot be used or reused easily in another project. Meanwhile, in certain cases, a whole set of information can be taken out of the context of the project to serve again in a different context, wherever it is similar to the current context or not. This kind of information, with a generic nature, gains from being identified independently of any project and from being capitalised in a specific knowledge base, which is different from the base where all the calculation projects are gathered. This gives any participant in the calculation easy access to generic information which can be reused in any other project.

Thus, the Instructional Case has a deeper meaning in the context of reuse as the ensemble of information that it contains is decontextualised from any project - particularly from the initial project which generated it - and then can be readapted to any specific project.

This concept stands at an intermediary level between the notion of rule, which is too general and has too wide an application domain to be adapted to our needs, and the notion of example, where its specific information is usable only in context and cannot be transferred to other projects.

Because the main target of the system of capitalisation under definition is to formalise and save the calculation knowledge to reuse it, it is obvious that the concept of Instructional Case is well adapted to these objectives and can be transposed to this system. However, the implementation of the Instructional Case and its confrontation with industrial calculations has shown a lack of structure in the method and the need for a more detailed typology of the Instructional Case.



__**The Instructional Case "Simulation"**__ :
How to represent a given physical problem, analytically or numerically and with the available methods and tool ?

- Each Simulation Model is independent form the others,
- Each Simulation Model uses distinct tools, methods or analysis parameters,
- The Conclusion shows the best adapted Simulation Model to fulfil the objectives defined in the Simulation Goal.

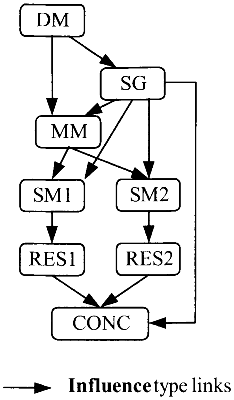
Figure -3. The Instructional Case "Simulation"

3.2.2 Different types of Instructional Case

In order to be efficient, the research work has been focused on the notion of mechanics domain related memory, in the particular case of computational mechanics. This mechanics domain related memory refers to the Instructional Case defined in 7 on which we base our work. To give the participants of the calculation a data reception structure which is specific, quickly identifiable and more easily reusable for each type of knowledge to capitalise through the concept of Instructional Case, the latter has been divided into three categories. The first two are the Instructional Case "Simulation" and the Instructional Case "Modelisation". They refer to knowledge of a technical order. The third one is the Instructional Case "Method" as it concerns concepts nearer to method or reasoning.

3.2.2.1 The Instructional Case "Simulation"

This Instructional Case proposes a structure for the capitalisation of technical information related to the research and implementation of a specific simulation model adapted to the calculation objectives. The idea is to compare several Simulation Models, belonging to a unique triplet Design *Model / Simulation Goal /Mechanical Model*, to find out which of the Simulation Models tested is the best adapted and make it appear in the Conclusion entity. Figure -3 gives a global view of the specific structuring of the Instructional Case "Simulation" with all the Influence type links which form it.



<u>The Instructional Case "Simulation"</u> :
How to represent a given physical problem, analytically or numerically and with the available methods and tool ?

- Each Simulation Model is an independent from the others,
- Each Simulation Model uses distinct tools, methods or analysis parameters,
- The Conclusion shows the best adapted Simulation Model to fulfil the objectives defined in the Simulation Goal.

Figure -4. Instructional Case "Modelisation"

3.2.2.2 The Instructional Case "Modelisation"

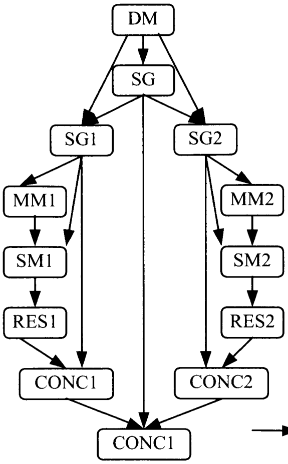
The Instructional Case "Modelisation" is similar to the Instructional Case "Simulation", but it concerns a specific Mechanical Model instead of a Simulation Model. Following the same idea, it makes a comparison between several Mechanical Models belonging to a unique couple *Design Model / Simulation Goal*, to find out which of the Mechanical Models tested is best adapted and make it appear in the Conclusion entity. Figure -4 proposes a possible representation of a Instructional Case "Modelisation".

The comparison of several Mechanical Models needs to implement several Simulation Models, which are adaptations of each Mechanical Model according to a method and a tool of calculation and which give results leading to a conclusion. It is however indispensable to implement Simulation Models which modify the mechanical hypotheses made during the mechanical model construction as little as possible, in order to obtain the best comparison possible.

3.2.2.3 The Instructional Case "Method"

In a calculation project, an initial objective has often to be broken down into a number of more elementary objectives giving elementary conclusions from which a final conclusion can be made. This particular structuring of calculation is the reason why we need the Instructional Case "Method", which proposes this division of one objective into elementary objectives. Figure -5 proposes an example of such a breakdown.

This Instructional Case needs to integrate new levels of refinement to model the embedding of elementary objectives in more general ones. Every level has to contain the couple *Objectives / Conclusion* which is indispensable for understanding of the general structure of the project and which prevents the simulation process from consistency loss.



<u>The Instructional Case "Method" :</u>
Which answers to elementary questions are able to solve a given problem ?

- The initial problem is broken down into several elementary problems, independent from one to another,
- Every elementary problem can be seen as a complex one and thus broken down,
- Every level contains at least the couple *Simulation Goal / Conclusion*.

→ **Influence** type links

Figure -5. The Instructional Case "Method"

4. CONCLUDING REMARKS

The increasing use of computational mechanics in the design processes of big industrial companies means that the latter are strongly preoccupied with back tracing, assistance and capitalisation of knowledge. The Work presented here is made within this context, in collaboration with a calculation department of the Renault company.

The SG3C method, derived from a previous work, was implemented in the company and the observations made showed the necessity to make the structure of data reception evolve, as it was seen as too general.

The work made led first to a detailed description of the concept of dependence through three different types : the influence type link which has to give a process to build simulation models, the evolution type link designed to trace the chronology of the process of simulation and last, the mean type link useful for reuse.

The work also gave a redefinition of the concept of Instructional Case used to capitalise mechanics domain related knowledge, which are methodologies, simulation methods and experience feedback on these methods. Three Instructional Cases have been created. Two of them, Instructional Case "Simulation" and Instructional Case "Modelisation", allow to capitalise technical information on different levels. A third one, Instructional Case "Method", is useful to model the process of creating elementary problems.

These evolutions actually suggest a need for structuring calculation processes at a more global level than that of the six basic entities presented in the article. The knowledge related to the domain of simulation seems to extend somewhat further than the only technical data which have been identified with the mechanics domain relating entities "Mechanical Model" and " Simulation Model ". In addition to that technical level, processes which lead to realisation of a whole set of simulations have to be modelled.

REFERENCES

1. Dieng R., Corby O., Giboin A., Golebiowska J., Matta N., Ribiere M. Méthodes et outils pour la gestion des connaissances, Editions DUNOD, avril, 2000.
2. Mahe S. PUMEO : un modèle actif pour la gestion des connaissances tacites et explicites dans l'entreprise, Ingénierie des connaissances, Plate-forme AFIA, Presses Universitaires de Grenoble, pp. 21-39, 2001.
3. Breitfield T., Kröplin B. An expert system for the verification of finite element calculations, Proceedings of the SAST conference, IEEE, 1996.
4. Rieg F., Koch F. Selection of finite elements considering loadcases and geometry, Proceedings of the International Conference on Engineering Design (ICED), Professional Engineering Publishing, Glasgow, 2001.
5. Troussier N. Contribution à l'intégration du calcul mécanique dans la conception de produits techniques : proposition méthodologique pour l'utilisation et la réutilisation, Thèse de Doctorat de Mécanique de l'Institut National Polytechnique de Grenoble, octobre, 1999.
6. Ballay J.F. Capitaliser et transmettre les savoir-faire de l'entreprise, Collection de la direction des Etudes et Recherche d'Electricité de France, 1997
7. Troussier N., Pourroy F., Tollenaere M., Trebucq B. Information structuring for use and reuse of mechanical calculations models in engineering design, Journal of Intelligent Manufacturing, vol. 10, n°1, pp. 61-72, 1999.
8. Prudhomme G., Boujut J.F., Pourroy F. Activité de conception et instrumentation de la dynamique des connaissances locales, Ingénierie des connaissances, Plate-forme AFIA, Presses Universitaires de Grenoble, pp. 41-60, 2001.
9. Tourtier P. Analyse préliminaire des métiers et de leurs interactions, Rapport intermédiaire du projet GENIE, INRIA - Dassault Aviation, 1995.

TOWARD NEW METHODS AND TOOLS TO FOSTER PRODUCT/PROCESS INNOVATION

J. Legardeur, J-F. Boujut and H. Tiger*
Laboratoire 3S, Grenoble, France
**Laboratoire CRISTO, Grenoble, France*

Abstract: This article presents the specifications and features of a new tool designed to co-ordinate the development of new solutions during the early phases of design projects. It is the result of 18 months of field work during which we played an active role in the development of an innovative solution. We were also able to closely analyze the activities of certain actors (called materials experts) who are often behind proposals for new materials and processes. The ID² (Innovation Development and Diffusion) tool is geared on the one hand to help innovative solutions emerge, and on the other to consolidate these solutions by circulating them to the company's different specialists on the other. The tool is based on a concepts/criteria table enabling the viewpoints of the different actors involved to be summarized during the design phase. In this article we shall show how functional features such as links, questions, alarms, information enquiries and the possibilities for exchanging information between actors can contribute to developing solutions for these different professionals. We will also see how the "innovative" solution proposed can be developed further by comparing and assessing it in relation to a list of criteria which is gradually drawn up by the actors involved in the project. This tool is designed to be managed by a specific actor (a coordinating actor), who may also, depending on his/her strategy, consolidate the proposal by involving an increasing number of people.

Key words: Innovation, information sharing, knowledge dynamics, network of actors, collaborative engineering

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 401–410.

1. RESEARCH METHOD AND BACKGROUND TO THIS STUDY

This article presents the results of pluri-disciplinary research into the integration of new technologies in upstream design stages carried out jointly by engineering design researchers and industrial sociologists. Our approach is based on intervention research. For over 12 months we took part in the development of a project within a large company specialising in industrial vehicles. This socio-technical study was the opportunity to closely observe the practices of actors faced with a proposal for an innovative technical solution (and backed up by an expert supplier). The technical aspect of the research project was related to the development of a new application using a composite material (SMC or sheet molding compound) which was not very well known and seldom used at that time in the design office we were observing. On the other hand, we very quickly found that the design office was in a relatively stabilised situation and the material choice turned out to be more often than not in the favour of steel. The situation showed that knowledge was not being equally shared between the design office, the material expert and the supplier, this last being an expert in the design and production of composite parts. However, this imbalance did not just concern the technology since the supplier was also in the process of learning about the product to be developed, and the set of associated constraints. We were thus able to observe and characterise the difficulties involved in integrating a material, different from the ones traditionally used, in a context where the actors did not produce a minimum of shared knowledge, particularly off SMC technology [1].

Following this study we felt it would be interesting to compare our own experience in the field with the system used by the customer when integrating new technologies. Indeed, we were able to extend our field study to include the material department of our partner by following and questioning specific actors, referred to as “materials experts” who were also in charge of putting forward new product/process alternatives to the design offices. This second study enabled us to precisely characterise the question of innovation during the pre-design stage, outlined in section 2, and also helped to draw up a proposal for a new tool, which is what we shall look at in section 3.

2. THE ROLE OF MATERIALS EXPERTS IN THE PREPARATORY PHASES OF INNOVATIVE PROJECTS

2.1 Fostering product/process innovation

It is today a well known fact that innovation is a key issue in any company's development and competitiveness. In this article we have looked at this issue from an unusual angle that of innovation in the upstream design phases of projects through product/process integration.

Traditionally, product/process integration is presented as an efficient way to plan ahead for the production stage and reinforce co-ordination between those who develop the product and those who design the process. This is why we talk about the simultaneous development of the product and manufacturing process since the process requirements have to be taken into account right from the start of the product design phases. Because of this, new services, new methods and new tools have been developed to meet these objectives. However, most of these systems are designed to be implemented in a relatively stabilised environment, i.e. in which the "product" actors and "process" actors both know the production processes.

In our field study this was not the case as the situation observed was characterised by a considerable amount of uncertainty about the product as well as the process. Furthermore, given that in the pre-development phases the technologies were not always defined, this uncertainty was accentuated by a lack of knowledge about the product and the process. And yet, it is during these phases in particular that innovative technological choices can be made from a set of several alternatives. But exploring new product/process alternatives then becomes very difficult and often off-putting as the actors find themselves devoid of knowledge in certain areas when traditional solutions are already relatively stable and advanced. However, we believe that the development and product/process integration of new materials during the upstream phases is a real vector of innovation just as innovation can emerge from new uses or new needs. We underline that looking ahead to new technologies is a way of finding new product/process concepts which can open the way to innovative product alternatives. We are therefore looking to develop an additional channel for innovation: that of product/process innovation.

2.2 Conditions for product/process innovation

An analysis of the innovative development cases that we were able to follow shows that product/process innovation can only move forward if several dimensions have been taken into account. Indeed, we observed that in situations where a new material is being introduced, the pros and cons of the proposed technology are not enough in themselves to get a new project accepted. The process of adopting or rejecting a new material is part of a more complex process that goes beyond simply confronting different technologies. We have thus chosen three main dimensions to characterise product/process innovation and the introduction of new technologies.

2.2.1 Organisational dimension: structural changes through actor networking

In product/process innovation situations, the goal of design work above all consists in managing a certain amount of tension between a "qualification" (or acceptance) system linked to the new materials set up by the promoters of a new solution, and a "de-qualification" (or rejection) system implemented by the promoters of a more routine-based solution. This opposition, which can also be found in [2] concerning innovation and organisation, to a certain extent destabilises the systems traditionally set up. In particular, we noticed that the design group is often called into question and finds itself faced with the need to reconfigure the networks of actors. Indeed, we underline that the dynamics of design within the framework of product/process innovation require a new network of skills to be set up and developed for the actors to deal with a totally new set of problems.

2.2.2 Dimension relating to the creation and sharing of knowledge

There can be no product/process innovation that does not involve the creation and sharing of knowledge. In our case, introducing a new technology leads most actors on the side of the design office to discover the material and the process. This double novelty is characterised by the customer design office's lack of knowledge about the technical and economic aspects of the process and explains the tensions which can get in the way of and put a stop to these innovative projects. In this context, using a new technology in design involves developing new knowledge, which has to be updated and discussed throughout the network and this involves setting up real co-development with the expert supplier. In that case, the goal is therefore to organise the emergence and confrontation of know-how and practices on the product and process.

2.2.3 Instrumental dimension through an analysis of the assessment systems

When new know-how is emerging and new actors are being introduced the product assessment and qualification systems are inevitably called into question. Indeed, for more routine-like projects, the "map" of actors involved and the main criteria used are defined and stabilised fairly quickly. However, in the case of innovative projects, we saw that this map is very unstable with, notably, the progressive arrival of new actors, bringing new assessment criteria with them at different stages in the project. We were able to observe the limits of the usual assessment systems with respect to the technical aspects (choice of solutions, calculations, simulations, etc.) and the economic aspects (estimates, quotes). These systems must therefore be modified and completed in order to back up the assessment grids to which these new criteria are added taking into account that such criteria are inherently unstable in the case of any innovative project.

2.2.4 Pre-developments outside of projects: the concept of preparatory phases

Following our field study concerning innovative development, we were able to extend our research in the field by accompanying actors playing very specific and strategic roles in product/process innovation, i.e. the materials experts. Indeed, these people sometimes have to work during pre-project periods, often with the help of several other actors, in order to explore, imagine and assess alternatives to the potential applications of a new idea or a new concept. Ideas for associating technology and application are thus developed during periods of negotiation and research, which are often informal and non contractual and which are referred to as *the preparatory phases*. At this level the official project has not been launched and the goal of these phases is first of all to be able to bring together a certain amount of data and information in order to justify and consolidate the idea put forward creating a configuration in which it is possible to launch a project. During these periods, the materials experts research the technical, economic, strategic elements and so on, relating to their new idea in order to define a "problematic area". This "problematic area" will be progressively analysed by the materials expert in relation to the capacities and characteristics of the new technology. This will be gradually tested with respect to the requirements and characteristics of the existing product and processes.

The preparatory data help to provide the first data that will go some way to rationalising the decision to launch a new project by providing the elements for comparing the alternatives available. This means that even if these phases are not clearly identified today in the development process traditionally presented by companies, they nevertheless constitute in themselves a real design work which is necessary in the development of innovative solutions. Furthermore, according to our hypothesis the efficiency of this *preparatory phase* work could be increased if the right tooling was used. Our work therefore consists in providing the instrumental tools for these *preparatory phases*.

3. A NEW TOOL FOR VIEWPOINTS SYNTHESIS AND INFORMATION SHARING: ID² (INNOVATION DEVELOPMENT & DIFFUSION)

3.1 Target objectives

Since our analyses concern the development of innovative projects and the practices of materials experts they show up a certain lack of tools offering help during these preparatory phases. This is why we have developed a tool designed to be used in a given structure with a view to improving the following four points [3]:

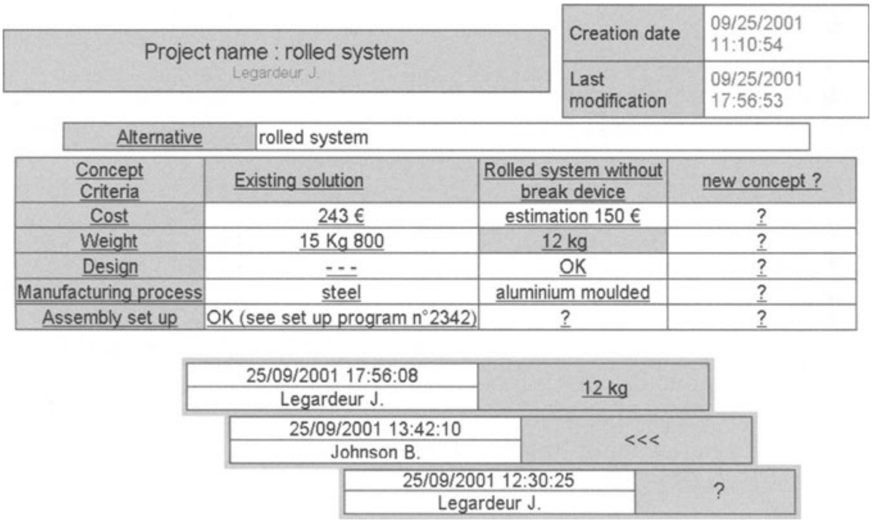
– Setting up a network of actors,
– Distributing information,
– Setting up a project log book,
– Learning and capitalising on experience during the project.

One of the main difficulties is to take into account the different strategies used by actors during the preparatory phases and generally in the upstream phases of a project. The tool we are offering is a tool mainly designed to be used mainly by the actor that is promoting the innovative solution (the co-ordinating actor) as we have already seen within the framework of the material expert's activity. However, the widely diffuse nature of the design activity [4] and the need to mobilise a wide variety of skills during the prospective phases has led us to propose a tool that is open to other participants operating in the network. Our main objective is to offer a tool to provide the co-ordinating actor with assistance in his/her strategy by taking into account the different points of view, rationale and reasoning of the other actors involved. Using such a tool the co-ordinator can define project access rights for the different actors in the company (and, if necessary, for the actors in external companies). Each participant must give his/her identity before being able to consult and act on the projects in the database for which s/he has been given authorised access by the co-ordinator.

Thus, following the identification procedure all information entered is linked to the author's name. However, any new information must be validated by the co-ordinating actor before being entered in the data base. The co-ordinator must check each declared participant's access by defining the read accesses to the project, for all information or just a certain number of specific points. The co-ordinator also checks the modifications put forward by the actors by validating or putting on hold information entered by the other project participants.

3.2 The Concepts/Criteria Table (CCT)

The ID² tool is mainly based on a Concepts/Criteria Table (CCT) which gives the alternative solutions in increasing order (presented in each column) together with the assessment criteria (presented in each line). Figure 1 below gives an example of this.



Project name : rolled system Legardeur J.		
	Creation date	09/25/2001 11:10:54
	Last modification	09/25/2001 17:56:53

Alternative	rolled system

Concept / Criteria	Existing solution	Rolled system without break device	new concept ?
Cost	243 €	estimation 150 €	?
Weight	15 Kg 800	12 kg	?
Design	- - -	OK	?
Manufacturing process	steel	aluminium moulded	?
Assembly set up	OK (see set up program n°2342)	?	?

25/09/2001 17:56:08 Legardeur J.	12 kg
25/09/2001 13:42:10 Johnson B.	<<<
25/09/2001 12:30:25 Legardeur J.	?

Figure -1. CCT and tree structure for one item of information

We propose to summarise the information from the different areas of expertise (engineering and industrial design, marketing, manufacture, etc.) in the concepts/criteria table so that all the participants can see the different elements called into play by the actors. In this table, the first column is for the reference solution, which is often the existing solution. The other columns are for the different alternatives (called concepts) which are proposed throughout the development phase.

So that the different points of view and areas of expertise can be compared, the idea is to provide a shared support tool enabling each actor to specify and explain his/her assessment criteria for the solutions. Thus, the structure of the table is dynamic and each actor can put forward new criteria (and thus new lines) as the project progresses. The criteria and information are therefore open to public viewing and discussion throughout the network.

The table is structured so that the different boxes can be gradually filled in. The most recent information is displayed in each box of the table while previous information given by the different participants is stored in a tree structure. Furthermore, the previous states of the table can be accessed using a case history and recalled for checking, explaining or justifying some of the project decisions.

Furthermore, we propose a structure with different levels of use so that actors can formulate information simply (a value, an appraisal, a word, etc.) until this information has a more complete specification (an exhaustive report). So, each box in the table can be linked to additional elements (texts, images, files enclosed) providing a broader description of and wider documentation on the solutions. The result of this structure is a summary tool giving access to different write and read levels as we shall see in what follows.

3.3 Annotations

We have seen that the design situations where partners were not sharing enough common knowledge, and more especially during the preparatory phases, often lead to communication and translation difficulties between the different fields of expertise and participants involved. The concepts/criteria table provides a multi-view support where each actor can react, comment on and request explanations with regard to any aspect of the project. One of the key features of this tool is the possibility of storing and sorting the results of different exchanges between project participants which helps to present the information in a global and structured way [5].

The tool thus provides an instrumental support to guide the interactions between different specialists via 4 annotation modes: links, alarms, questions and information enquiries.

1. The links (shown by segments on the CCT in figure 2) allow actors to link two dependent pieces of information, whose dependency is not obvious and is worth being underlined.
2. The warnings (shown by the $\boxed{W}$ icons on the CCT in figure 2) indicate an actor's remark about a specific point.

3. The questions (shown by the $\boxed{\text{Q}}$ icons in the CCT in figure 2) show the need to look for information about a specific point and thus also help to mark the degree of uncertainty surrounding the solution.
4. The information enquiries (shown by the $\boxed{\text{dI}}$ icons in the CCT in figure 2) allow actors to express their interest in a specific point of the project.

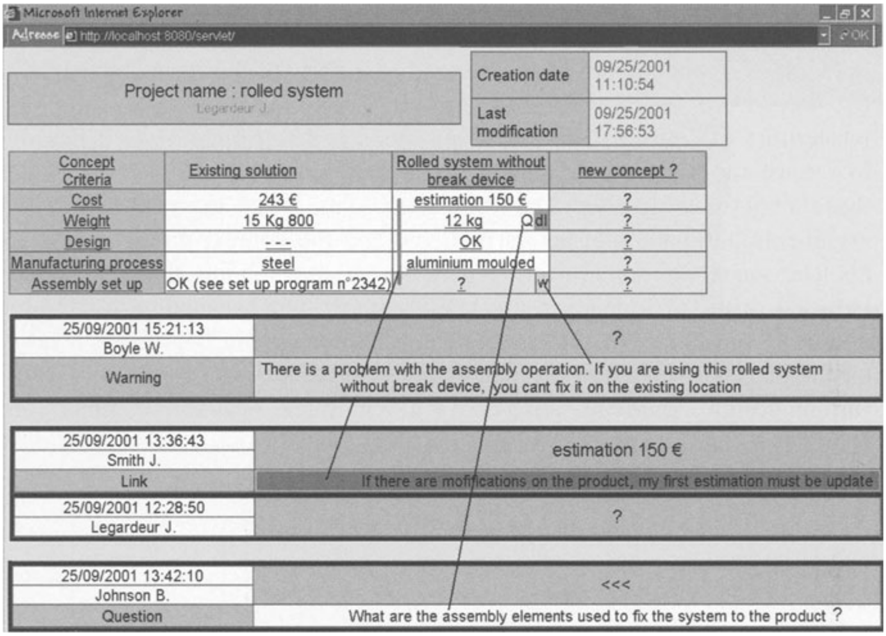


Figure -2. Examples of links, alarms, questions and information enquiries

4. CONCLUSIONS

The tool is structured to encourage actors to formulate and explain their own criteria thus facilitating discussion within the network. Each new actor adds his or her vision of the solution, which may be positive, neutral or negative, resulting in a certain number of assessment criteria. The innovation is the result of an adaptation process, the success of which depends on the actors who will be progressively involved in the design procedure [6]. Thus, the ID² tool does not aim to offer help in looking for new ideas but encourages innovation by distributing information about the solution and putting it to the test.

The tool provides all the actors with the different views of the project enabling them to learn about the specialised field of the other actors.

Collective learning processes [7] are very important in design and especially when a new material is being introduced. They help to create shared knowledge thus encouraging co-operation mechanisms for future projects. With ID^2, the co-ordinating actor is no longer the only one to benefit from the new knowledge built into the action. For example, questions raised by actors on certain specific points (e.g. can the proposed material be recycled?) can lead to discussions (technical, economic, strategic, etc.) and possibly bring answers. This new information can then be shared and re-used by the network of actors and possibly extended to the rest of the company thanks to the key word research features of ID^2. The different tree structures concerning the information, links, questions and warnings make it possible to retrace the history of exchanges between actors and discussions which these may have led to. The tool goes some way to memorising key arguments, the names of actors involved and the context giving rise to the choices made on the project. We believe that showing these exchanges between different specialists creates a support which can improve the co-operation and learning of actors in design situations. Our hypothesis is based on the fact that when a new technology is introduced, the capacity to re-use former design experience and, above all, mobilise available company and supplier skills are often strategic elements that encourage the process of innovation.

REFERENCES

1. J. Legardeur, J.F. Boujut, H. Tiger, « Innovating in a routine design process - Empirical study of an industrial situation », IDMME 2000, Montreal, May 16-19, 2000.
2. N. Alter, « Innovation et organisation : deux légitimités en concurrence », Revue française de sociologie, avril/juin, pp. 175-196, 1993.
3. J. Legardeur, J.F. Boujut, H. Tiger, « An Interface Tool for driving innovation during preparatory phases: Application in the design of composite parts », ICED 01, Glasgow, August 21-23, 2001.
4. O. Garro, D. Brissaud, S. Tichkiewitch, « Toward a design cooperative system », Actes COOP'96, 12-14 juin, pp. 375-390, 1996.
5. S. Finger, S. Konda, E. Subrahmanian, « Concurrent design happens at the interfaces », Artificial Intelligence for Engineering Design, Analysis and Manufacturing 95, 9, pp. 89-99, 1995.
6. M. Akrich, M. Callon, B. Latour, "A quoi tient le succès des innovations: l'art de l'intéressement, l'art de choisir les bons porte-parole", Annales des Mines, série Gérer et Comprendre, n°11, n°12, 1988
7. A. Hatchuel, « Apprentissages collectifs et activités de conception », Revue Française de Gestion, juin/juillet/août, pp. 109-119, 1994.

RELIABLE DESIGN OF THIN AXISYMMETRIC SHELLS USING PARTIAL FACTOR CALIBRATION

N. Gayton, M. Pendola and M. Lemaire
Laboratoire de Recherches et Applications en Mécanique Avancée, IFMA et UBP, Clermont-Ferrand, France
Φ -MECA Engineering, Clermont-Ferrand, France

Abstract: The mechanical behavior of thin axisymmetric shells subjected to buckling largely depends on variability (loading, material properties, geometrical imperfections,...). The design of this kind of structure is traditionally based on deterministic analytical calculations and on the use of the well-known empirical safety factors listed in the British Standard BS5500 [1]. This empirical standard could be improved by a probabilistic design rule based on the use of the λ_a knock-down factor, which only takes into account geometrical imperfections, and on the use of *n* partial safety factors noted γ_i which consider all the structural variability except geometrical imperfections. The Arbocz method used to calculate the λ_a-factor has been efficiently applied [2] to an industrial axisymmetric shell requiring a tool which links two types of software: INCA-STANLAX for the numerical finite element mechanical modeling and RYFES for the calculation of the reliability level. This present paper focuses on the γ_i-factors. The calibration of partial factors is treated, with the aim of proposing values to bring to the design rule for thin shell structures subjected to buckling.

Key words: axisymmetric shells, reliable design, buckling, calibration, partial factors.

1. INTRODUCTION

Structural design and more especially dimensioning is widely based on mechanical computations from characteristic values $x_{i_k},\ldots,x_{n_k}$ of design

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 411–420.

variables and from the so-called “safety factors” listed in the design rules. These design rules are the results of the designers know-how. Their deterministic use hides the existence of the structural probability of failure due to several structural imperfections and uncertainties.

The use of probabilistic approaches is in full expansion in theoretical ideas [3] as well as in mechanical applications [4]. They allow us whether to justify existing factors used in structural design or to improve design rules by computing new partial factors for a target structural reliability level: this second procedure is the so-called “partial factor calibration”. The knowledge of probabilistic modeling still being incomplete, a simply algorithmic procedure cannot be satisfactory and a calibration procedure is performed to obtain the same implicit reliability level of the existing rules.

Thin axisymmetric shells submitted to external pressure are subjected to geometrical instability, buckling is very sensitive to several types of variability (geometry, loading,…) [5]. In this field, serious failure consequences of these kinds of structures like submarines, make the use of probabilistic approaches very interesting and justified. The design of such structures is traditionally performed using British Standard BS5500 [1]. This standard could be improved by a probabilistic rule whose format is:

$$G = G(x_{1_d},\dots,x_{n_d},\lambda_a) = \lambda_a P_{adm}(x_{1_d},\dots,x_{n_d}) - P \geq 0$$

where P is the external loading pressure, P_{adm} is the safe buckling load which is computed from the perfect structure without geometrical imperfections and from n design values $x_{1_d},\dots,x_{n_d}$. This kind of rule format has the advantage of separating the different types of structural variability. The λ_a-factor only considers geometrical imperfections like defects in roundness. This factor can be computed using the probabilistic approach from the Arbocz method [2]. The x_{i_d}-values are design values different from characteristic values and permit the structural target reliability level to be satisfied. The designer only knows characteristic values and design values may be expressed as a function of characteristic values using partial factors noted $\gamma_1,\dots,\gamma_n$. As a convention, $\gamma_i = x_{i_d} / x_{i_k}$ if X_i is a loading type variable (an increment of x_{i_d} decreases the G-function) and $\gamma_i = x_{i_k} / x_{i_d}$ if X_i is a resistance type variable (an increment of x_{i_d} increases the G-function). The design values and consequently the partial factors computations are the goal of the calibration procedure.

The goal of this paper is to compute partial factors to bring to the design rule of thin axisymmetric stiffened shells. This paper considers neither geometrical imperfections (that is to say the λ_a-factor) nor other factor taking eventual mechanical modeling uncertainties. It is divided into two parts. The first one reminds us reliability concepts and more particularly calibration method principles. The second one is the implementation of calibration methods on a particular complex stiffened shell requiring a numerical mechanical modeling. This part compares the reliability index of structures designed using different calibrated rules and BS5500.

2. PARTIAL FACTOR CALIBRATION METHODS

2.1 Partial factor computation

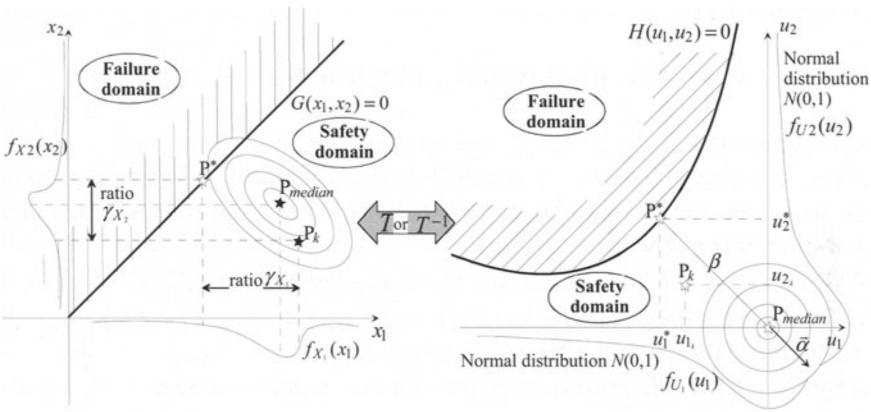


Figure -1. Reliability concepts.

From the structural mechanical modeling, the failure mode and the probabilistic modeling of each random variable X_i, the reliability analysis gives the Hasofer-Lind reliability index β [3] from the following optimization problem:

$$\beta = \min \sum_{i=1\ldots n} u_i^2 \quad \text{under} \quad G(x_i) \le 0 \;\text{ et }\; u_j = T_j(x_i) \qquad (1)$$

where T is the iso-probabilistic transformation from the X space of the physical variables to the U space of the uncorrelated standard Gaussian variables, $G(x_i)$ is the performance function ≤ 0 if the set of data $\{x_i\}$ is in the failure domain and > 0 otherwise. The optimal point resulting from the problem eq.(1) is the most likelihood failure point noted P^* whose coordinates are noted $\{u^*\}$ and $\{x^*\}$ respectively in the U and X spaces. Direct cosines $\{\alpha\}$ are very important values because they quantify the importance of each random variable on the structural reliability. The relation giving the most likelihood failure point is $\{u^*\} = -\beta\{\alpha\}$. Figure 1 summarizes reliability concepts in the case of a two random variables problem.

In most of the design methods, the most likelihood failure point coordinates are chosen as the design values and then:

$\gamma_i = x_i^* / x_{i_k}$ (loading type variable),

$\gamma_i = x_{i_k} / x_i^*$ (resistance type variable)

Figure 1 illustrates the computation of the γ_i-factors in the case of a two variables problem. For a given design situation, with a given *G*- function, the computation of the γ_i-factors is simple, it results from a single reliability study. Nevertheless, the aim of a calibrated design rule is to spare the designer the reliability analysis for each design situation giving him a rule which still works for a set of design situations.

2.2 Calibration methods principle [6]

Calibration methods are divided into six important steps:

Step 1: define the class of structure (the mechanical element, the failure mode(s) and the working area of the rule) for which the rule is aimed. This definition may contain limitations.

Step 2: define the goal of the rule by an expected target reliability index $\bar{\beta}$ of structures within the structure class defined in the first step. Existing rules must be used for evaluate the target reliability index.

Step 3: choose *L* specific design situations and evaluate their relative frequencies ω_j $(\sum_{j=1}^{L} \omega_j = 1)$ in the structure class. Experimental design theory can be used.

Step 4: design each situation *j* selected in step 3 according to the goal of the rule. Calculate, for each situation *j*, partial factors $\gamma_1^{(j)},...,\gamma_n^{(j)}$ using the method stated in section 2.1. These sets of partial factors are not calibrated because they just work for single situations.

Step 5: calculate *n* calibrated partial factors $\gamma_1,...,\gamma_n$ minimizing the fit between each situation reliability index and the target one.

Step *6:* verify the calibrated rule using random situation sampling within the structure class and calculate their corresponding reliability index.

The first three steps are very important to define the working area and the goal of the calibrated rule. Steps 4 and 5 are detailed in the following sections. The radial expansion method [7] seems to be the most realistic for step 4.

2.3 Radial expansion method [7] (step 3)

The radial expansion method gives partial coefficients $\gamma_1^{(j)},...,\gamma_n^{(j)}$ to be used in a particular design situation according to the target reliability level.

Figure 2 illustrates this method. From a design situation j defined by its performance function $G_j(x_i)$ in the X space and by $H_j(x_i)$ in the U space, a reliability analysis gives the coordinates of the most likelihood failure point noted $P^{*(j)}$ whose coordinates are noted $\{u^{*(j)}\}$ and the reliability index $\beta^{(j)}$. The design point noted $P_d^{(j)}$ whose coordinates are $\{u_d^{(j)}\}$ may be chosen according to the target reliability index and consequently may be chosen on the $\bar{\beta}$-radius sphere in the U space. An infinite number of choices are possible. The radial expansion method suggests choosing the projection of $P^{*(j)}$ onto the $\bar{\beta}$-radius sphere:

$$\{u_d^{(j)}\} = \bar{\beta}/\beta^{(j)} \{u^{*(j)}\}$$

Then, non-calibrated (because they come from a single situation j) factors $\gamma_1^{(j)},\ldots,\gamma_n^{(j)}$ are computed according to section 2.1.

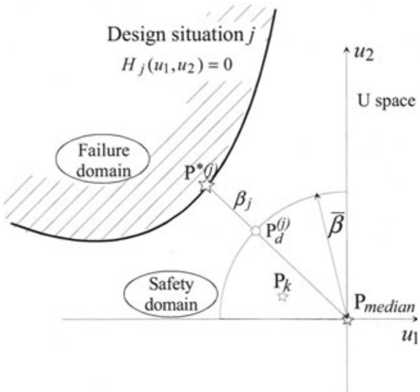


Figure -2. Radial expansion method principles.

2.4 Calibrated partial factors evaluation (step 4)

The goal is to find a set of partial factors $\gamma_1,\ldots,\gamma_n$ that work for all situations of the structure class.

The first possibility is to choose the most conservative rule that is:

$$\gamma_i = \max_j \gamma_i^{(j)}$$

It consists in choosing the most conservative rule. It generally has the disadvantage of increasing the reliability index of designed structures.

The second possibility is to take the weighted mean values:

$$\gamma_i = \frac{1}{L}\sum_{j=1}^{L}\omega_j \gamma_i^{(j)}$$

It leads to situations for which the target reliability level is not satisfied.

The third possibility, is to compute the better set of partial coefficients from the following optimization problem:

$$\underset{\gamma_i}{\text{Min}} \quad \Delta = \sum_j \omega_j (\tilde{\beta}_j(\gamma_i) - \bar{\beta})^2 \tag{2}$$

where $\tilde{\beta}_j(\gamma_i)$ is the reliability index of the design situation j modified by the set of partial coefficients. $\tilde{\beta}_j(\gamma_i)$, still hard to evaluate, is approximated using assumptions on the performance function $G_j(x_i)$. The problem eq. 2 becomes [6]:

$$\underset{\delta}{\text{Min}} \quad \Delta = \sum_j \omega_j (\bar{\beta}\langle\alpha_j\rangle.\{\delta\} - \bar{\beta})^2 \tag{3}$$

where $\{\alpha_j\}$ are the direct cosines of the design situation j and the optimization variables $\{\delta\}$ defines the design point coordinates in the U space $\{u^d\} = -\bar{\beta}\{\delta\}$. It is easier to compute the design point coordinates in the X space and then the calibrated partial factors.

3. PARTIAL FACTOR CALIBRATION OF THIN AXISYMMETRIC STIFFENED SHELLS

3.1 Step 1: structure class

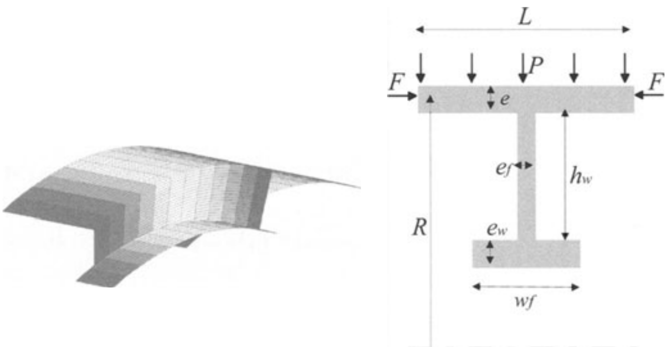


Figure -3. Structure class geometry, thin stiffened axisymmetric shell.

The structure class is composed by thin axisymmetric shells with a single stiffener (see fig. 3) and submitted to an external pressure P and to the edge force F coming from the effect of P on the bulkheads. The calibrated design rule concerns all the structures whose characteristic values are in the ranges table 1. The performance function is assumed to be:

$$G = P_{euler}(L_S, e, R, e_f, e_w, h_w, w_f, E, \upsilon, f_y) - P$$

where P_{euler} is the limit buckling load with the Euler assumptions (perfect elastic material properties, small displacement fields before buckling) computed from the following eigenvalue problem:

$$[K_0 + \lambda K_\sigma]v = 0, \qquad v \neq 0 \text{ (displacement field)} \tag{4}$$

where $[K_0]$ is the stiffness matrix of the initial geometry and $[K_\sigma]$ is the initial stress matrix. P_{euler} is computed multiplying the lowest eigenvalue λ by the initial stress load. For complex structures like stiffened shells, the eigenvalue problem eq. 4 may be solved using finite element method. INCA-STANLAX software [8] gives P_{euler} for complex thin axisymmetric shells.

Random variables are normally distributed, the coefficient of variation is 0.03 for geometrical variables and 0.02 for material variables and Poisson's ratio can be considered as deterministic [5]. For more simplicity, characteristic values are chosen here as the mean value of each random variable. The choice of fractile values could be another judicious possibility. The reliability software RYFES [9] is used for the probabilistic modeling and for the reliability analysis.

The stiffened shell reliability analysis requires a numerical link between RYFES and INCA-STANLAX using an external C++ routine [2].

Table - 1. Characteristic values of the structure class.

Geometry variables (m)							Material variables		
L_{sk}	e_k	R_k	e_{fk}	e_{wk}	h_{wk}	w_{fk}	E_k (Pa)	ν_k	f_{yk} (Pa)
[0.6-1]	[0.024-0.034]	[2.5-3.5]	0.024	0.01	0.18	0.12	200.10^9	0.3	290.10^6

3.2 Step 2: goal of the calibrated rule

The goal of the calibrated rule may be according to the implicit reliability index of the British Standard BS5500. Thirty random designs in the range of the structure class are achieved according to BS5500. Reliability analysis using RYFES and INCA-STANLAX gives the reliability index of each designed structure. Results can be found in fig. 5. The mean value of the reliability index is 4.00 and the standard deviation 0.17 (see table 4). The lowest obtained reliability index could be taken as the target of the rule. Nevertheless, it is better to propose a higher reliability index according to the standard deviation of the obtained BS5500 reliability indexes. The target reliability index of the calibrated rule is taken as 2 standard deviations smaller than the mean value that is to say:

$$\bar{\beta} = 3.66$$

3.3 Step 3: choice of L design situations

The choice of L extreme situations of the structure class seems to be interesting. Experimental design theory could give information on the best number of selected situations in order to limit this number. It is proposed to use a two-level experimental design (situations number 1 to 8 table 2) and the mean situation (number 9 table 2). The selected situations can be found in table 2. Their relative frequencies ω_j are the same. Other characteristic values are constant (see table 1).

Table - 2. Selected design situations (step 3).

Situations	1	2	3	4	5	6	7	8	9
L_{S_k}	0.6	0.6	0.6	0.6	1	1	1	1	0.8
e_k	0.024	0.024	0.034	0.034	0.024	0.024	0.034	0.034	0.029
R_k	2.5	3.5	2.5	3.5	2.5	3.5	2.5	3.5	3

3.4 Step 4: radial expansion method

After the radial expansion method, the partial factors can be found in table 3. Partial coefficients are normally higher than 1.

3.5 Step 5: calibrated partial coefficients

The three possibilities have been achieved. The first possibility is to consider the highest partial coefficients (in bold) in table 3. The calibrated rule is then:

$$G = P_{euler}\left(1.080R_k, 1.028L_{s_k}, \tfrac{1}{1.012}e_k, \tfrac{1}{1.004}e_{w_k}, \tfrac{1}{1.070}h_{w_k}, \tfrac{1}{1.021}e_{fk}, \tfrac{1}{1.021}w_{fk}, \tfrac{1}{1.013}E_k\right) - P \geq 0 \tag{5}$$

where P_{euler} is computed using INCA-STANLAX software from characteristic values in the bracket.

The second possibility is to compute the weighted mean value of partial coefficients. The calibrated rule becomes:

$$G = P_{euler}\left(1.078R_k, 1.027L_{s_k}, \tfrac{1}{1.010}e_k, \tfrac{1}{1.004}e_{w_k}, \tfrac{1}{1.068}h_{w_k}, \tfrac{1}{1.020}e_{fk}, \tfrac{1}{1.019}w_{fk}, \tfrac{1}{1.013}E_k\right) - P \geq 0 \tag{6}$$

The third possibility is to solve the optimization problem eq. 3. Results could be found in table 3. The calibrated rule becomes:

$$G = P_{euler}\left(1.082R_k, 1.024L_{s_k}, \tfrac{1}{1.022}e_k, \tfrac{1}{1.016}e_{w_k}, \tfrac{1}{1.076}h_{w_k}, \tfrac{1}{1.025}e_{fk}, \tfrac{1}{1.032}w_{fk}, \tfrac{1}{0.973}E_k\right) - P \geq 0 \tag{7}$$

Table - 3. Radial expansion method results.

Sit.		R	L_s	e	e_w	h_w	e_f	w_f	E
1	$\gamma^{(1)}$	1.077	1.027	1.010	1.004	**1.070**	1.019	1.019	1.013
2	$\gamma^{(2)}$	1.079	1.026	1.008	**1.004**	1.067	1.020	1.020	1.013
3	$\gamma^{(3)}$	1.077	1.027	**1.012**	1.004	1.069	1.018	1.019	1.013
4	$\gamma^{(4)}$	1.079	1.027	1.011	1.004	1.067	1.020	1.019	1.013
5	$\gamma^{(5)}$	1.077	1.026	1.009	1.003	1.069	1.020	1.020	1.013
6	$\gamma^{(6)}$	**1.079**	1.025	1.007	1.003	1.067	1.021	1.019	1.013
7	$\gamma^{(7)}$	1.076	1.028	1.011	1.003	1.069	1.020	1.020	1.013
8	$\gamma^{(8)}$	1.078	1.027	1.011	1.004	1.067	**1.021**	**1.021**	**1.013**
9	$\gamma^{(9)}$	1.078	**1.028**	1.010	1.003	1.068	1.019	1.019	1.013
	γ_{mean}	1.078	1.027	1.010	1.004	1.068	1.020	1.019	1.013
	γ_{opt}	1.082	1.024	1.022	1.016	1.076	1.025	1.032	0.973

3.6 Step 6: rules checking

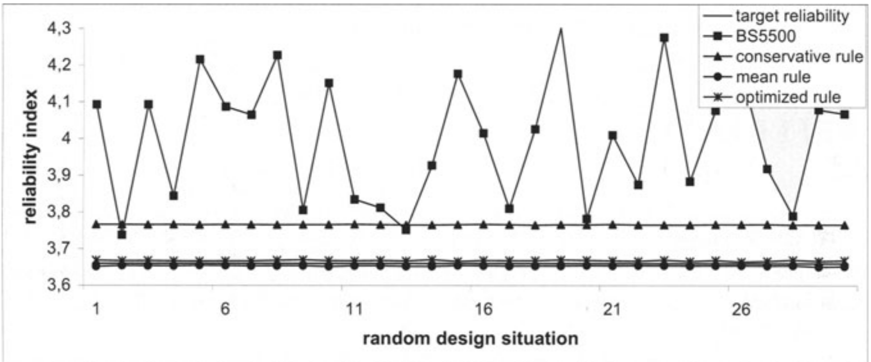


Figure -4. Reliability indexes of random designed structures.

Thirty random designs in the range of the structure class are achieved according to the three design rules eq. 5, 6 and 7. Then, a reliability study of each designed structure gives reliability indexes (see figure 4). The mean value and standard deviation of the reliability indexes distribution can be found in table 4.

Table - 4. Reliability results from BS5500 and calibrated rule eq. 5, 6, 7.

	BS5500	$\overline{\beta}$	Conservative rule (eq. 5)	Mean rule (eq. 6)	Optimized rule (eq. 7)
Mean value	4.00	3.66	3.77	3.65	3.67
Std deviation	0.17	-	≈ 0.00	≈ 0.00	≈ 0.00

Calibration methods give calibrated rules leading to structures which have more homogeneous reliability indexes than using BS5500. The conservative rule (eq.5) is not very efficient in this case because it leads to reliability indexes much higher than the target reliability index. The mean calibrated rule (eq.6) and the optimized one (eq.7) give reliability indexes near to the target reliability index and whose standard deviation is near to zero. Nevertheless, the mean rule is generally uninteresting because it gives reliability indexes lower than the target reliability index.

4. CONCLUSION

Calibration methods seems interesting in order to give partial “safety” factors according to a target reliability index. Such methods are numerous and it is still hard to propose a single best method because it is very problem dependent. They provide partial coefficients in order to design structures with more homogenous reliability indexes than those produce by empirical rules like BS5500.

The use of such methods on the field of shell stability is efficient and the optimized rule seems to be the best in this case.

REFERENCES

[1] *British Standards Institution*, *Specification for unfired fusion welded pressure vessels*, BS 5500, HMSO, 1996.

[2] Gayton N., Lambelin J.P., Lemaire M., Dimensionnement d’une coque de révolution conique raidie par une approche probabiliste, XVème Congrès Français de Mécanique, AUM, 3-7 Septembre 2001, Nancy, France.

[3] Ditlevsen O., Madsen H.O., *Structural Reliability Methods*, John Wiley and Sons, 1996.

[4] Lemaire M., Mohamed A., Finite element and reliability: a happy marriage?, 9th IFIP WG 7.5 Working Conference, Reliability and optimization of structural systems, September 2000, Ann Arbor, USA, published by the University of Michigan.

[5] Morandi A.C., Das P.K., Faulkner D., Finite element analysis and reliability based design of externally pressurized ring stiffened cylinders, RINA International Symposium, 1996.

[6] Kroon I.B., Decision theory applied to structural engineering problems, PhD thesis, Instituttet for bygningsteknik, Aalborg Universitet, Danmark, 1994.

[7] Devictor N., Marques M., Probabilistic optimization of safety coefficient : Comparison between the design point method and a global optimization method, Safety and Reliability, Schuëller & Kafka (eds), Balkema, Rotterdam, 1999.

[8] Combescure A., Static and dynamic buckling of large thin shells, In Nuclear Engineering and Design 92,North-Holand, Amsterdam, 1986.

[9] Mohamed A., Suau F., Lemaire M., A New Tool for Reliability Based Design with Ansys FEA, In ANSYS Conference and Exhibition, Houston, USA, pp. 3.13-3.23, 1996.

SOURCES OF ERROR IN THE DESIGN PROCESS

J. A Cooke, C. A McMahon and M. R North
Faculty of Engineering, University of Bristol, England

Abstract: This paper reports on an investigation into the causes of error in the design process based on a series of case studies in three companies. An initial study attempted to identify the points of occurrence of error in the case studies based on formal process models of the design activities. However, it was found to be difficult, if not impossible, to make a clear identification of a point of introduction of error in any of the case studies. In general, there was no single error source, but instead there was an accumulation of a number of contributory factors that when combined led to the error situation. Furthermore, it was generally not process issues that led to error but human factors. The paper identifies four categories of error inducing circumstance: unrecognised change of circumstances; lack of design effort; lack of integration of information and want of knowledge. Overall, it is asserted that the underlying issue is one of uncertainty handling and risk management.

Key words: design error, risk management, uncertainty, case studies.

1. INTRODUCTION

This paper arises out of a research project that set out to investigate the design processes followed by three companies, with the aim of identifying changes in procedures that would allow reduction in design error and rework. The approach that was initially adopted for the investigation was a very mechanistic "process orientated" methodology. The research team set about developing process and product models to represent the observed design activity and the product data developed during the design process. The process models decomposed the design process into series of activities

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 421–430.

with clearly defined inputs and outputs. It was postulated that, by using these models, it would be possible to identify where errors were introduced into the design process, since for each error found in a design activity, there must either be errors in the input information to the activity, or the activity itself must introduce error. The aim was to use these results to identify which activities or information sources were likely to lead to error, and thus to suggest procedural improvements to reduce the likelihood of the occurrence of error in future.

In order to carry out the investigation, the collaborating companies identified case studies in which error had occurred, and a programme of investigation comprising interviews and documentation studies was set up. However, despite clearly identifiable problems in the designs, it proved difficult if not impossible to make a clear identification of a point of introduction of error in any of the case studies. As the process and information models were developed it became apparent that in general there was no single error source, but instead it was found that in each case, although there were main candidates, a number of contributory factors combined to lead to the error situation. In this respect the findings reflect those identified in failure and disaster investigation [1][2], and also the motivation behind techniques such as the Toyota "Five Whys" process [3], which suggest that errors often arise from an accumulation of events, any of which by themselves may not lead to error. Further development of the study suggested that while there was not a clearly identifiable classification of errors in activities, it was possible to identify common circumstances that pertained in the design teams where errors occurred.

2. ERROR AND REWORK IN DESIGN

The present study was particularly concerned with rework in the design process – i.e. those circumstances in which it was necessary to repeat some aspect of the process, to correct an error or to deal with some change in requirements, but in which the eventual outcome in the product was satisfactory. Most other work on error in design has concerned errors that persist in the designed artefact – in particular those errors leading to failure. The subject has been investigated most widely by those concerned with the consequences of design error in structural failure, and these have tended to concentrate on the nature and effect of human error. For example, Stewart [4] suggests that reviews of statistical data indicate that up to 75% of structural failures are human errors [5], while Petroski [6] argues that human error is the most likely cause of fundamental errors made at a conceptual stage, which can be the most serious and elusive of design errors.

The theme of human error in structural failure is taken up by a number of authors, in particular through the analysis of failure case studies [7][8], and these and other works lead to a number of classifications and suggestions for terminology. Kaminetzky divided human errors into errors of knowledge (ignorance), errors of performance (carelessness and negligence) and errors of intent (greed). He contends, as does Blockley [1] that in fact all failures are ultimately due to human error, and he identified education and training as the only effective ways of reducing error. In further definitions, Norman [9] distinguishes between slips (unconscious errors) and mistakes (errors due to conscious action), and Ravindra [10] defines gross errors as deviations from standard practice which have an impact on plant safety or risk. He classifies errors as random (i.e. occur at random in time or space) or systematic (persistent), a distinction that is also made in experimental activity and by Kaminetzky. Ravindra also distinguishes between design errors (e.g. errors in modelling, assumptions, design data etc.), construction errors (e.g. misreading of drawings, defective installation etc.) and state-of-the art errors (e.g. failure modes or loads not envisaged, discovery of new external factors). There are other similar classifications in the literature (e.g. [11][12]).

These classifications influenced the present authors in the view of design that provided the initial theoretical basis of the current study. The overall view on which this was based is that a design evolves from a set of functional requirements at the start of the design process to an information set at the end of the process which explicitly describes the product and its manufacture [13]. This evolution process occurs by a set of activities that transform information during the evolution. Errors may be present in the information supplied to the activities or may be introduced by the activities themselves. As the information must be represented in some form in order to be communicated between activities a third source of error – in the representation and communication process - is introduced. We expected that it would be possible to identify errors in information in activities or in representation and communication, and to categorise them using the categories described above. The information-transformation view was based on McMahon's observations on design change [14]. Design errors were classified according to where (in which information) they occur, and how (in which activities) [15]. It was proposed that it should be possible to map the instances of error on to the attribute and process descriptions, and that, for each error identified in an output from an activity, there should be an error in an input to the activity, or an error should be introduced by the activity itself.

During the investigation it became apparent that this view is not entirely satisfactory. Often the evidence was incomplete and conflicting. Where a decision about the causes of an error was reached, it often came down to a number of possible causes that might or might not, in isolation, have led to problems, but which probably led to error as a combination of circumstances. It was difficult to judge the extent to which an individual circumstance was in itself a likely cause of a problem. This was for two reasons. Firstly, many design decisions involve making selections from a continuum, such as choosing a value for a safety factor in a calculation, and in these cases the boundary between satisfactory and unsatisfactory decisions is often very unclear. Secondly, many choices are based on limited and perhaps conflicting available evidence. In the case studies there were several instances where marginal but apparently justifiable decisions combined to lead to problems that required rework – and even with the benefit of hindsight the engineers involved felt that given the same circumstances they would make the same choices. Recognition that a problem might be developing required recognition that a particular combination of circumstances was emerging, or alternatively recognition that the decisions that had been taken were based on sufficiently uncertain information that a problem was likely. As the study developed it was observed that failures to recognise these likely error conditions seemed to be associated with particular circumstances in the companies being studied in which the likelihood or risk of error was not recognised.

3. ERROR CASES

3.1 Case study 1 - test failure

This case study involved the fatigue failure of components in a dynamically loaded assembly. The assembly was a development of an existing production component and the changes included a different material specification and a new feature - a bolted joint. Previous development design studies supported by testing had been carried out on the bolted joint aspect of the design over a number of years. During qualification testing of the new design a component in the region of the bolted joint failed and the investigation that followed found that the expected bolt preload had not been achieved. This had allowed movement at the joint which led to fretting at the joint surfaces resulting in a fretting fatigue failure. The bolts were tightened using a company standard torque tightening procedure and there was much experience in the method in other applications. However, in this case the tightening procedure under-performed.

The immediate cause of this failure was that the torque tightening procedure did not achieve the required preload as predicted by the analysis model. This resulted in gapping under high dynamic test loads and produced conditions where fretting could occur. In this case, however, the situation was compounded by bolt lubricant that did not perform as expected.

3.2 Case study 2 - manufacturing failure

A design improvement was introduced in which a panel that was originally reinforced by fabrication to provide the required thickness profile was redesigned to be a single panel chemically machined to the required profile. However despite the assurances of the manufacturer concerning the capability of the process it was found that the panel could not be formed into shape without warping.

The manufacturer was the instigator of the change and at the time was short of work. There was also considerable pressure on the design organisation to reduce costs and the new design was believed to be a good method of achieving this. However, the process to be used was new to the manufacturer and relatively untried. There was also a considerable time lapse of two years between the consultation with the manufacturer and award of the contract. By this time the manufacturer had plenty of work on and so the costs had gone up considerably.

The immediate cause of this failure was that differences in forming behaviour between a composite construction and a single piece construction were not recognised. However, underlying this were various political and cost pressures that caused a lack of consideration of the technical aspects of the design and the associated risks

3.3 Case study 3 - change of supplier

A component required redesign to introduce a more complex series of bends. The designers knew that the design was close to the limits of what could be achieved by the available manufacturing process. However, after consultation with the supplier, who had specialist machinery, a design was developed that was feasible. However, when the design was complete the supplier was changed. Although the new supplier had experience with manufacturing similar components in the past the extra demands made by the new component caused many manufacturing problems. Experienced designers would have known this to be the case but were not consulted when the decision was made to change the supplier.

The immediate cause of this failure was that the people responsible for choosing the supplier were not fully aware of the design implications. This was compounded by a component design that was on the limit of manufacturing capability for a chosen supplier with specialist machinery.

3.4 Case study 4 - CAD modelling failures

A customer required a piece of design work to be performed by a design sub-contractor, and one of the main project deliverables was CAD models of the design, constructed using a specified CAD system. The design team were generally very familiar with CAD systems including the one that the customer used. This was therefore considered to be a low risk part of the project. However, the customer required the design team to follow the customer's conventions and, in particular, to use parametric modelling rather than the more traditional non-parametric or explicit approach. It was found that the combination of the use of the (non-standard) customer's conventions, the use of an unfamiliar modelling method and also problems within the CAD system itself, along with a short timescale, meant that there was much difficulty in achieving the deliverable.

The immediate cause of this failure was that the combinatorial effect of three low risk requirements joining to result in a high risk was not recognised.

4. ERROR INDUCING CIRCUMSTANCES

As noted in section 2, the case studies indicate that errors occur when decisions are made based on conflicting or inadequate information, or where a series of decisions combine to move a design into error. The common circumstance is that the risk in the design has been high, but what are the conditions that are likely to lead to unrecognised high risk? The case studies suggest that a number of factors are important in this respect.

The first factor considered is *unrecognised change of circumstances*, which may occur when designers consider themselves to be experts in an aspect of the design which gives them confidence. This can lead to confirmation bias, where contradictory evidence is ignored [16]. Often this contradictory evidence occurs at a detailed level in the design process and may be as a result of many factors coming together which, on their own appear unimportant but, when combined, move the design away from the understood design domain. Confidence can lead to these factors being overlooked. The tendency of human behaviour to do this is understandable; the designer generally relies on reuse of previous designs and design experience. Small differences that may invalidate this approach are overlooked. Busby also highlights this issue stating that design re-use is an error prone process because designers did little research on the existing designs that they re-used [17].

The second factor considered is *lack of design effort*. This error inducing circumstance differs from the previous one in that here the designer does recognise the change in circumstances but does not sufficiently explore the design to ensure technical success and justify the design reasoning. It can be an issue of uncertainty handling. The designer uses approximate analyses based on uncertain information but does not quantify the level of uncertainty in the results. Alternatively, the designer may just dismiss certain factors as unimportant.

The third factor considered to be important is *lack of integration of information*, which causes what can often be considered as fairly basic mistakes and oversights. Designing is a process of bringing together information from multiple sources and resolving conflicts between requirements. Because there are many sources it is difficult for a single designer to have knowledge of them all. Systems are therefore required to allow the exchange of information to take place in a timely manner. The complexity of these systems means that they break down easily or contain loopholes and as a result the information is either not timely or not supplied at all. Because the systems are complex and numerous, errors can occur with relatively simple pieces of information.

In a number of the case studies, design decisions were made on the basis of limited and perhaps inadequate information, and this appears to be a normal state of affairs in design. Often in the design process the design has to progress based upon necessarily uncertain information. If the information is not available or does not exist it is important for the company to recognise and mitigate this by either taking steps to generate information or to make allowances for the greater level of uncertainty in the design. If the information is not accessible (e.g. because records are not indexed properly or because information is kept in inaccessible personal records) then the company can take steps to improve its knowledge systems. If the information is not timely (e.g. because analysis results are late) then the processes followed need to be addressed.

Errors due to unavailability of information are surprisingly rare, perhaps because it is relatively easy to identify this problem. More likely is that the designer has failed to consult the information. Examples of errors that arise from information being difficult to access include design modifications that propose previously rejected concepts and experts not being consulted. The authors have seen many examples of late design changes having to be made owing to the results of analyses arriving late.

The final error-inducing factor proposed here is *want of knowledge*. Designers cannot be expected to be experts in all the domains that they may encounter during their design work. Integrated teams are one method of overcoming this issue. However designers must have a certain level of competency. There must be a certain level of knowledge of routine design tasks and also enough knowledge to enable them to determine when the design reaches the limit of their knowledge and that external help is required. Ignorance of the limit of one's own knowledge is perhaps the most dangerous of all. Want of knowledge can also occur when design experts fail to keep abreast of current practice. In the case of CAD systems, experienced designers may be ignorant of the new tools. Various other examples have been noted in the case studies where mistakes were made because of out-of-date expertise.

4.1 Design failure as a combination of circumstances

In the case studies investigated, it has been found that failure generally occured due to a combination of two or more error inducing circumstances. There were no cases where a single event could be established as being entirely to blame for the failure. The one common theme was the failure to correctly identify when the uncertainty in the design was becoming unacceptably large so that high levels of risk were introduced into the project.

5. RISK AND UNCERTAINTY

Essentially, all of the above error inducing circumstances occur because the risks inherent in the emerging design are poorly understood. The risks can either not be identified or can be incorrectly assessed.

Risk results when the occurrence of an event is uncertain and there is a penalty associated with its occurrence. The design process is full of uncertainty. There is uncertainty in the external loads applied to the design, in the accuracy of the analysis, in the material properties and in the geometry. All these factors lead to uncertainty in the predicted attribute values of the performance of the design. Each attribute has a constraining value associated with it - the allowable stress is constrained by the strength of the material. Not only is the actual stress in the component uncertain but the actual strength of the material has some uncertainty associated with it. Design failure occurs when the attribute value exceeds its constraint. The purpose of applying safety factors is to try to take account of the uncertainties by ensuring that the predicted attribute value is far enough away from its predicted constraining value. The value of the safety factor is most often determined by empirical evidence based on historical design practice and by the judgement of the designer. Because each attribute is generally dependant on many other attributes, each of which is uncertain, it is very easy for the uncertainties to accumulate to a level not anticipated or recognised by the designer. It is under these circumstances that error likely conditions occur.

The importance of risk and uncertainty suggest that through modelling of these factors we may be able to obtain indications of the aspects of a design that may be subject to accumulated errors, and perhaps also to obtain early indication of where the decisions taken and values chosen for a design may be leading the design team into circumstances where there is a high risk of error. In order for such an approach to be successful, however, engineers need to significantly increase the extent to which they collect and collate information about uncertainties and imprecision in engineering data and in engineering methods.

6. CONCLUSIONS

The authors initially approached this research from a very mechanistic, process orientated viewpoint assuming that errors could be traced back to specific events that caused them. This has proved not to be the case and it has been found that errors arise from an accumulation of events leading to an unforeseen circumstance in the design.

In the five case studies encountered the sources of errors have been complex and have resulted from a combination of circumstances. A categorisation of the circumstances leading to error has identified four categories as: unrecognised change of circumstances, lack of design effort, lack of integration of information and want of knowledge. The error categories generally do not result from a failure of the processes but are more related to the attitude, training and competence of the designers. This demonstrates that the human factors are the most important aspects in the design process. Processes are secondary to these.

The importance of uncertainty handling has been identified in that a design fails when the level of uncertainty is such that a design attribute can exceed its constraining value. Because there are few formal methods for handling uncertainty the skill and judgement of the designers is relied upon. All of the above error circumstances result from a failure to handle uncertainty correctly.

The categories identified demonstrate that process change is not the most suitable method of reducing error. More important are the human factors. For a product to have more chance of succeeding requires the design process to proceed in an atmosphere of information sharing, risk awareness and design visibility to enable the application of expertise.

7. ACKNOWLEDGEMENTS

The authors are grateful for the support and contribution of their industrial collaborators, and to the UK Engineering and Physical Sciences Research Council for financial support under grant GR/K 68028.

REFERENCES

1. Blockley, D.I., *The Nature of Structural Design and Safety*, Chichester: Ellis Horwood, 1980
2. Petroski, H, *To Engineer is Human*, New York: St. Martins Press, 1982.
3. Shibata, H, Problem solving: definition, terminology and patterns, http://www.mediafrontier.com/Article/PS/PS.htm, 1998
4. Stewart, M.G., Simulation of human error in reinforced concrete design, Research in Engineering Design 4(1), 1992, pp.51-60
5. Brown, C.B. and Yin, X., Errors in structural engineering, ASCE Journal of Structural Engineering, 114, 1988, pp.2575-2593
6. Petroski, H., Paconius and the pedestal for Apollo: a case study of error in conceptual design, Research in Engineering Design, 3, 1991, pp.123-128
7. Piésold, D. D. A., *Civil Engineering Practice: Engineering Success by Analysis of Failure*, London: McGraw-Hill, 1991
8. Kaminetzky, D., *Design and Construction Failures: Lessons From Forensic Investigations*, New York: McGraw Hill, 1991
9. Norman, D.A., Categorisation of action slips, Psychological Review, 88, 1981, pp.1-15
10. Ravindra, M.K., Gross errors, seismic margins and seismic PRAs, Modeling Human Error in Structural Design and Construction, ASCE, Ann Arbor, Mi, 1986, pp.113-121
11. Fitzsimmons, N., On modeling errors in the construction industry, Modeling Human Error in Structural Design and Construction, ASCE, Ann Arbor, Mi, 1986, pp.190-194
12. Kletz, T., *An Engineer's View of Human Error*, 2nd Edition, London: Institution of Chemical Engineers, 1991
13. Bañares-Alcántara, R., Representing the engineering design process: two hypotheses, Computer-aided Design, 23, 1991, pp.595-603
14. McMahon, C.A., Observations on modes of incremental change in design, Journal of Engineering Design, 5(3), 1994, pp.195-209
15. McMahon, C.A., Cooke, J.A. and North, M.R., A classification of errors in design, International Conference on Engineering Design, ICED'97, Tampere, 1997, pp.119-124
16. Prabhu, P., Sharit, J. and Drury, C., Classification of temporal errors in CIM systems: development of a framework for deriving human-centred information requirements, International Journal of Computer Integrated Manufacturing, 5(2), 1992, pp.68-70
17. Busby, J.S., Why designers don't learn effectively from feedback, International Conference on Engineering Design, ICED'97, Tampere, 1997, pp.105-110.

A MODEL-BASED FRAMEWORK FOR ROBUST DESIGN

K. Al-Widyan, J. Angeles and J. Cervantes-Sanchez
Department of Mechanical Engineering and Centre for Intelligent Machines McGill University, Montreal, Quebec, Canada

Abstract: In this paper we develop an as-yet-missing theoretical framework as well as a general methodology for model-based robust design. At the outset, a distinction is made between three sets: the set of design variables, grouped in the n-dimensional vector $\mathbf{x}$, which are to be assigned values as an outcome of the design job; the set of design-environment parameters (DEP), grouped in the ν-dimensional vector $\mathbf{p}$, over which the designer has no control; and the set of performance functions, arrayed in the m-dimensional vector $\mathbf{f}$, representing the functional relations among performance, design variables and DEP. Resorting to the mathematical model available for the object under design, an $m \times \nu$ design performance matrix $\mathbf{F}$, mapping the space of relative variations of $\mathbf{p}$ into that of relative variations of $\mathbf{f}$, is derived. Moreover, two pertinent concepts are introduced: the *design sensitivity matrix*, which plays a major role in the transmission of the variations of $\mathbf{p}$ into variations of f, and its associated *bandwidth*, defined as the logarithm of the square root of the ratio between the maximum to the minimum singular values of the design performance matrix, measured in decades. A result stating the relation between the bandwidth of a matrix and its inverse is shown. Consequently, the aforementioned bandwidth represents an index for evaluating the robustness of a design. To demonstrate our approach, case studies are included

Key words: Robust design, sensitivity, performance, and environment

1. INTRODUCTION

The impressive development of mathematical programming in the sixties, mostly due to the work of Bellman in the USA [1], paved the way toward the

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 431–442.

development of *optimum design*. In this context, the design task is formally defined by means of an *objective function* that quantifies a performance index whose minimization or maximization leads to an optimum performance. Techniques to solve the problem of optimum design are well developed, but they suffer of some inherent shortcomings, mainly the difficulty of guaranteeing a *global optimum* in the most general cases. More recently, with the development of *genetic algorithms* (GA) by John Holland [2], as well as other techniques like *simulated annealing* (SA) [3], some of those shortcomings have been overcome. Indeed, GA and SA are paradigms of what is known as *statistical methods* of optimization. While the scope of these methods is much broader than that of methods based on mathematical programming, they entail also some drawbacks, mostly the lack of general criteria guaranteeing convergence.

Whether a mathematical programming or a statistical approach is used to solve a design problem, the nature of the task is essentially the same: minimize a cost or maximize a profit. Nothing in the underlying formulation really addresses the issue of sensitivity of the performance of the optimum design to changes in the design operation conditions, also referred to as *nominal design conditions*. Indeed, when formulating a design task with the above approaches, the designer must assume *nominal values* of parameters defining the design conditions, e.g., loads, working temperatures, cruising speeds and altitudes, and so on. In real life, however, actual load values as well as other *design parameters* defining the operation conditions are seldom known, for those conditions can vary over broad ranges. How sensitive the performance of the optimum design to changes in the values of the design parameters is cannot be anticipated by the foregoing approach. Robust design is an alternative approach that aims at the design of products whose high-quality performance is maintained even in the presence of large variations in the design environment. The sensitivity of the performance to variations in the design environment is, generally speaking, a major issue within the methodology of robust design. Along these lines, Gadallah and El-Maraghy [4] identified the design sensitivity matrix as the Hessian of the objective function. However, associating the Hessian with the design job is limiting in that this ties robust design to the objective-function approach, which need not be the case. On the other hand, Zhu and Ting [5] consider variations in the design variables instead of variations in design environment parameters.

The roots of robust design can be traced to Taguchi's work on quality control and the design of experiments in the fifties [6]. However, Taguchi's work a) was not available in English but up until the eighties, and b) lacks a model-based theoretical framework. Indeed, the philosophy of robust design was originally proposed in connection with the design of experiments, in areas where a mathematical model was not available; it has not as yet fully impacted on design engineering, mostly because it regards the design object as a black box, while current design engineering is largely model-based. Generally speaking, robust design is a philosophy aiming at rendering a product performance as insensitive to variations in the operation conditions as possible. In his work, Taguchi introduced two concepts that are fundamental to engineering design [7] [8]:

a) The signal-to-noise (S/N) ratio, that measures the sensitivity of a design performance to changes in both environment and operation conditions. The S/N ratio has been used in the communication industry since the turn of last century as a means of assessing the quality of communication systems. Dr. Taguchi brought the concept into the realm of quality engineering and developed it to be the index for the quality of all kinds of products (the larger the S/N ratio, the more robust the performance), and
b) The loss function, that measures the loss of society by virtue of a flawed design. The importance of the loss to society caused by a flawed design cannot be overstated; for example, the catastrophe of a Concorde in July 2000 is attributed to the flawed location of the main landing gear, before the engine inlet. The crash of the Concorde caused heavy losses to society in terms of life and property.

2. FORMULATION OF THE ROBUST DESIGN PROBLEM

Engineering design is increasingly becoming model-based, in that their complexity calls for a mathematical model involving multiple quantities, some of which are to be decided on by the designer with the purpose of meeting performance specifications—e.g., the thrust that an aircraft engine must deliver at a given rpm—under given environment conditions—engine must operate at a specified ambient temperature and at a given ambient pressure. We thus classify the various quantities occurring in the model into

• *Design variables*: Those quantities that the designer has to find, as an outcome of the design job, with the purpose of meeting performance specifications under the given conditions. We shall denote by **x** the *n*-dimensional vector of design variables:

$$\mathbf{x} \equiv [\, x_1 \quad x_2 \quad \cdots \quad x_n \,]^T$$

• Design Environment Parameters (DEP): Those quantities over which the designer has no control, and that define the conditions under which the designed object must operate. They will be grouped in the array:

$$\mathbf{p} \equiv [\,p_1 \quad p_2 \quad \cdots \quad p_v\,]^T$$

• Performance Functions: Those that are used to represent the performance of the design in terms of design variables and environment parameters. The performance parameters will be grouped in the m-dimensional vector $\mathbf{f}$, namely,

$$\mathbf{f} \equiv [\,f_1 \quad f_2 \quad \cdots \quad f_m\,]^T$$

Let us assume that the components $\{x_i\}_1^n$ of $\mathbf{x}$ have been normalized so that they are all dimensionless, the normalization being done by dividing all physical design variables by their *nominal* values, in Taguchi's terminology. We have, in this context, a set of functions $\{f_i(\mathbf{x};\,\mathbf{p})\}_1^m$ of the design variables and the environment parameters that represent the *performance* of the *object under design*. Thus, we have the functional relation

$$\mathbf{f} = \mathbf{f}(\mathbf{x};\mathbf{p})$$

Robust design aims at rendering the performance vector $\mathbf{f}$ of a design as insensitive to variations $\Delta\mathbf{p}$ as possible. Thus, if we introduce a variation $\Delta\mathbf{p}$ in the DEP, while keeping $\mathbf{x}$ fixed, then,

$$\mathbf{f}(\mathbf{x};\mathbf{p}_o + \Delta\mathbf{p}) = \mathbf{f}(\mathbf{x};\mathbf{p}_o) + \nabla\mathbf{f}(\mathbf{x};\mathbf{p}_o)\Delta\mathbf{p} + \cdots + \mathrm{HOT} \tag{1}$$

where HOT represents higher-order terms. Taking into account only the first-order terms, we can write

$$\nabla\mathbf{f} = \mathbf{F}\,\nabla\mathbf{p} \tag{2a}$$

where $\Delta\mathbf{p} \equiv \mathbf{p} - \mathbf{p}_o$ and $\mathbf{F}$ is the $m \times v$ Jacobian matrix of $\mathbf{f}$ with respect to $\mathbf{p}$, i.e.,

$$\mathbf{F} \equiv \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \tag{2b}$$

We call $\mathbf{F}$ *the performance matrix* of the design at hand. Let S be the *sensitivity* of the design performance to changes in the DEP parameters, i.e.,

$$S \equiv \frac{\|\Delta \mathbf{f}\|}{\|\Delta \mathbf{p}\|} \Rightarrow S^2 = \frac{\Delta \mathbf{p}^T \mathbf{F}^T \mathbf{F} \Delta \mathbf{p}}{\Delta \mathbf{p}^T \Delta \mathbf{p}} = \frac{\Delta \mathbf{p}^T \mathbf{S} \Delta \mathbf{p}}{\Delta \mathbf{p}^T \Delta \mathbf{p}} \tag{3}$$

with the *sensitivity matrix* **S** defined as

$$\mathbf{S} \equiv \mathbf{F}^T \mathbf{F} \tag{4}$$

which is apparently a positive-definite $\nu \times \nu$ matrix. Consequently, matrix **S** has ν linearly independent eigenvectors denoted as $\{\mathbf{e}_i\}_1^\nu$ and ν nonnegative real eigenvalues $\{\lambda_i\}_1^\nu$ [9]. Moreover, matrix **S** can be written in diagonal form as

$$\mathbf{S} = \mathbf{E}^T \mathbf{\Lambda} \mathbf{E} \tag{5}$$

where the columns of matrix **E** are the ν linearly independent eigenvectors of **S**, namely,

$$\mathbf{E} = [\mathbf{e}_1 \; \mathbf{e}_2 \; \ldots \; \mathbf{e}_\nu] \tag{6}$$

and $\mathbf{\Lambda}$ is given by

$$\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_\nu) \tag{7}$$

Moreover, if we let

$$\Delta \mathbf{p} = \mathbf{E}^T \mathbf{y} \tag{8}$$

then, eq. (3) can be cast in the form

$$S^2 = \frac{\lambda_1 y_1^2 + \cdots + \lambda_\nu y_\nu^2}{y_1^2 + \cdots + y_\nu^2} \tag{9}$$

Letting $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_\nu$ be the eigenvalues of the design sensitivity matrix **S**, S^2 is bounded in the range

$$\lambda_1 \le S^2 \le \lambda_\nu \tag{10}$$

Accordingly, the sensitivity S is never below λ_1 and never above λ_ν . For a robust design, we want $\lambda_{max}(=\lambda_\nu)$ to be a minimum, i.e., the problem can be formulated as a minimax problem of the form

$$\min_{\mathbf{x}} \max_{i} \{\lambda_i\}_1^\nu \tag{11}$$

Obviously, λ_{max} is a minimum when all λ_i are equal, i.e., when **S** is proportional to the identity matrix and **F** is *isotropic*. An isotropic matrix is one with identical and nonzero singular values [10].

Geometrically speaking, the above relation defines an n-axis ellipsoid, namely, the *sensitivity ellipsoid*, whose axes have lengths inversely proportional to the square roots of the eigenvalues of **S**. As a consequence, the more the eigenvalues of **S** are widespread, the more sensitive the design is to changes in he DEP. The least sensitive design is, then, one with an "ellipsoid" of identical axis-lengths, i.e., a sphere.

We now introduce

Definition: The *bandwidth b* of a matrix **A is** defined as

$$b = \log\sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$

where λ_{max} and λ_{min} are the maximum and minimum eigenvalues of $\mathbf{AA}^T$, respectively.

Remark 1: The above definition measures b in decades.

Remark 2: The bandwidth of an isotropic matrix is zero, while that of a singular matrix tends to infinity, i.e.,

$$0 \le b \le \infty$$

Remark 3: The bandwidth of a matrix **S** measures the *elongation* of the sensitivity ellipsoid, defined as the ratio of its longest to its shortest semiaxes.

Definition: The sensitivity of a design, measured in decades, is the bandwidth of the design sensitivity matrix.

Now we have

Theorem: *A matrix and its inverse have identical bandwidths.*

The proof of the foregoing theorem is straightforward, and not be included here.

3. EXAMPLES

In order to illustrate the application of the proposed methodology, we include two examples.

3.1 The Robust Design of a Low-Pass Filter

We consider here the robust design of the *RL* circuit shown in Fig. 1, first studied by Taguchi [6] and then by Wilde [11]. The design variables are the resistance R and the inductance L, to be determined with the aim of keeping the current I_o at a nominal value of 10 amperes, while the amplitude V_o of the excitation voltage $v(t)=V_o\cos(\omega t)$ and its frequency ω undergo considerable variations beyond the control of the designer.

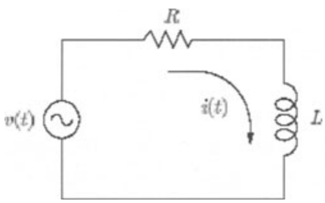


Figure -1. A low-pass filter

For this filter, the steady-state current i(t) is harmonic: $i(t)=I_o\cos(\omega t+\phi)$, where I_o and are ϕ the magnitude and the phase of $i(t)$. These are given by

$$I_o=\frac{V_o}{\sqrt{R^2+\omega^2L^2}},\quad \phi=\tan^{-1}\left(\frac{\omega L}{R}\right) \tag{12}$$

The design vector, the DEP vector, and the performance vector are

$$\mathbf{x}=\begin{bmatrix}R\\L\end{bmatrix},\quad \mathbf{p}=\begin{bmatrix}V_o\\\omega\end{bmatrix},\quad \mathbf{f}=\begin{bmatrix}I_o\\\phi\end{bmatrix}$$

Assume that R and L are fixed but ΔV_0 and $\Delta\omega$ induce variations ΔI_0 and $\Delta\phi$. Moreover, let

$$\Delta\mathbf{p} = \begin{bmatrix} \Delta V/V \\ \Delta\omega/\omega \end{bmatrix}, \quad \Delta\mathbf{f} = \begin{bmatrix} \Delta I_o/I_o \\ \Delta\phi \end{bmatrix}$$

be the nondimensional variations of vectors $\mathbf{p}$ and $\mathbf{f}$. The performance matrix is, then

$$\mathbf{F} = \begin{bmatrix} 1 & -\alpha^2 \\ 0 & \alpha\sqrt{1-\alpha^2} \end{bmatrix}, \quad \alpha = \left(\frac{I_o}{V_o}\right)\omega L \tag{13}$$

It can be noted that $\mathbf{F}$ relates variations in the DEP vector with those in the performance vector. Now, the sensitivity matrix $\mathbf{S}$ is calculated as

$$\mathbf{S} = \mathbf{F}^T\mathbf{F} = \begin{bmatrix} 1 & -\alpha^2 \\ -\alpha^2 & \alpha^2 \end{bmatrix} \tag{14}$$

Figure (2) shows a plot of the bandwidth b of the sensitivity matrix $\mathbf{S}$ versus the inductance L. It is noteworthy that $b \to \infty$ when $L = 28.9$ mH, where $\mathbf{S}$ is singular. The minimum of the bandwidth of $\mathbf{S}$ occurs at

$$\alpha = \frac{\sqrt{3}}{3} \Rightarrow L = \frac{\sqrt{3}\,V_o}{3I_o\omega} = 16.7 \text{ mH}$$

The foregoing solutions are compared with those found by Taguchi and by Wilde in Table 1, for V_o=100 V and ω = 55 Hz. Note that the bandwidth of Taguchi's solution is about 30% higher than that of our solution, while that of Wilde's solution is about 22% higher than the bandwidth of our solution.

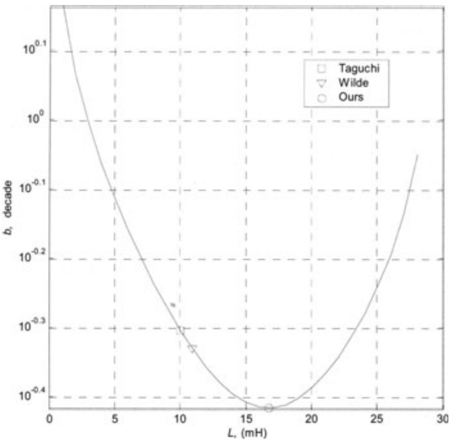


Figure -2. Variation of b vs. L

Table 1. Comparison of results; Taguchi's, Wilde's and ours

Variable	Taguchi	Wilde	Ours
R (Ω)	9.5	9.5	8.2
L (mH)	10.0	10.9	16.7
B	0.4959	0.4671	0.3828

All designs meet the target of $I_o = 10$ amperes, but, as appearing in Table 1, variations in the DEP leading to smaller variations in the performance are associated with a lower bandwidth. For instance, if the components of **p** vary by, say, 10%, then the corresponding variations in **f** are summarized in Table 2.

Table 2. Variations of **f** for $\Delta V_o/V_o$=0.1 and $\Delta\omega/\omega$=0.1

Variable	Taguchi (%)	Wilde (%)	Ours (%)
$\Delta I_o/I_o$	8.8058	8.5811	6.6694
$\Delta\varphi$	3.2428	3.4893	4.7131
$\Delta\mathbf{f}$	7.3000	7.1400	6.3300

3.2 Robust Design of a Helical Spring

Figure 3 shows a helical spring loaded by the axial force F. We denote by D the mean spring diameter, d the wire diameter, and N the number of turns. In addition, the stiffness k of the spring, the shear stress in the spring, and the natural frequency ω of vibration of the spring are given by

$$k = \frac{d^4 G}{8D^3 N}, \quad \tau = K_s \frac{8FD}{\pi d^3}, \quad \omega_n = \frac{1}{2}\sqrt{\frac{kg}{w}} \tag{15}$$

where G is the shear modulus, w is the weight of the spring, and K_s is the shear stress correction factor and g is the gravity acceleration.

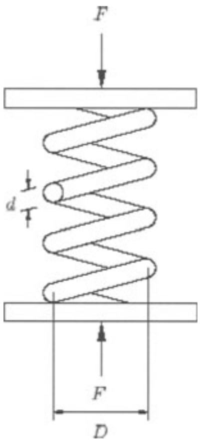


Figure -3. Axially loaded helical spring

Based on the dynamics model, we have

$$\frac{X_o}{F} = \frac{k}{1-r^2} \tag{16}$$

where X_o and F_o are the amplitudes and $r(=\omega/\omega_n)$ is the frequency ratio, with ω denoting the excitation frequency of the harmonic excitation $F(t)=F_o\cos(\omega t)$. Moreover, the design vector, the DEP vector, and the performance vector are now

$$\mathbf{x} = [d, D, N]^T, \ \mathbf{p} = [F_o, \ \omega]^T, \ \mathbf{f} = [\tau, \ X_o]^T$$

Since X_o is a function of both F_o and ω_o, we can write

$$\frac{\Delta X_o}{X_o} = \frac{\Delta F_o}{F} + \frac{2r^2}{1-r^2}\left(\frac{\Delta\omega}{\omega}\right), \quad \frac{\Delta\tau}{\tau} = \frac{\Delta F}{F} \tag{17}$$

In vector form

$$\begin{bmatrix} \Delta\tau/\tau \\ \Delta X/X \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & \dfrac{3r^2}{1-r^2} \end{bmatrix} \begin{bmatrix} \Delta F/F \\ \Delta\omega/\omega \end{bmatrix}$$

Figure (4) shows a plot of the bandwidth b of the sensitivity matrix **S** versus the frequency ratio r. The minimum bandwidth of the design sensitivity matrix occurs when $r = 0.6436$, with a bandwidth of 0.3828 decades. Besides, $b \rightarrow \infty$ when $r = 1$, at which value **S** is singular. Note also that the bandwidth remains essentially constant for $r \geq 5$.

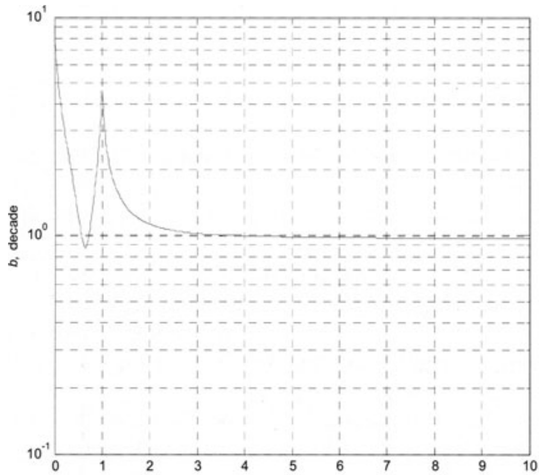


Figure -4. Variation of b vs. r

4. FURTHER REMARKS

The two examples discussed above are quite simple. Both involve two performance parameters and two design environment parameters. As a consequence, the performance matrix is square, of 2×2 in fact. In general, **F** is an $m \times \nu$ matrix. If $m > \nu$, then the definition of **S** from eq. (4) is straightforward; this definition leads, for a full-rank **F**, a nonsingular $\nu \times \nu$ square, positive-definite matrix **S**. If, however, $m < \nu$, then **S**, as defined in eq. (4), is singular; however, if **F** is of full-rank, then **S** will have exactly $(\nu - m)$ zero singular values; its m positive singular values are the semiaxes of the sensitivity ellipsoid. Note that these positive singular values are the square roots of the eigenvalues of $\mathbf{FF}^T$ instead.

5. CONCLUSIONS

A theoretical framework for robust design, as applied to engineering systems, through the minimization of the sensitivity of the performance against the operation and environmental conditions, was proposed here.

We showed that minimum sensitivity, and hence, maximum robustness, implies a minimum bandwidth of the sensitivity matrix. The robustness of the design is quantified by the bandwidth of the design sensitivity matrix. Two examples are used to illustrate the formulation and demonstrate the approach.

REFERENCES

[1] Bellman, R. and Dreyfus, S., *Applied Dynamic Programming*, Princeton University, New York, 1962.

[2] Holland, J., *Adaptation in Natural and Artificial Systems*, Michigan Press, Ann Arbor, Mich., 1975.

[3] Kirkpatrick, S., Gelatt, C., and Vecchi, M., ``Optimization by Simulated Annealing,'' *Science*, Vol. 220, pp. 671-680, 1983.

[4] Gadallah, M., El-Maraghy, H.,``A Statistical Optimization Approach to Robust Design," in Utpal, R., Usher, J., Hamid, R. (editors), *Simultaneous Engineering*, Gordon and Breach Science Publishers, Singapore, pp. 135-158, 1999.

[5] Zhu, J. and Ting, K., ``Performance Ditribution Analysis and Robust Design", *Trans. ASME, Journal of Mechanical Design*, Vol. 123, pp. 11-17, 2000.

[6] Taguchi, G., ``The Development of Quality Engineering,", *The ASI Journal*, Dearborn, MI, Vol. 1, No. 1, pp. 5-29, 1988.

[7] Taguchi, G., *Taguchi on Robust Technology Development: Bringing Quality Engineering Upstream*, ASME Press, New York, 1993.

[8] Wu, Y. and Wu, A, *Taguchi Methods for Robust Design*, ASME Press, NY, 2000.

[9] Golub, G. and Van Loan, C, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1989.

[10] Strang, G., *Linear Algebra and Its Applications*, Third Edition, Harcourt Brace Jovanovich College Publishers , New York, 1988.

[11] Wilde, D, ``Monotonicity Analysis of Taguchi's Robust Circuit Design Problem", *Trans. ASME, Journal of Mechanical Design*, Vol. 1

STRUCTURAL DESIGN OF UNCERTAIN MECHANICAL SYSTEMS USING INTERVAL ARITHMETIC

O. Dessombz, F. Thouverez, J-P. Laîné and L. Jézéquel
Laboratoire de Tribologie et Dynamique des Systèmes, Ecole Centrale de Lyon, France

abstract : This paper deals with a method to analyze the modal characteristics of structures including bounded uncertain parameters. These parameters can be uncertain, variable or not set at a design stage. We will use interval theory to propose a formulation adapted to finite element mechanical problems, taking into account the construction of mass and stiffness matrices. We will consider two kinds of problems: finding the bounds of static problem solutions, and the envelopes for transfer functions. A new numerical method will be used, based on the formulation proposed above. We will emphasize the efficiency of this method on simple mechanical problems. A frame structure will be studied, for which an overlap phenomenon appears on the transfer function envelope.

1. INTERVAL ARITHMETIC, INTERVAL MATRIX ARITHMETIC

The interval arithmetic was first introduced by Moore [7], who was interested in error propagation due to truncation of the mantissa in computer computations. Many publications (in particular the book of Alefeld and Herzberger [1]) give the basic and advanced concepts of this theory. We will just mention the basic concepts required for a good comprehension of this paper.

A real bounded and closed interval is defined by a real pair a and b, $a \leq b$: $\mathbf{x}=[a,b]=\{x:a \leq x \leq b\}$

Boldface, lower cases, underscores and overscores respectively denote Intervals, scalars, lower bounds of intervals and upper bounds of intervals.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 443–452.

$$\mathbf{x} = [\underline{\mathbf{x}}, \overline{\mathbf{x}}] \tag{1}$$

The set of real intervals is denoted as $\mathbb{I}\mathbb{R}$. An interval can then be identified to a couple of real numbers, for which arithmetic laws can be built.

1.1 Operations on intervals

The four classical arithmetic operations are defined:

$$\begin{aligned}
&\mathbf{x}+\mathbf{y} = [\underline{\mathbf{x}}+\underline{\mathbf{y}}, \overline{\mathbf{x}}+\overline{\mathbf{y}}], \qquad \mathbf{x}\text{ - }\mathbf{y} = [\underline{\mathbf{x}}-\overline{\mathbf{y}}, \overline{\mathbf{x}}-\underline{\mathbf{y}}] \\
&\mathbf{x} * \mathbf{y} = [\min(\underline{\mathbf{x}}\underline{\mathbf{y}}, \underline{\mathbf{x}}\overline{\mathbf{y}}, \overline{\mathbf{x}}\underline{\mathbf{y}}, \overline{\mathbf{x}}\overline{\mathbf{y}}), \max(\underline{\mathbf{x}}\underline{\mathbf{y}}, \underline{\mathbf{x}}\overline{\mathbf{y}}, \overline{\mathbf{x}}\underline{\mathbf{y}}, \overline{\mathbf{x}}\overline{\mathbf{y}})]
\end{aligned} \tag{2}$$

After the definitions of these simple operations on intervals, we are likely to use functions of interval variables. Evaluation of functions leads to a (sometimes too important) problem of conservatism.

For instance, let us calculate the result of the operation $2\mathbf{x}-\mathbf{x}$. We should get $\mathbf{x}$ as a result, but if the previous laws are applied, the result is quite different: $2\mathbf{x}-\mathbf{x} = [2\underline{\mathbf{x}}-\overline{\mathbf{x}}, 2\overline{\mathbf{x}}-\underline{\mathbf{x}}]$ has been calculated as $2\mathbf{x}-\mathbf{x} = \{2x - y \,/\, x \in \mathbf{x}, y \in \mathbf{y}\}$. The dependence between the multiple occurrences of $\mathbf{x}$ has been forgotten.

Actually all of the arithmetic expressions concerning intervals suffer from that conservatism problem. If the interval variables appear only once in the expression, the calculation gives the exact solution of the problem. (Moore, see [7]).

The commutativity and associativity properties are preserved for the sum and multiplication operation, but the distributivity is not. Only an inclusion rule is true (sub-distributivity): $\mathbf{x}(\mathbf{y}+\mathbf{z}) \subset \mathbf{xy}+\mathbf{xz}$

If x is a real number, this is an equality. As mentioned previously, it is the problem of dependence (on the left side, the variables appear only once while $\mathbf{x}$ appears twice on the right side).

1.2 Extension to n-dimension

Definitions given above can be extended to n-dimensional problems. Interval arithmetic can be applied to interval vectors and interval matrices.
An interval vector $\{\mathbf{x}\}$ is a vector whose components are intervals.
$\mathbb{I}\mathbb{R}^n$ will denote the set of n-dimensional interval vectors. An interval vector is an n-dimensional parallelepiped whose sides are parallel to the axis.

An interval matrix $[\mathbf{A}]$ is a matrix whose components are intervals $\mathbb{IR}^{mxn}$ denotes the set of interval matrices of dimension mxn.
The special properties of interval matrices have been investigated for example by Ning & al and Rohn in [8, 11].

1.3 Solving linear systems

If we are interested in the static and dynamic behavior of an industrial mechanical structure, one has to consider Finite Element Modeling, which leads to matrices (such as stiffness, mass, or damping matrix). Thus, linear systems of equations are to be solved. If some of the mechanical parameters are uncertain at the design stage, or are variable, such as the weight of a tank, they can be modeled using interval theory. The uncertain parameters can be geometrical ones (length, thickness, clearance...), or physical ones (Young's Modulus...). Then the matrices given by the Finite Element theory are interval matrices, and the problem is written as $[A]\{x\}=\{b\}$, with $[A]\in[\mathbf{A}]$ and $\{b\}\in\{\mathbf{b}\}$.

Although several problems can be distinguished, as achieved by Chen and Ward in [2] and by Shary in [13], we will focus exclusively in this paper on the outer problem which is defined as $\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})$, where $[\mathbf{A}]$ is an interval matrix and $\{\mathbf{b}\}$ an interval vector:

$$\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})=\{x\in R^n / [A]\in[\mathbf{A}],\ \{b\}\in\{\mathbf{b}\}\} \qquad (3)$$

In general, this set is not an interval vector. It is a non convex polyhedron. The Oettli and Prager theorem [9] gives the exact solution set.

Nevertheless, this method is quite difficult to use with matrices corresponding to real physical cases in an n-dimensional problem. Most of the time, we will consider the smallest interval vector containing $\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})$, which is defined as $\square\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})$. In this case, this ensures that the true solution is included in the numerical solution found.

2. FORMULATION ADAPTED TO FINITE ELEMENTS METHODS

The existing algorithms used to solve $\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})$ have been formulated for reliable computing from a numerical point of view. In an interval matrix, for instance, each term can vary independently of each other in its interval, which is generally sharp.

If the interval formulation has then to be adapted to mechanics, the dependence between the parameters must be taken into account. Many of the terms of the matrices depend on the same parameters. For example if Young's modulus varies in $\mathbf{E}$, the stiffness matrix can formally be written $\boldsymbol{\alpha}\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$ which is not the same as $\begin{bmatrix} \boldsymbol{\alpha} k_{11} & \boldsymbol{\alpha} k_{12} \\ \boldsymbol{\alpha} k_{21} & \boldsymbol{\alpha} k_{22} \end{bmatrix}$, that is treated as $\begin{bmatrix} \alpha_1 k_{11} & \alpha_2 k_{12} \\ \alpha_3 k_{21} & \alpha_4 k_{22} \end{bmatrix}$, with $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ varying in $\boldsymbol{\alpha}$ independently.

When including the parameters in the terms of the matrices and vectors, the width of $\Sigma_{\exists\exists}([\mathbf{A}], \{\mathbf{b}\})$ grows substantially (see example in 3.3).

Moreover, the stiffness matrices are symmetric positive and definite. If all the matrices $[K] \in [\mathbf{K}]$ are considered, we must notice that many of them do not physically correspond to stiffness matrices.

The problem is formulated as:

$$[A] = [A_0] + \sum_{n=1}^{N} \boldsymbol{\varepsilon}_n [A_n] \qquad \{b\} = \{b_0\} + \sum_{p=1}^{P} \boldsymbol{\beta}_p \{b_p\} \tag{4}$$

N and P are the number of parameters to be taken into account when building the matrix $[A]$ and the vector $\{b\}$. $\boldsymbol{\varepsilon}_n$ and $\boldsymbol{\beta}_n$ are independent centered intervals, generally $[-1,1]$. $[A_0]$ and $\{b_0\}$ correspond to the matrices and vector built from the mean values of the parameters.

This transcription of the equations enables to keep the particular form of the matrices and vectors. For example, considering the following stiffness matrix:

$$[K] = [K_0] + \sum_{n=1}^{N} \boldsymbol{\varepsilon}_n [K_n] \tag{5}$$

for each value of ε_n in $\boldsymbol{\varepsilon}_n$, $[K]$ remains symmetric positive and definite, due to the physical character of this matrix.

A special algorithm is required, because the solution set is not given by a combination of the bounds of the parameters.

3. OBTAINING AN INCLUDING SOLUTION BY MEANS OF A NEW ALGORITHM

3.1 Rump's inclusion

The inclusion method of Rump [6] relies on the fixed point theorem.

The problem to be solved is $[A]\{x\}=\{b\}$, where $[A]$ is a square matrix. For an arbitrary non singular matrix $[R]$, and a vector $\{x_0\}$, $\{x\}=\{x_0\}+\{x^*\}$, where $\{x^*\}$ verifies $\{x^*\}=[G]\{x^*\}+\{g\}$ with $[G]=[I]-[R][A]$, $\{g\}=[R](\{b\}-[A]\{x_0\})$

In practice, $[R]\approx[A]^{-1}$, and $\{x_0\}=[R]\{b\}$, so that $[G]$ and $\{g\}$ are of small norms, and $\{x^*\}$ is close to 0.

In the general case, this algorithm converges rapidly, with a high level of accuracy (see [10]).

Rohn and Rex have shown in [12] that the algorithm converges if and only if $\rho(|[G]|)<1$, where $\rho(|[A]|)$ is the spectral radius of the absolute value of $[A]$. The proof can also be found in a technical report of Rohn [10].

3.2 Adaptation to the mechanical formulation

The proposed algorithm has successfully been used for numerical analysis. In order to apply this algorithm to the mechanical formulation some adaptations are required, due to the specific differences of the mechanical problems highlighted previously.

Let us consider a system in which only one parameter is an interval. The general equation of this system is:

$$\left([A_0]+\alpha[A_1]\right)\{x\}=\{b\} \qquad \alpha\in \boldsymbol{\alpha} \tag{6}$$

and the particular form of the matrix has to be taken into account [3].
The method proposed in [3] is easily generalized to the problems:

$$[\mathbf{A}]\{x\}=\{\mathbf{b}\}\text{, where } [\mathbf{A}]=[A_0]+\sum_{n=1}^{N}\boldsymbol{\alpha}_n[A_n],\ \{\mathbf{b}\}=\{b_0\}+\sum_{p=1}^{P}\boldsymbol{\beta}_p\{b_p\} \tag{7}$$

3.3 Comparison between the general algorithm and the adapted one

The results found with the modified Rump's algorithm are much sharper than the ones found with the classical formulation.

Let us consider for instance a clamped free beam:

Figure -1. Clamped free beam

F and M are respectively the shear force and bending momentum applied at the free end of the beam, d and θ correspond to the displacement and slope at the free end of the beam.

The characteristics of the beam are:
The Young's modulus $E \in [2.058e11, 2.142e11] \quad 2.1e11 \pm 2\%$
The Inertia $I \in [8.82e-8, 9.18e-8] \quad 9e-8 \pm 2\%$
The length l=1

If we consider the elementary Finite Element matrix of the Euler Bernoulli theory [4], the static matrix equation of the problem is given by:

$$\begin{bmatrix} \frac{2EI}{9l} & \frac{-EI}{3l^2} \\ \frac{-EI}{3l^2} & \frac{2EI}{3l^3} \end{bmatrix} \begin{Bmatrix} d \\ \theta \end{Bmatrix} = \begin{Bmatrix} F \\ M \end{Bmatrix} \tag{8}$$

and from a numerical point of view, the stiffness matrix is an interval matrix.

Let us consider the interval vector $\{\mathbf{f}\} = \begin{Bmatrix} [-10.2, -9.8] \\ [29.4, 30.6] \end{Bmatrix}$.

The first problem that can be solved is finding the solution set corresponding to the numerical equation:

$$\begin{bmatrix} [4033.68,4369.68] & [-6554.52,-6050.52] \\ [-6554.52,-6050.52] & [12101.40,13109.04] \end{bmatrix} \begin{Bmatrix} d \\ \theta \end{Bmatrix} = \begin{Bmatrix} [-10.2,-9.8] \\ [29.4,30.6] \end{Bmatrix} \quad (9)$$

The Oettli and Prager lemma gives the exact solution set $\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})$ and $\square\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})$ (dotted line) shown in Figure 2. All the terms in the matrix are said to be independent.

On the other hand, we can consider the mechanical problem with factorized interval parameters:

$$\mathbf{EI}\begin{bmatrix} \dfrac{2}{9l} & \dfrac{-1}{3l^2} \\ \dfrac{-1}{3l^2} & \dfrac{2}{3l^3} \end{bmatrix} \begin{Bmatrix} d \\ \theta \end{Bmatrix} = \begin{Bmatrix} \mathbf{F} \\ \mathbf{M} \end{Bmatrix} \quad (10)$$

As this system is quite simple, the analytical solution can be found analytically. The exact mechanical solution set is given in Figure 2. It is called a mechanical exact solution set. The hull of this set (which is an interval vector) has also been drawn. The mechanical exact solution set is included in $\Sigma_{\exists\exists}([\mathbf{A}],\{\mathbf{b}\})$, and is really small in comparison. This shows how important factorization is in solving mechanical problems.

To test our algorithm, we have computed the result of the modified Rump's algorithm. It is illustrated in Figure 2. As we can see, it covers the exact solution, but it gives a very good idea of the size of the solution, without large overestimation.

As had been noticed in [2], a large overestimation is obtained when including the parameters in the elements of the matrices. For finite element matrices, this overestimation can become critical, and often leads to an insolvable problem. As we have shown above, even on 2x2 matrices, the overestimation can reach a factor of 10 or more. Such an adaptation of this algorithm enables its use in industrial problems involving huge matrices.

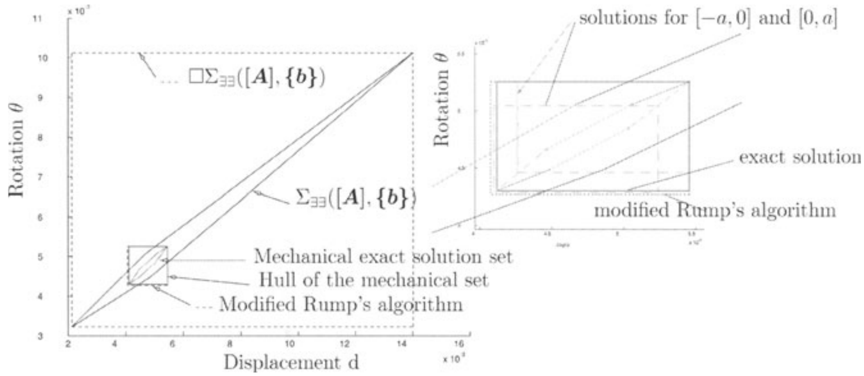


Figure -2. Solution sets for the clamped free beam. EI is uncertain (±2%). Numerical global problem, and reduced mechanical problem, and their respective hulls.

4. FRAME STRUCTURE

We can consider some kind of realistic structure. It is a two dimensional frame structure (Figure 3). It can model for instance the structure of a building, as in [5].

It is a 18 element structure, with 12 degrees of freedom. Only a concentrated load on the beams is considered, applied on the DOFF 3, the torque $F = 10^3\,\mathrm{Nm}$. The parameters of the model are the lengths of the beams $L_1 = L_2 = 1\mathrm{m}$, the inertia $I = \pi 10^{-8} / 4\mathrm{m}^4$ and their area $S = \pi 10^{-4}\,\mathrm{m}^2$. We assume that the bending rigidity E is uncertain ($E = 210 \pm 10\%$ GPa$).

We will study the two-dimensional frame structure from a dynamical point of view. In the Finite Element Model, we use a Euler Bernoulli Beam model, leading to a stiffness and a mass matrix, calculated with the numerical values given at the beginning of paragraph. We suppose that there is hysteretic damping ($\eta = 2\%$) in the model. Because all of the beams are identical, some of the modes are relatively close to each other. This means that when the bending rigidity varies, the eigenmodes can overlap. This behavior is illustrated in Figure 4 , where a harmonic torque is applied on node 3 of the truss. The collocated transfer function $H(3,3)$ is computed thanks to the proposed algorithm, and compared to deterministic transfer functions calculated for various values of the Young's modulus.

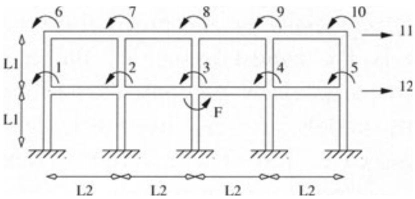


Figure -3. Two dimensional frame structure with 12 DOFF. A torque F is applied on node 3.

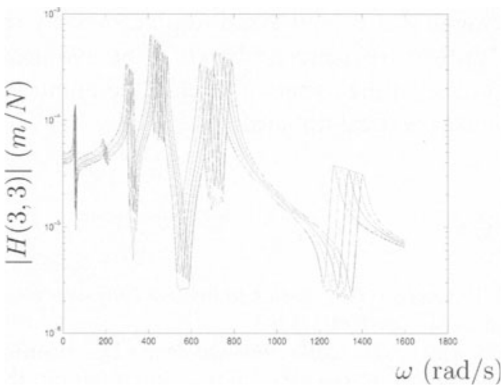


Figure -4. Modulus of the collocated transfer function $H(3,3)$ for the frame. The bending rigidity is uncertain ($E = 210 \pm 10\%$ GPa). The min and max values calculated with the modified algorithm are represented, wrapping the transfer function for several values of E.

As we can see in Figure 4, the proposed algorithm can take overlapping eigenfrequencies into account. It leads to an envelope of the modulus of the transfer function. Some of the deterministic transfer functions have been plotted to show that the envelope found does not overestimate the real solution too much.

5. CONCLUSION

The algorithm developed in this paper for taking into consideration uncertain parameters in mechanical models, and particularly the Finite Elements Models, has been applied to typical examples of vibrating systems. For a discrete system in which several parameters are uncertain, the algorithm gives some interesting results. It leads to an envelope both for the static response and for the dynamic transfer function. The relevance of such an envelope is that one can be certain that all of the solutions corresponding to each parameter varying between the pre definite bounds is definitely in this envelope. This is the robust aspect of the method. Moreover, the numerical cost is low, especially compared to Monte Carlo simulations. Large Finite Elements models have also been tested, and the algorithm gives good results in these cases too. The solution is accurate, wrapping the solution set without overestimating it too much. The necessary amount of computation remains reasonable, and it can be reduced drastically if the required accuracy is lower.

The algorithm is able to handle vibrating systems with natural frequencies close together. This behavior leads to an overlap phenomenon on the transfer functions, that is well taken into account by the algorithm. This enables to find "critical frequencies" bands to be avoided for example, and gives a superior bound of the dynamic level of the output, which in turn is of benefit when calculating ideal dimensions.

REFERENCES

[1] G. Alefeld and J. Herzberger. *Introduction to Interval Computations*. Academic Press, Inc, 111 Fifth Avenue, New York 10003, 1983.

[2] R. Chen and A. C. Ward. *Generalizing Interval Matrix Operations for Design*. Journal of Mechanical Design, 119:65-72, mar 1997.

[3] O. Dessombz, F. Thouverez, J.P. Laîné and L. Jézéquel. *Analysis of Mechanical Systems using Interval Computations applied to Finite Elements Methods*. Journal of Sound and Vibration, 239(5): 949-968, feb 2001.

[4] J.-F. Imbert. *Analyse des structures par éléments finis*. Cépaduès Editions, 1995",

[5] H.U. Koyluoglu, A.S. Cakmak and S.R.K. Nielsen. *Interval Algebra to Deal with Pattern Loading and Structural Uncertainties*. Journal of Engineering Mechanics, 1149-1157, nov 1995.

[6] U. Kulisch and W. Miranker. *A New Approach to Scientific computation, chapter Solving algebraic problems with high accuracy* (S.M. Rump), pages 51-120. Academic Press, New York, 1983.

[7 R.E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

[8] S. Ning and R. B. Kearfott. *A comparison of some methods for solving linear interval equations*. SIAM Journal on Numerical .Analysis., 34(4): 1289-1305, aug 1997.

[9] W. Oettli and W. Prager. *Compatibility of Approximate Solution of Linear Equations with Given Error Bounds for Coefficients and Right-Hand Sides*. Numerische Mathematik, 6:405-409, 1964.

[10] J. Rohn. Technical report 620. *Validated Solutions of Linear Equations*. Institute of Computer Science, Academy of Sciences of the Czech Republic, jan 1995.

[11] J. Rohn. . Technical report 686. *Checking Properties of interval Matrices*. Institute of Computer Science, Academy of Sciences of the Czech Republic sep 1996.

[12] J. Rohn and G. Rex. Technical report 666. *Enclosing Solutions of Linear Equations*. Institute of Computer Science, Academy of Sciences of the Czech Republic, feb 1996.

[13] Sergey P. Shary. *On optimal solution of interval linear equations*. SIAM Journal on Numerical .Analysis, 32(2):610-630, apr 1995.

INTEGRATION OF RELIABILITY-BASED DESIGN OPTIMIZATION WITHIN CAD AND FINITE ELEMENT MODELS

G. Kharmanda, A. Mohamed and M. Lemaire
LaRAMA, IFMA/UBP, Clermont-Ferrand, France

Abstract: High technology designs aim to define the best compromise between cost and safety. The Reliability-Based Design Optimization (RBDO) allows us to design structures which satisfy economical and safety requirements. However, in practical applications, the coupling between the geometrical modeling, the mechanical simulation, the reliability analyses and the optimization methods leads to very high computing time and weak convergence stability. Traditionally, the solution of the RBDO model is achieved by alternating reliability and optimization iterations (sequential approach). This approach leads to low numerical efficiency, especially when the structure geometry is described by efficient CAD models (such as B-spline or NURBS) which necessitate a great amount of data. In this case, the integration of the sequential RBDO approach with Finite Element Analysis requires lengthy computing time. In order to avoid this difficulty, we propose an efficient method, called the hybrid RBDO method, based on the simultaneous solution of reliability and optimization problems. In this paper, the efficiency of the proposed methodology is demonstrated on a finite element problem using CAD model description.

Key words: reliability-based design optimization, CAD, finite element analysis

1. INTRODUCTION

In deterministic structural optimization, the designer aims to reduce the construction cost without caring about the effects of uncertainties concerning materials, geometry and loading. In this way, the resulting optimal

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 453–462.

configuration may present a lower reliability level and lead to higher failure rates. The equilibrium between cost minimization and reliability maximization is a great challenge for the designer. The purpose of Reliability-Based Design Optimization (RBDO) is to design structures that should be economic and reliable, by introducing the safety criteria in the optimization procedure.

In the RBDO model, we distinguish between two kinds of variables:

1) The design variables $\vec{x}$ which are deterministic variables to be defined in order to optimize the mechanical system.

2) The random variables $\vec{y}$ which represent the structural uncertainties, identified by probabilistic distributions.

The solution of coupled optimization and reliability problems requires very large calculation resources that seriously reduce the applicability of the classical approach [1]. Thus, there is a strong motivation to develop efficient techniques with the aim of reducing the computational time. In this paper, we propose a new approach which is based on the simultaneous solution of the reliability and optimization problems by using a new Hybrid Design Space (HDS).

2. STRUCTURAL RELIABILITY

The design of structures and the prediction of their correct operation come from the verification of a certain number of rules resulting from the knowledge of physics and mechanical experience of designers and constructors. These rules traduce the necessity to limit the loading effects such as stresses and displacements. Each rule represents an elementary event and the occurrence of several events leads to a failure scenario. In addition to the vector of deterministic variables $\vec{x}$ to be used in the system control and optimization, the uncertainties are modeled by a vector of stochastic physical variables affecting the failure scenario. The knowledge of these variables is not, at best, more than statistical information and we admit a representation in the form of random variables. For a given design rule, the basic random variables are defined by their probability distribution associated with some expected parameters; the vector of random variables is noted herein $\vec{Y}$ whose realizations are written $\vec{y}$.

Safety is the state where the structure is able to fulfil all the functioning requirements: mechanical and serviceability, for which it is designed. To evaluate the failure probability with respect to a chosen failure scenario, a performance function $G(\vec{x}, \vec{y})$ is defined by the condition of good operation of the structure. The limit between the state of failure $G(\vec{x}, \vec{y}) \leq 0$ and the state of safety $G(\vec{x},\vec{y})>0$ is known as the limit state surface $G(\vec{x},\vec{y})=0$ (fig. 1). The failure probability is then calculated by:

$$P_f = \Pr\left[G(\vec{x},\vec{y}) \leq 0\right] = \int_{G(\vec{x},\vec{y})\leq 0} f_{\vec{Y}}(\vec{y})\, dy_1 \cdots dy_n \tag{1}$$

where P_f is the failure probability, $f_{\vec{Y}}(\vec{y})$ is the joint density function of the random variables $\vec{Y}$ and Pr[.] is the probability operator.

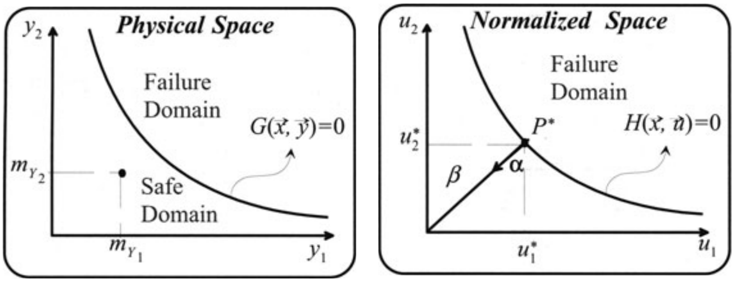


Figure -1. Physical and normalized spaces

The evaluation of integral (1) is carried out by First and Second Order Reliability Methods *FORM/SORM* [2]. They are based on the reliability index concept, followed by an estimation of the failure probability. The invariant reliability index β was introduced by Hasofer and Lind [3], who proposed to work in the space of standard independent Gaussian variables instead of the space of physical variables. The transformation from the physical variables $\vec{y}$ to the normalized variables $\vec{u}$ is given by:

$$\vec{u} = T(\vec{x},\vec{y}) \quad \text{and} \quad \vec{y} = T^{-1}(\vec{x},\vec{u}) \tag{2}$$

This transformation $T(.)$ is called the probabilistic transformation. In this standard space, the limit state function takes the form:

$$H(\vec{x},\vec{u}) \equiv G(\vec{x},\vec{y}) = 0 \tag{3}$$

In the *FORM* approximation, the failure probability is simply evaluated by:

$$P_f \approx \Phi(-\beta) \tag{4}$$

where $\Phi(.)$ is the standard Gaussian cumulated function. For practical engineering *FORM* gives a sufficiently accurate estimation of the failure probability. The reliability index β is evaluated by solving the constrained optimization problem:

$$\beta = \min d(\vec{u}) = \sqrt{\vec{u}^T \cdot \vec{u}} \qquad \text{under the constraint: } G(\vec{x}, \vec{y}) \leq 0 \qquad (5)$$

where $d(\vec{u})$ is the vector modulus in the normalized space, measured from the origin. The solution of the problem is called the *design point*: $\vec{u} = -\beta\vec{\alpha}$ with $\vec{\alpha}$ the unit normal vector to the limit state at the design point $\vec{u}$, as illustrated in figure (1).

3. RBDO MODELS

3.1 Classical approach

The sequential RBDO procedure is solved in two spaces: the space of design variables, known as the *physical space* and the space of Gaussian random variables, known as the *normalized space*. The RBDO is calculated by nesting the two following problems:

1) optimization under deterministic and reliability constraints:

$$\min: \ f(\vec{x}) \qquad \text{subject to: } g_k(\vec{x}) \leq 0 \ \text{ and: } \ \beta(\vec{x}, \vec{u}) \geq \beta_t \qquad (6)$$

where $f(\vec{x})$ is the objective function, $g_k(\vec{x})=0$ are the associated deterministic constraints, $\beta(\vec{x}, \vec{u})$ is the reliability index of the structure and β_t is the target reliability.

2) calculation of the reliability index:

$$\min: \ d(\vec{u}) \qquad \text{subject to: } H(\vec{x}, \vec{u}) \leq 0 \qquad (7)$$

where $d(\vec{u})$ is the distance in the normalized space and $H(\vec{x}, \vec{u})$ is the limit state function as shown in section 2.

3.2 Hybrid RBDO

In order to avoid the large computational time of the nested problems given in section 3.1, we propose a new formulation by combining deterministic and random spaces. The new form of the objective function $F(\vec{x}, \vec{y})$ integrates cost and reliability aspects:

$$F(\vec{x},\vec{y}) = f(\vec{x}) \cdot d_\beta(\vec{x},\vec{y}) \tag{8}$$

where $d_\beta(\vec{x}, \vec{y})$ is the image of $d(\vec{u})$ in the physical space; this means that: $d_\beta(\vec{x}, \vec{y})=d(T(\vec{x}, \vec{y}))$. The new problem is now formulated as:

$$\begin{array}{lll} \text{min} & : & F(\vec{x},\vec{y}) \\ \text{subject to} & : & g_k(\vec{x}) \leq 0 \;;\; G(\vec{x},\vec{y}) \leq 0 \quad \text{and}: \; \mathrm{d}_\beta(\vec{x},\vec{y}) \geq \beta_t \end{array} \tag{9}$$

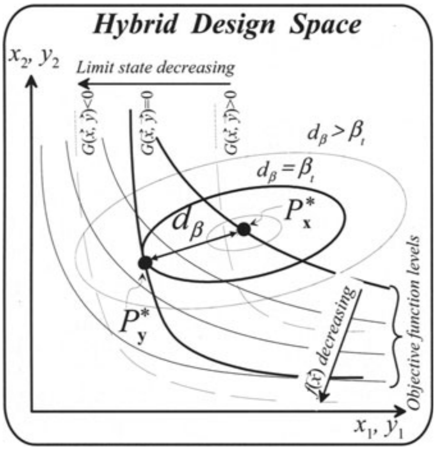


Figure -2. Hybrid Design space

The minimization of the function $F(\vec{x}, \vec{y})$ is carried out in the hybrid space of deterministic variables $\vec{x}$ and random variables $\vec{y}$. An example of this hybrid design space (HDS) is given in figure (2), containing design and random variables, where the reliability levels d_β are represented by ellipses (case of normal distribution), the objective function levels are given by solid curves and the limit state function is represented by dashed lines. We can see two important points: the optimal solution P_x^* and the reliability solution P_y^* (i.e. the design point found on the curves $G(\vec{x}, \vec{y})=0$ and $d_\beta=\beta_t$). All the information about the RBDO problem can be found in this space (e.g. optimal points, sensitivities, reliability levels, objective function iso-values and constraints). The efficiency of the proposed approach with respect to the classical one has been tested on several examples [4,5].

4. RBDO MODELS WITH FEA

In this section, we show how the hybrid RBDO model becomes an efficient tool when the mechanical model is represented by Finite Element Analysis (FEA). After the discussion of sensitivity equations in FEA, the hybrid RBDO is extended to nonlinear problems in order to demonstrate the efficiency of the hybrid methodology.

4.1 Sensitivity operators

Let us consider the case of RBDO using a finite element model based on a geometrical and material linear elastic displacement method. For a given failure scenario, the limit state function is written as:

$$H(\vec{x},\vec{u},\vec{b}(\vec{x},\vec{u},\vec{q}(\vec{x},\vec{u}))) = 0 \tag{10}$$

where $\vec{q}$ is the nodal displacement vector and $\vec{b}$ is a vector of response parameters associated with the limit state function, e.g. internal forces, stresses, strains or displacements. In the optimization algorithms for the design point search, the gradients of *H*(.) with respect to $\vec{u}$ are needed. The efficiency of the use of sensitivity operators in reliability analysis has been shown in previous works [6].

In the RBDO problem with linear elastic analysis, it is seen that, at the sub-iteration level, the calculation of the limit state function and its gradient requires only one solution of the finite element equilibrium equations for each sub-level (i.e. for each $\vec{x}$), as long as the stiffness matrix is independent of $\vec{u}$. The main advantage of estimating the sensitivities of β instead of a simple numerical finite difference scheme is that a very large number of β calculations and stiffness assemblies and inversions can be avoided, thus reducing considerably the computational time.

4.2 Efficiency in non-linear analysis

In the cases where material or geometrical non-linearities in the finite element model are involved, it is also possible to perform the RBDO but the computational time will increase significantly because the iterations must be performed at 3 levels:

1) Deterministic optimization in the design space $\vec{x}$,
2) Reliability analysis in the normalized space $\vec{u}$,
3) Nonlinear equilibrium iterations in the nodal displacement space $\vec{q}$.

For nonlinear analysis, the hybrid RBDO is very efficient because the number of derivatives is largely reduced and a lot of non-linear iterations are avoided, when compared to sequential approach.

5. CAD MODELS

The parameterization step allows us to define the search directions of the optimization process. In the shape optimization, these parameters are chosen among the design variables that define the geometry at the domain boundary. In fact, the shape optimization process is driven by the information corresponding to the geometrical boundary perturbation. The structural geometry that will be modified during the optimization process, can be described by several methods such as element list (circular arcs and straight segments), Bézier, B-spline or NURBS descriptions.

The element list technique is very simple to implement; the design variables such as arc radius and center, angles or coordinates of straight segment ends can be chosen as the optimization parameters. The boundary is described by the assembly of the elements in the list. The perturbation of the boundary design variables does not imply the change of all boundaries. But the discontinuity in the intersection of the different elements constitutes a major problem for the optimization procedure. Because of these discontinuities, the geometric irregularities of the boundary highly influence the evaluation of certain variable fields defined at the boundary. These irregularities represent a serious disadvantage for the functional minimization, as artificial singularity is created in the model. Furthermore, the use of the element list such as straight segments, circular arcs, parabolic curves represented by mathematical equations, does not ensure the free change of topology during the optimization process.

Therefore, it is necessary to describe the structural geometry by using flexible curves and surfaces. When using Bézier curves, two ways exist: in the first one, high-degree curves are used; in the second one, Bézier curves of modest order are pieced together using simple geometric rules to ensure continuity at the different joints. For instance, to achieve zero-order continuity at a joint, it is sufficient to make the end control points of the curve to coincide. First-order continuity can be obtained by stating that the edges of the two polygons adjacent to the common end point must lie on a straight line. But, the Bézier curves do not provide local control: moving any control point will change the shape of every part of the curve. However, the B-splines provide on the one hand that local control of the curve shape can be achieved and on the other hand, that additional control points can be introduced without increasing the degree of the curve. B-splines offer more parameters to the designer than Bézier curves: the degree can be selected, as well as the multiplicity of the control points. The B-spline parametric curves representation is a very attractive tool for shape optimization by the design element technique. By using the possibility of automatically piecing together splines of modest order to define a B-spline, and by controlling the continuity, we are able to describe any complex geometry. A small number of design elements are generally sufficient to fully describe the region that is modified during optimization. Moreover, selecting the degree according to the variation diminishing property provides a rational scheme to avoid unrealistic designs [7].

6. PRACTICAL APPLICATIONS

The following applications illustrate the interest of the RBDO model with respect to the deterministic design optimization model. The Deterministic Design Optimization (DDO) problem is to minimize the structure volume subject to the mechanical stress constraint, however, the RBDO model aims to minimize the structural volume subject to the target reliability β_t=3.8.

6.1 Anchor bracket

A bracket is clamped at its two lower ends to a rigid foundation, and acted upon by a concentrated load transmitted through a rigid axle. All the boundaries are allowed to be modified, except the internal circular hole (the radius of the axle is fixed). The external boundary is presented by segments and B-Spline curves. The coordinates of the control points (x,y) representing the curves' polygons are considered as optimization variables (random and design variables). Uncertainties are introduced by geometrical tolerances and applied load fluctuations.

Table 1 shows that the reliability level of the deterministic point is very low with respect to the target reliability level. However, the RBDO solution respects the required reliability level of 3.8 but with a small increase of volume. Fig. 3 shows the final shapes of the two solutions.

Table -1. Optimal results of the bracket

Model	DDO	RBDO
$V_{optimal}$ (mm^3)	93710	95760
β	1.41	3.82

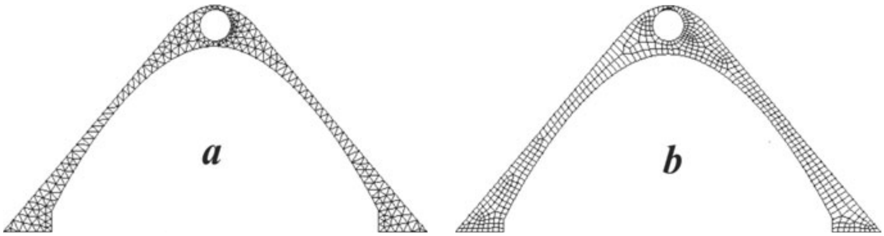


Figure -3. a-Optimal solution using DDO, b-Optimal solution using RBDO

6.2 Supporting plate

The supporting plate shown in Fig. 4 is clamped at the four external holes *1*, *2*, *3* and *4* and acted upon by a concentrated load transmitted through the central hole.

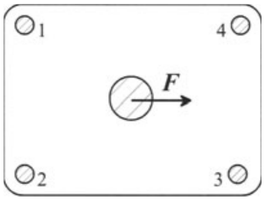


Figure -4. Design model of the supporting plate

The optimal volumes are found to be 148000 mm³ and 168000 mm³ for DDO and RBDO respectively. This increase of volume is accompanied by improving the reliability level from 1.74 to 3.83. Fig. 5 shows also the final shapes of the two solutions, we can see that the DDO solution leads to very thin supporting arms. The results indicate that the reliability at the deterministic optimum is quite low and needs to be improved by considering the RBDO which reduces the structural weight in uncritical regions.

7. CONCLUSIONS

In classical RBDO, many repeated searches are needed in the physical and normalized spaces that leads to very large computational time and to a complex coupling between geometric modeling, mechanical modeling, reliability analysis and optimization methods. Because of these two difficulties, few researchers integrate the RBDO model within the design process, especially for large-scale problems. The proposed method allows this coupling because it consists in solving the RBDO problem in a hybrid space containing random and deterministic variables and it leads to parallel convergence for both problems. Using this approach, reliability-based design optimization becomes a practical tool for designers.

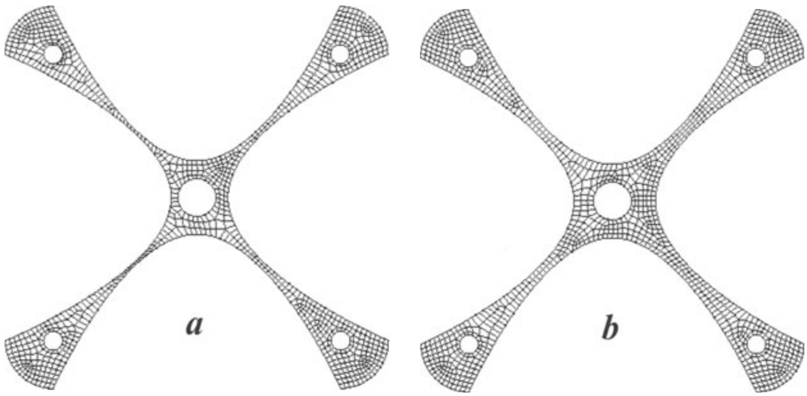


Figure -5. a-Optimal solution using DDO, b-Optimal solution using RBDO

REFERENCES

1. G. Gheng, G. Li, Y. Cai. *"Reliability-based structural optimization under hazard loads"*, Structural Optimization 16(2-3), 1998, pp. 128-135.
2. O. Ditlevsen, H. Madsen. *"Structural Reliability Methods"*, John Wiley & Sons, 1996.
3. A.M. Hasofer, N.C. Lind. *"An Exact and Invariant First Order Reliability Format"*, J. Eng. Mech., ASCE, EM1 100, 1974, pp. 111-121.
4. G. Kharmanda, A. Mohamed, M. Lemaire. *"New hybrid formulation for reliability-based optimization of structures"*, The Fourth World Congress of Structural and Multidisciplinary Optimization (WCSMO-4), Dalian, China, June 2001.
5. G. Kharmanda, A. Mohamed, M. Lemaire. *"Efficient reliability-based design optimization using a hybrid space with application to finite element analysis"*, accepted for publication in Structural Optimization, 2002.
6. A. Mohamed, M. Lemaire. *"The use of sensitivity operators in the reliability analysis of structures"*, Computational Stochastic Mechanics, Spanos, P.D., Balkema, Greece, 1998, pp. 211-220.
7. L. Piegl, W. Tiller. *"The NURBS Book"*, Springer-Verlag, USA, 1995.

PLANNING COMPLEX ENGINEER-TO-ORDER PRODUCTS

C. Earl, D. Ping Song and C. Hicks
Open University, UK
Imperial College London, UK
Newcastle University, UK

Abstract: The design and manufacture of complex Engineer-to-Order products is characterised by uncertain operation durations, finite capacity resources and multilevel product structures. Two scheduling methods are presented to minimise expected costs for multiple products across multiple finite capacity resources. The first sub-optimises the operations sequence, using mean operation durations, then refines the schedule by perturbation. The second method generates a schedule of start times directly by random search with an embedded simulation of candidate schedules for evaluation. The methods are compared for industrial examples.

Key words: engineer-to-order, schedule, stochastic, resource constraints

1. INTRODUCTION

Planning Engineer-to-Order products [1,2,3,4] takes into account several factors: significant uncertainty in operation times; limited resources maintained by companies in response to fluctuations in demand; concurrent development of multiple products and complex product (Figure 1). The problem addressed here is to create a plan with minimum expected costs (both earliness and lateness). The majority of planning research in this area is limited to simple systems such as single machines [5], serial structures or flow lines [6,7,8,9], two stage assembly systems and two stage distribution type networks [10,11]. Research on planning multilevel products has optimised operation sequences for deterministic operation durations. However, with stochastic durations, the other factors, namely; finite capacity resource constraints, precedence constraints, assembly co-ordination and due

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 463–472.

date constraints; all cause the actual schedule to deviate significantly from plan. A further difficulty is that although the schedule starts as a useful guide, it can rapidly become a constraint as uncertainties take effect, operations are completed and new conditions for the remaining operations come into play. Two methods are developed to optimise schedules for multiple complex products over multiple finite resources. The first optimises sequences (using mean operation duration estimates) then further optimises timings for this sequence. The second optimises timings directly. The methods are tested on examples of ETO products in the power generation industry.

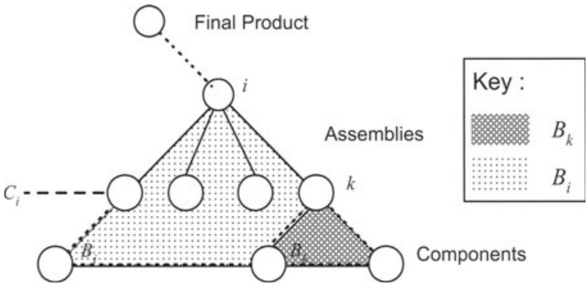


Figure -1. General product structure whose nodes represent manufacture/assembly (with component design at leaves and final assembly at the root)

2. PLANNING

Planning in stochastic systems needs to take account of how plans are implemented. For example, when a resource finishes an operation but the next planned operation has not arrived (but several other operations are queuing at this resource); which operation should the resource do next? Two choices present themselves. The first keeps the original sequence [12]. The resource ignores the queuing operations, keeps idle and waits for the next planned operation. The second choice selects one of the queuing operations by a priority rule. These choices give rise to the two methods developed here for planning. Broadly the first optimises sequences then refines the associated timings. The second optimises timings directly.

2.1 Notation

An implementation is a sample process (or real execution) of a schedule of start times $\boldsymbol{s} = \{s_i\}$ of operations i, with durations x_i sampled from their probability distributions (or given by their real values). An implementation is described by the 'actual' operation start times a_i and completion times c_i of operations on many products and resources. The following notation is used:

- Γ set of all operations over all products and all resources,
- L set of the final assembly operations over all products,
- C_i operations which immediately precede operation i (figure 1),
- $\rho(i)$ operation immediately after i in the product structure,
- $r(i)$ resource used to undertake operation i
- $\varphi(i)$ operation immediately preceding operation i on resource $r(i)$;
- d_i due date of operation i.
- h_i, h_i^- costs of earliness and lateness for operation i.

The basic relations among starting and completion times are:

$$a_i = \max\,(s_i,\, c_{\varphi(i)},\, \{c_j : j \in C_i\}) \qquad c_i = a_i + x_i \tag{1}$$

The cost function which is minimised is the expected value of:

$$\textstyle\sum_{i\in\Gamma\setminus L} h_i(a_{\rho(i)} - c_i) + \sum_{i\in L} h_i \max(d_i - c_i, 0) + \sum_{i\in L} h_i^- \max(c_i - d_i, 0) \tag{2}$$

2.2 Industrial ETO examples

The planning methods were tested on examples from an Engineer-to-Order company which designs and manufactures power generation equipment. Although the examples are not complete products they are significant functional subassemblies such as bearing pedestals, casings and rotor/blade assemblies. The data available covered both product structures and estimates of operation times for manufacturing processes and assembly. As mentioned above these times are not known with certainty. But the difficulties are more profound. Because the products are engineer-to-order, operations are not repeated across products and so estimations of stochastic characteristics and distributions are problematic due to small sample sizes. However, from historical data on sets of similar operations, means and variances can be approximated but estimates of distributions are more difficult to obtain. In this paper we assume that the distributions are normal.

3. TWO-PHASE OPTIMISATION METHOD

The first phase (figure 2) sets the values of operation durations at the estimated means. A sequence of operations on each resource is then determined by either: (i) a finite loading heuristic with a priority rule or, (ii) random search using Simulated Annealing (SA) or Evolution Strategy (ES). The second phase takes operation sequences from the first phase as constraints and refines operation timings using a Perturbation Analysis Stochastic Approximation (PASA). Although SA and ES can also be used to refine the timings in the second phase, PASA was found to converge faster.

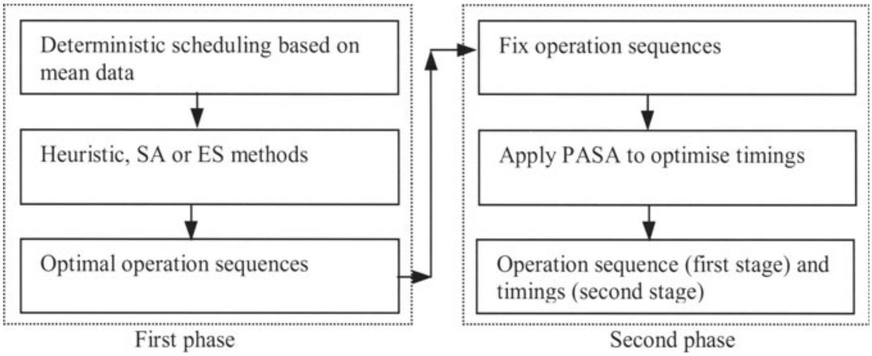


Figure -2. Two-phase scheduling

3.1 First phase heuristics

First phase heuristics are based on backwards scheduling which aims to start each item as late as possible. If more than one operation is waiting to be loaded onto a resource then a latest possible completion time (LCT) priority rule is used to decide which is loaded first. Note that if an operation is loaded first in the backwards scheduling then it will be processed last when the plan is implemented. The LCT priority rule tries to reduce waiting and thus holding costs. In cases where LCT cannot decide, the highest cost ratio h_i/x_i (HCR) selects the next operation to be loaded.

Backwards finite loading with the LCT priority rule has computational complexity less than $O(n^2)$ in the number of operations n. Several examples of ETO products, (major subassemblies), from the power generation sector were tested. One with 113 operations and 13 resources took 2 seconds and another with 239 operations on three concurrent products and 17 resources took 4 seconds. See table 1 and figure 4 for details of these products. Finite loading gives much lower costs (up to a factor of five) than infinite loading.

3.2 First phase random search

Random search methods are used to optimise start times of operations under the assumption that durations are set at their estimated means. These start times give a sequence which is then adopted in the second phase refinement of the schedule. Two different random search techniques have been applied. Simulated Annealing (SA) [13,14] numerically optimises [15] operation start times $\{s_i\}$. The performance (total cost) of a candidate schedule within a SA iteration is determined by simulating its implementation using the priority rule that the operation with the earliest planned start time (EPST) is started first if there is competition for resources. Two types of constraint are applied. Physical constraints specify that a starting event on an operation cannot occur before all immediately preceding operations C_i are completed. Planning constraints specify that an operation cannot start before its planned start time ($a_i \geq s_i$). The actual schedule $\{a_i\}$ given by the simulation is likely to deviate significantly from $\{s_i\}$ because of the finite capacity resources. Neighbouring schedules are obtained by randomly choosing new start times from a uniform distribution in the ranges $[s_i - 1/2\gamma_i, s_i + 1/2\gamma_i]$ where γ_i is the current step size. Step sizes and annealing temperatures are reduced linearly.

Evolution Strategy (ES) has iterative procedures with "selection", "crossover" and "mutation" for generating offspring from a parent population. Offspring are selected for further generation. ES uses continuous variables and is thus suitable for numerical optimisation [16] of schedules. The chromosome of an individual schedule is represented by the vector $\mathbf{s} = \{s_i\}$ of start times. Evaluation of offspring is by simulation as in SA.

Results from running SA and ES on industrial data for a single multilevel subassembly with 113 operations and 13 separate resources indicate that SA converges faster than ES, although in the long term ES gives lower costs. See table 1 and figure 4(a) for details of this product and associated operations. Compared with the heuristic method ES and SA achieve marginally lower costs, in the region of 10%, but higher computational costs (by a factor of 100). Similar results were obtained from experiments on similar subassemblies from the power generation sector.

3.3 Second phase perturbation analysis

Perturbation analysis (PA) is used in the second phase. Consider a sample implementation (or nominal path NP) of a schedule with start times $\mathbf{s} = \{s_i\}$ and parameters (eg durations) ω, in the whole sample space Ω, determining start times $\{a_i\}$ and completion times $\{c_i\}$ by equations (1). The cost of the sample process $V(\mathbf{s}, \omega)$ is given by (2) [17,7]. Let planned start time s_j be perturbed to $s_j + \Delta$. The sample process for $\{s_j + \Delta, s_i, i \neq j, i \in \Gamma\}$ with the same ω is a perturbed path (PP). The perturbed path with start and completion times $\{a_i'\}$ and $\{c_i'\}$ can be constructed directly from the nominal path without repeating the simulation.

Perturbation generation is described by (i) if $a_j > s_j$, then $a_i' \equiv a_i$ and $c_i' \equiv c_i$, for $i \in \Gamma$ (ii) if $a_j = s_j$, then $a_j' = a_j + \Delta$ and $c_j' = c_j + \Delta$. Perturbation propagation and disappearance are described by (iii) if $a_i = s_i$ $(i \neq j)$, then $a_i' = a_i$ and $c_i' = c_i$ (iv) if $a_i = c_{\varphi(i)}$ $(i \neq j)$, then $a_i' = a_i + (c_{\varphi(i)}' - c_{\varphi(i)})$ and $c_i' = c_i + (c_{\varphi(i)}' - c_{\varphi(i)})$, (v) if $a_i = c_k$ $(i \neq j, k \in C_i)$, then $a_i' = a_i + (c_k' - c_k)$ and $c_i' = c_i + (c_k' - c_k)$. The whole perturbation gain Δ will be propagated along the perturbed path. Define $I(i) := 1\{a_i' \neq a_i\}$, where $1\{.\}$ takes 1 if $\{.\}$ is true, and 0 otherwise. Note that a_i' and c_i' have the same perturbation gain. The sequence $\{I(i), i \in \Gamma\}$ determines

the difference between PP and NP. A recursive procedure implements the perturbation propagation rules and determines $\{I(i),\ i\in\Gamma\}$. Song et al [12] show that for any $\omega\in\Omega$ and $j\in\Gamma$
$\partial V(\mathbf{s},\ \omega)/\partial s_j = \sum_{i\in\Gamma\setminus L} h_i\cdot(I(\rho(i))-I(i))+\sum_{i\in L} h_i\cdot I(i)\cdot 1\{d_i>c_i\}+\sum_{i\in L} h_i^{-}\cdot I(i)\cdot 1\{d_i\leq c_i\}$
is an unbiased estimator of gradient, that is:
$\partial\{\mathrm{E}\ V(\mathbf{s},\ \omega)\ \}/\partial s_j = \mathrm{E}\{\partial V(\mathbf{s},\ \omega)/\partial s_j\}$and $\mathrm{E}|\partial V(\mathbf{s},\ \omega)/\partial s_j| < \infty$, for any $j\in\Gamma$. Thus a PA-based Stochastic Approximation [18,19] is $\mathbf{s}_{n+1} = \mathbf{s}_n - \gamma_n \cdot\nabla J_n$, where $\mathbf{s}_n$ are start times at the beginning of iteration n, ∇J_n is the gradient estimator, and step $\gamma_n>0$, $\gamma_n\to 0$, $\Sigma\ \{\gamma_n\}$diverges and $\Sigma\ \{\gamma_n\}^2$ converges (eg $\gamma_n = 1/n$). The gradient estimator is calculated using K sample processes.

Some results from applying PASA (in conjunction with first phase methods and using K = 100 sample processes) to complex products (figure 4) are shown in figure 5. They indicate that application of the second phase PASA yields an extra reduction in costs of between 10-20% with moderate computation.

4. ONE-PHASE SCHEDULING

Simulated Annealing (SA) and Evolution Strategy (ES) methods can be extended from their application in the two-phase process to provide one-phase optimisation methods (figure 3). Multiple sample processes are required to estimate expected cost.

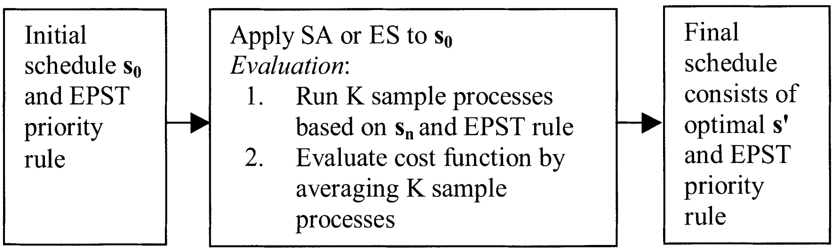


Figure -3. One phase method for scheduling

5. NUMERICAL EXAMPLES

Two examples are described here (Table 1). Example A consists of a single product and example B has three products designed and manufactured concurrently. The products are major subassemblies of power generation plant. One simplification is introduced, namely that the concurrent products are all assumed to start at the same time, with the same delivery date. In practice, for major subassemblies on a single product this is a reasonable assumption but for concurrent products, staged start and delivery will be usual practice. Example B has more pressure on resources - a common feature of ETO where resource constraints become a significant problem as the number of concurrent products increases.

Table -1. Examples of ETO used for scheduling

Example	Products	Components	Operations Manufacture/Assembly	Resources
A	1	9	100/13	13
B	3	47	210/29	17

5.1 Single product

Consider a single product (figure 4a) which is a major subassembly of power generation plant. The numbers on the nodes are references to particular operations, components and products. Normally distributed operation times have increasing variance as assembly progresses. Holding costs are set at 1% x {sum of operation times (in days) already spent on the item}x £1000. Lateness costs for final product were set at twice holding costs of the final product. The due date for this single subassembly was 180 days. Parameters for the search algorithms were established by experiment over several simulations of a range of examples, including the specific ones described here. For each method K=100 sample processes are used to assess stochastic effects. The three two-phase methods are compared in figure 5a, where the second phase PASA is compared for the various inputs provided by the different first phase methods. For single products an extra cost reduction of about 20% is achieved by applying the second phase PASA.

5.2 Multiple products

Consider three different products with the multi-level assembly structures shown in figure 4(a), (b) and (c). The aim is to plan concurrent development of these three products. The same regime of holding and lateness costs as for the single product was assumed but with a due date of 900 days. Normal distributions were assumed for all operation times with variances increasing for assembly operations closer to the finished product. In all cases K=100 samples were used. The three two-phase methods are compared in figure 5(b). The second phase PASA is compared for the various inputs provided by the different first-phase methods. For multiple products an extra 10% cost reduction is achieved by applying the second phase PASA.

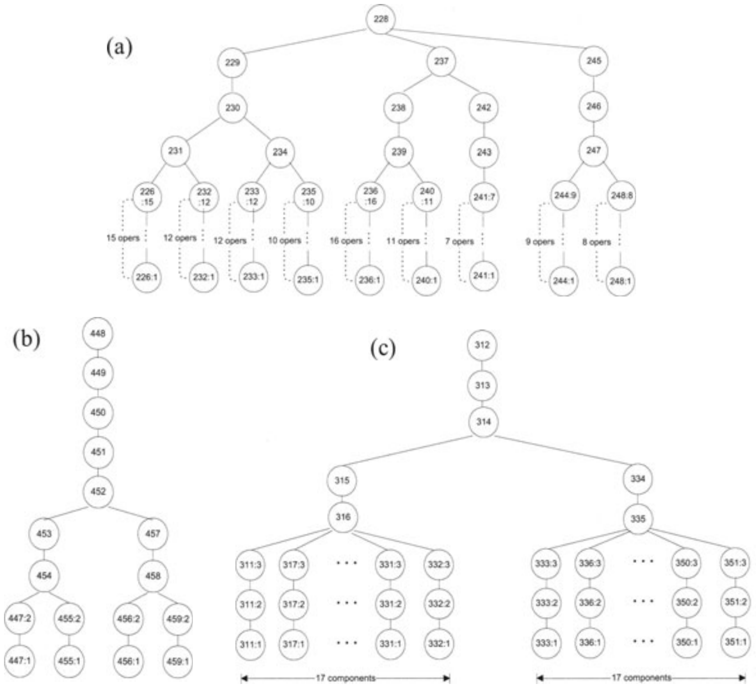


Figure -4. Examples of multilevel product structures

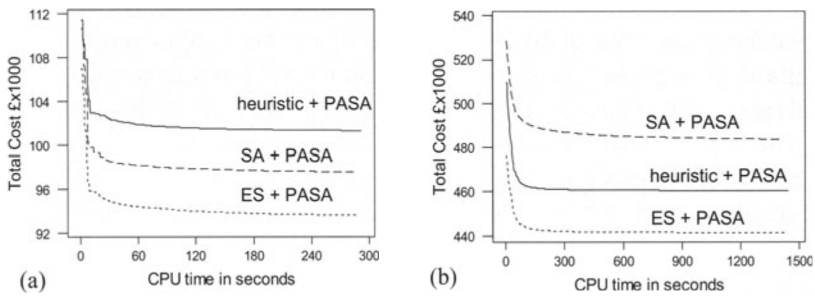


Figure -5. (a) Total costs versus CPU times at the second phase for (a) assembly of product in figure 4(a) and (b) concurrent assembly of three products in figure 4(a), (b) and (c)

For one phase methods the initial schedule $\mathbf{s}_0$ is obtained by a shifted backwards scheduling assuming infinite capacity and mean duration times. The shift, corresponding to letting the mean product completion time meet the product due date, helps to reduce SA and ES search time. The choice of parameters for SA and ES is made by repeated experiments on these and other similar complex examples. One-phase methods are applied to the single and multiple product cases (figure 4) to find optimal operation timings with the EPST priority rule (figure 6).

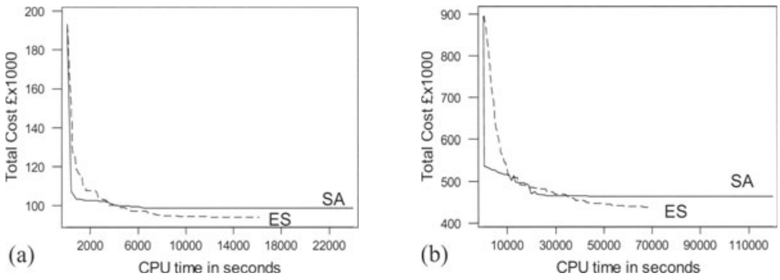


Figure -6. Total costs versus CPU time by one-phase methods for (a) assembly of product in figure 4(a) and (b) concurrent assembly of three products in figure 4(a), (b) and (c)

6. CONCLUSION

Planning large and complex products is a difficult problem especially when the operations have large uncertainties in duration. This is likely to happen for engineer-to-order products which are customised to a particular client's specification. Furthermore, companies which design and manufacture these types of product often have an unpredictable level of orders. Thus they retain core resources, but when several products are undertaken concurrently, significant competition for the resources occurs. The aim of an optimum plan is to minimise expected cost, which includes lateness and holding (i.e. work in progress or earliness) costs.

The methods developed in this paper go some way to solving this problem. One method uses a two-phase process. The first phase optimises the sequence of operations and the operation the start times (assuming means of operation times are set deterministically at estimates of their means). An heuristic is described which is computationally two orders of magnitude quicker than alternative random search simulated annealing and evolutionary strategy methods, with only a marginal reduction in cost of schedules for the industrial examples tested. The second phase uses the sequence of operations from the first phase and optimises timings for this sequence using perturbation analysis giving between 10-20% cost reduction. The other method optimises timings directly through random search but due to the stochastic operation times a priority rule is required both to implement the final plan as well as evaluate candidate solutions. Repeated simulation of sample processes for this evaluation makes the one-phase method computationally expensive. Both the one- and two-phase methods were implemented and validated on several cases with industrial data. Further research on re-planning when new orders arrive is currently being undertaken by the authors.

7. REFERENCES

[1] Hendry L C, Kingsman B G 1989 "Production planning systems and their applicability to make-to-order companies, *European Journal of Operational Research*, 40, 1-15.

[2] Bertrand J W M, Muntslag D R 1993 "Production control in engineer to order firms" *International Journal of Production Economics*, 30, 3-22.

[3] Wortmann J C, Muntslag D R, Timmermans P J M 1997 *Customer-driven manufacturing* Chapman & Hall.

[4] Giebels M 2000 *EtoPlan: a Concept for Concurrent Manufacturing Planning and Control*, Ipskamp BV

[5] Liao C 1992 "Optimal release time on a stochastic single-machine" *International Journal of Systems Science*, 23 1693-1701.

[6] Yano C A 1987 "Stochastic lead-times in two-level assembly systems" *IIE Transactions* 19(4), 371-378.

[7] Gong L G, Kok T, Ding J 1994 "Optimal leadtimes planning in a serial production system" *Management Science*, 40(5), 629-632.

[8] Homem de mello T, Shapiro A, Spearman M L 1999 "Finding optimal material release times using simulation-based optimisation" *Management Science*, 45(1), 86-102.

[9] Hasan C N, Spearman M L 1999 "Optimal material release times in stochastic production environments" *International Journal of Production Research*, 37(6), 1201-1216

[10] Yano C A 1987 "Setting planned lead-times in serial production systems with tardiness costs" *Management Science*, 33(1), 95-106.

[11] Yano C A 1987 "Stochastic lead-times in two-level distribution type networks" *Naval Research Logistics*, 34, 831-843.

[12] Song D P, Hicks C, Earl C F 2001 "Setting planned job release times in stochastic assembly systems with resource constraints" *International Journal of Production Research*, 39(6), 1289-1301.

[13] Aarts E, Korst J 1989 *Simulated Annealing and Boltzmann Machines*, Wiley, NY

[14] Azencott R (Ed.) 1992 *Simulated Annealing Parallelization Techniques*, Wiley, NY

[15] Brandimarte P, Maiocco M 1999 "Job shop scheduling with non-regular objective: a comparison of neighbourhood structures based on a sequencing/timing decomposition" *International Journal of Production Research*, 37, 1697-1715

[16] Schwefel H P 1995 *Evolution and Optimum Seeking*, Wiley, NY

[17] Baker K R, Scudder G D 1990 "Sequencing with earliness and tardiness penalties: A review" *Operations Research*, 38(1), 22-36.

[18] Rubinstein R 1986 *Monte Carlo Optimisation, Simulation and Sensitivity of Queuing Networks*, Wiley, NY.

[19] Fu M, Hu J 1997 *Conditional Monte Carlo: Gradient Estimation and Optimisation Applications* Kluwer, Dordecht

A FRAMEWORK FOR INTEGRATED PRODUCT AND PROCESS DESIGN

R. Holt and C. Barnes
School of Mechanical Engineering, University of Leeds, UK

Abstract: Simultaneous design and manufacture planning of new products can offer significant benefits across the product life cycle, such as reduced lead-time, more efficient manufacture, and lower costs. Research at the University of Leeds has focused on developing a new method for achieving this through the use of decision-support tools. These tools will help designers identify suitable manufacturing routes for their product, allowing early manufacturing planning and feedback so that designs can be made more suitable for manufacture. These tools will be underpinned by a series of units, each representing a set of design requirements or process capabilities. By selecting and integrating the appropriate units, designers will be able to compare their requirements with the capabilities of a range of manufacturing processes. This paper presents a methodology for constructing the data models that will make up these units. A General Purpose data model has already been developed and is presented here, with a discussion of how it forms the basis of the Situation-Specific data models needed for individual units. A method for using these units is also presented, and demonstrated for the case of Selective Laser Sintering. Finally, future work and potential improvements will be discussed.

Key words: Concurrent Engineering; Process Selection; Product Design; Data Modeling

1. INTRODUCTION

The advent of concurrent engineering has seen decisions on how a product will be manufactured taken progressively earlier in its development cycle. This early consideration allows manufacture planning as the product is designed, and helps avoid time-consuming and expensive iterations to adapt

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 473–482.

the product to the chosen manufacturing process [1]. However, as materials and processing technology evolve, the range of options available becomes wider and ever more complicated [2], and designers become more dependent on the experience of manufacturing engineers to inform their decisions.

This has led to the development of decision-support tools to capture manufacturing expertise so that design and processing decisions can be made quickly and effectively. A lot of work in this area concentrates on process selection, choosing the most suitable combination of manufacturing processes to make the product, and sometimes methods for optimising the process and calculating production parameters. Little emphasis is placed on feedback to make designs better suited to the preferred manufacturing route.

This paper focuses on the development of robust and extendable data models to underpin a decision support system that provides such feedback. These models will provide a framework for advising design decisions, and their impact upon the manufacturability of a product. This will support the adaptation of both the product and its manufacturing process during the design stages to ensure that the optimal combination is used.

2. OVERVIEW

Many process selection techniques are available, and a range of these [1,3-9] has been investigated in the course of this research. These take different approaches, from the generic methodology of Lovatt and Shercliff [3,4] to the domain-specific tools of Kaschka [5] and Er and Dias [6].

Each manufacturing process has its own set of capabilities and limitations and these are not always analogous or easy to compare. Lovatt and Shercliff [3,4] approach this by gradually rejecting possible manufacturing processes as the design progresses, leaving a small number – typically in a single domain –to be considered in more detail. Kaschka [5] adopts a similar approach, selecting a general "technological strategy", then considering the processes within that strategy in more detail. Er and Dias [6] have developed a selection technique specific to casting processes. However, a number of methods apply to manufacturing processes in general: Gayretli and Abdalla's constraint-based model [1] or Perzyk and Meftah's technique [7] of combining various domain-specific databases.

Most process selection techniques are reactive; identifying how suitable a process is for making a fixed design. However, some do offer the chance to adapt the design in one form or another. Er and Dias [6] use a series of decision levels, each concentrating on a different aspect of the design. At each level, processes unable to meet that aspect are rejected, and the user is offered feedback to explain why, and given the chance to amend their requirements and repeat the level. This method, however, only applies to casting processes. Perzyk and Meftah [7] propose a method that considers a number of possible designs as well as potential processes, and selects the optimum combination of design and process, but offers no specific feedback to the user. Brechet *et al.* [8] offer a loop for revising requirements based on the results of process and material selection – this, however, is to identify a valid set of requirements, by checking for expected and unexpected results. None of these methods offers specific feedback on how to adapt a design to make it more suitable for a given process: something this framework is intended to address.

3. THE FRAMEWORK

3.1 The General Purpose Model

The foundation of the framework is the General Purpose Model, shown in Figure 1. This is an information model – developed using the EXPRESS information-modelling language [10] – designed to capture the information necessary to contrast manufacturing processes and materials with the requirements of a given product design, in a process similar to that of Lovatt and Shercliff [3,4]

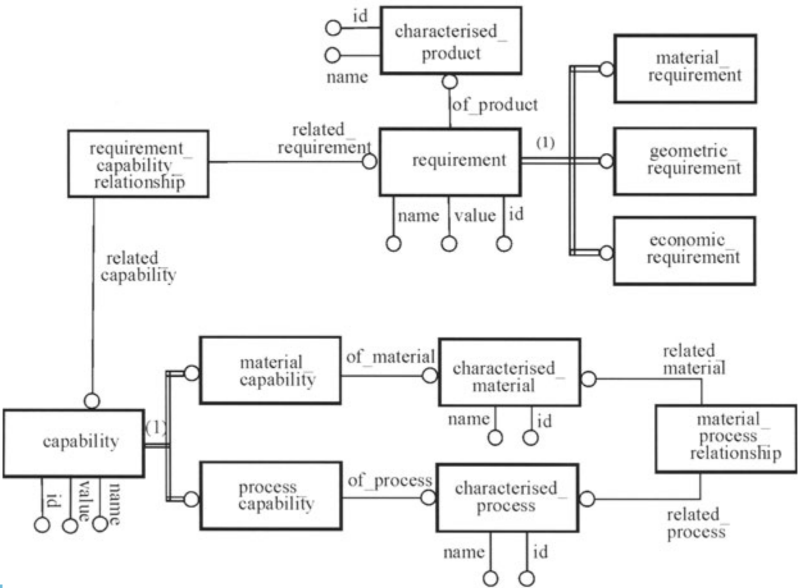


Figure -1. EXPRESS-G Representation of the General Purpose Model

The model tries to represent requirements, as they appear in the design, being material, geometric or economic requirements. The capabilities of the materials and processes are captured separately. This is not always accurate: the two capabilities can be dependent. For example, the accuracy of a manufacturing process can vary according to the material specified. The machine used can also affect the size of part that can be manufactured. However, it is felt that this adds significant complexity to the model and is outside the scope of the work at present.

Each requirement and capability is captured as an instance of the appropriate entity, with a name describing it, a unique identifier and a value – which may be a numeric value, or an alphanumeric string. For every requirement, there must be a related capability and vice-versa. These are linked by an instance of the *requirement_capability_relationship* entity. All requirements are attached to a product and all capabilities to a material or process. These are referred to as *characterised_process* and *characterised_material*, as they are only represented by their relevant capabilities, instead of every single capability that they possess. For example, Selective Laser Sintering's ability to produce conformal cooling channels is not relevant if the product does not have cooling channels, and should not be considered.

3.2 Situation-Specific Models

For a given situation, this model must be populated to reflect the specific product, processes and materials under consideration. This means careful analysis of these entities, to establish the exact requirements and capabilities that must be used, and requires expertise in the processes in question.

The following three-stage procedure is used to populate the model:

IDENTIFY REQUIREMENTS: The product requirements, including functional requirements, are identified and classified as "geometric", "material" or "economic".

RELATE TO CAPABILITIES: For each product requirement, an instance of the *requirement* entity is created, with a name and appropriate value. A corresponding instance of *capability* is added and linked to this requirement with an instance of the *requirement_capability_relationship* entity. The user must choose whether the requirement is a material or process entity. However, it is possible to link multiple capabilities to a single requirement – and vice-versa - if this is appropriate.

SPECIFY PROCESS LIMITATIONS: For each material and manufacturing process under consideration, an instance of the appropriate "characterised" entity is needed. All the limitations of each material and process must be identified, and added as instances of the capability entity. A corresponding related instance of the requirement entity and the *requirement_capability_relationship* entity is also needed so that the appropriate information from the design can be captured.

The approach adopted here will support the development of a library of standard instance data for a range of materials, machines and processes. These standard models will capture the limitations specific to these entities and will be placed directly into the General Purpose Model.

The capability and requirement entities already have attributes of name, value and id. It is likely that the user will wish to populate the name (a convenient description for the requirement or capability) and id (a unique identifier for convenience, when that entity is referenced by other) attributes as the instances are added. Unless appropriate values are known for the entities, then the value attribute will be left blank until the model is finished, and the remaining data is acquired.

3.2.1 Implementation and Population

Once the Situation-Specific Model has been established, it can be translated into some implementation form and populated with values for the identified requirements and capabilities. Requirement instances are populated using data taken from the current product design. Some of this information will come directly from the product requirements themselves, whilst others – particularly for those related to the specific limitations of a manufacturing process or material – will have to be acquired by examination of the design under consideration. To fully populate the capability instances, expert knowledge of the processes, materials and machines involved will be needed. Just as a library of expansion models can be developed for the specific limitations of various processes and materials, it would be possible to capture values for these limitations, and capabilities related to common requirements (common material properties, maximum dimensions etc.).

With the Situation-Specific Model prepared and populated, the requirements must be compared with their associated capabilities to generate feedback about the design relative to the processes available.

3.2.2 Comparing Requirements and Capabilities

Each requirement will be compared with the associated capability of each process and material under consideration. A status will be allocated to the capability of that machine, process or material using a feedback token – a flag indicating whether or not the capability can meet the requirement. At present, two levels of feedback token are available: *Go* (indicating that the capability can meet the requirement), and *No Go* (indicating that the capability is not able to meet the requirement).

The feedback tokens generated for each process and material are presented to the user. Based on this information, the user is able to identify which processes and materials meet which requirements. This will support both the process selection decision and the indication of which design attributes must change to make other processes or materials viable. It is left to the discretion of the designer to decide if the combination of process, machine and material is able to meet all the requirements, or whether the design must be amended to make it more suitable to another process.

These decisions would also help to highlight suitable manufacturing process chains for any component. For example, an injection-moulding die may require extremely fine features, but also conformal cooling channels [12]. In this case, SLS would be one of the few processes able to produce the conformal cooling channels, but it would not be able to produce the fine features required. Equally, EDM – a viable alternative in the production of dies – can achieve fine detail, but could not make the conformal cooling channels. The two processes are clearly complimentary.

4. CASE STUDY

The University of Leeds has been involved in creating a number of injection moulding dies by Selective Laser Sintering, and the use of this method has been illustrated by applying it to the design of such a die. The die was comprised of two parts: a fixed half, and a moving half. Selective Laser Sintering (SLS) had already been chosen as the production technique for the die in question, using the material RapidSteel™ 2.0, as this allows the production of conformal cooling channels, which can offer significant reductions in cycle times [11]. The proposed models were used as a validation check to investigate whether the process of SLS was suitable for the die geometry, with the feedback advising whether post processing would be required.

4.1 The Product

Table 1 shows the requirements identified for the two halves of a typical injection moulding die, with their related capabilities, the appropriate values used to populate the model, and the tokens allocated by the feedback process (as described in Section 4.3). For ease of comparison, units have been kept consistent. In this case, all lengths were measured in metres, irrespective of their magnitude. Consistency only needs to be strictly maintained between related capabilities and requirements, but it is easier – and less confusing – to use the same units for all measurements of the same type. In this case, the only other unit is µmRa, used to measure Surface Roughness.

Table -1. Requirements and related capabilities for the die halves, with values.

PRODUCT REQUIREMENT	VALUE - Fixed Half	VALUE – Moving Half	RELATED CAPABILITY	PROCESS CAPABILITY	FEED-BACK TOKEN
Conformal Cooling Channels ?	Yes	Yes	Can produce conformal cooling channels?	Yes	Go
Surface Finish (Moulding Surfaces)	0.05 µmRa	0.05 µmRa	Surface Finish	6µmRa	No Go
Height	0.053m	0.053m	Maximum Part Height	0.365m	Go
Length	0.133m	0.133m	Maximum Part Length	0.171m	Go
Width	0.118m	0.118m	Maximum Part Width	0.171m	Go
Corner Radii	0.01m	0.01m	Minimum Feature Size	0.002m	Go
Smallest Feature Dimension	0.00280m	0.00214m	Minimum Feature Size	0.002m	Go
Moulding Surface Tolerance	±0.0002m	±0.0002m	Accuracy	±0.0002m	Go

Features have been taken as geometric requirements, and only the smallest feature dimension was considered. This was on the basis that, if the smallest feature could be produced, all the features would be manageable (barring the specific exceptions given below). If SLS proved unable to manage this feature dimension, then the procedure could be repeated, listing every feature as a requirement to highlight exactly which ones the process could and could not manage. A similar approach was adopted for wall thickness (see Table 1).

4.2 The Process

Selective Laser Sintering (SLS) is an additive rapid-prototyping process, capable of manufacturing production tooling using materials such as RapidSteel™ [13,14]. In this case, SLS was carried out on a DTM Sinterstation 2000 machine using the material RapidSteel™ 2.0 [12]. Based on work carried out at the University of Leeds [13], the additional requirements to be added to take account of the specific limitations of SLS, and the values identified these requirements and capabilities are as shown in Table 2.

Table -2. Capabilities representing the specific limitations of SLS, their associated requirements, and values.

PRODUCT REQUIREMENT	VALUE - Fixed Half	VALUE – Moving Half	RELATED CAPABILITY	PROCESS CAPABILITY	FEED-BACK TOKEN
Minimum Wall Thickness	0.001m	0.006m	Narrowest Wall	0.0027m	Go
Maximum Overhang	0.02m	0.00225m	Largest Overhang	0.00225m	Go

4.3 Feedback

The General Purpose model was translated into a relational database using Microsoft Access™ and populated with the data described above. Every requirement for each die half was tested against its related capability and feedback tokens allocated as shown in Table 1 and Table 2.

The database allocated *Go* feedback tokens to every requirement/capability pair except the surface finish required for the moulding surfaces of the die, which was allocated a *No Go* token. In this case, the surface finish of the die was inherited from the requirements of the moulded product. Changing this would mean altering the moulded product, and this was considered unacceptable. However, instead of rejecting SLS outright, a post-processing operation (EDM) was selected as a means of providing the required surface finish.

4.4 Case Study Conclusions

The case study demonstrates successful application of the framework to a real situation. The General Purpose Model was easily translated into a relational database, though a better method for implementing the comparison of requirements with capabilities is needed, as the use of queries is laborious. The decision support mechanism using the SLS situation specific model guided the user towards the conclusions reached by industrial experts. This reflected the manufacturing method actually used and the finished die was able to make parts to the required standard, once the moulding surfaces had been finished using EDM.

The General Purpose Model was able to handle all the necessary information, though several issues were raised. There is a need to build measurement units into the model, and ensure they are consistent between requirements and capabilities. It is also necessary to highlight the nature of the relationship between the requirement and capabilities. For example, whether the capability must be greater than the requirement (as with largest overhang, in the case study) or less than the requirement (as with smallest wall in the case study), or one from a number of given values (Yes or No for Conformal Cooling channels, for example).

5. CONCLUSIONS AND FURTHER WORK

This paper has outlined the framework for the development of a feedback-based decision support tool, intended to help adapt product designs to better suit available manufacturing processes. The first stage of this method – a General Purpose information model that captures information about the product and available processes – has been presented here, along with a development process for Situation-Specific models. These models could also be used to support a variety of existing process selection methodologies; particularly those based around direct comparison of requirements with process capability. A case study has shown that these models can be used to support the design decisions throughout the development process.

The General Purpose model still requires some refinement. Some features that could be included are as follows:

- Verification of consistency between units using Select types;
- Validation to ensure that requirements and capabilities are comparable types (for example, both text, or both numbers, instead of trying to compare a text requirement with a numeric capability);
- The ability to represent requirements and capabilities by range of values;
- The extension of the *requirement_capability_relationship* entity to show the nature of the relationship (i.e. greater than, less than, exactly equal to, one of, etc.);

In addition to the improvement of the models, a more robust method is required for comparing the requirements with their related capabilities and generating feedback. The present method of simply comparing each requirement with its capability in turn does not provide sufficient information to support the decision-making process.

The framework presented here only helps the designer to screen processes on the basis of their compatibility with the given design, and adapt the design to be compatible with desirable processes. Ideally, the framework should also give some indication of *which* manufacturing processes are desirable - based on the importance of the design requirements, and how well they are satisfied - and how the design can be optimized. For this purpose, multi-criteria decision-making techniques such as Multi-Attribute Utility Theory [15] or the Analytical Hierarchy Process [16] may prove useful.

Finally, a suitable implementation form must be identified. While relational databases are very useful for storing the information, as they can relate directly to the EXPRESS model, a better mechanism for making the comparison and generating feedback is required.

REFERENCES

1. A. GAYRETLI, H. S. ABDALLA, "A feature-based prototype system for the evaluation and optimisation of manufacturing processes", Computers & Industrial Engineering, N°37, 1999, pp. 481-484.
2. A.M. LOVATT, H.R. SHERCLIFF, "Manufacturing process selection in engineering design. Part 1: the role of process selection", Materials and Design, N°19, 1998, pp. 205-215.
3. A.M. LOVATT, H.R. SHERCLIFF, "Manufacturing process selection in engineering design. Part 2: a methodology for creating task-based process selection procedures", Materials and Design, N°19, 1998, pp. 217-230.
4. A.M. LOVATT, H.R. SHERCLIFF, "Selection of manufacturing processes in design and the role of process modelling", Progress in Materials Science, N°46, 2001, pp. 429-459.
5. U. KASCHKA, P. AUBERCH, "Selection and Evaluation of Rapid Tooling Process Chains with PROTOOL" Rapid Prototyping Journal, Vol. 6, N°1, 2000, pp. 60-65
6. A. ER, R. DIAS, "A rule-based expert system approach to process selection for cast components", Knowledge-Based Systems, N°13, 2000, pp. 225-234.
7. M. PERZYK, O.K. MEFTAH "Selection of manufacturing process in mechanical design", Journal of Materials Processing Technology, N°76, 1998, pp. 198-202.
8. Y. BRECHET, D. BASSETTI, D. LANDRU, L. SALVO, "Challenges in materials and process selection", Progress in Materials Science, N°46, 2001, pp. 407-428.
9. N. RAMAN, D. CHAJED, "Simultaneous determination of product attributes and prices, and production processes in product-line design", Journal of Operations Management, N°12, 1995, pp. 187-204.
10. ISO 10303-11: "Industrial Automation Systems and Integration – Product Data Representation and exchange – Part 11: Description Methods: The EXPRESS Language Reference Manual", 1993
11. D. SCHENCK, P. WILSON, "Information Modeling the EXPRESS Way", Oxford University Press, ISBN 0195087143, 1994.
12. K.W. DALGARNO, T.D. STEWART, "Manufacture of Production Injection Mould Tooling Incorporating Conformal Cooling Channels via Indirect Selective Laser Sintering", Proceedings of the Institution of Mechanical Engineers, Vol 215, Part B: Management and Engineering Manufacture, 2001, pp. 1323 - 1332
13. K.W. DALGARNO, T.D. STEWART, "Production tooling for polymer moulding using the RapidSteel process", Rapid Prototyping Journal, Vol. 7, N°3, 2001, pp. 173-179.
14. A. ROSOCHOWSKI, A. MATUSZAK, "Rapid tooling: the state of the art", Journal of Materials Processing Technology, N°106, 2000, pp. 191-198.
15. D. L. THURSTON, J. V. CARNAHAN, T. LIU, "Optimisation of Design Utility", Journal of Mechanical Design, N° 116, 1994 pp. 801-808
16. E. H. FORMAN, S. I. GASS, "The Analytic Hierarchy Process - An Exposition", Operations Research, N° 49, 2001, pp. 469-486

A TOOL FOR THE DESIGN PROCESS TECHNICAL CONTROL

O. Povéda*, D. Brissaud* and O. Garro**
** Laboratoire 3S, INPG, Grenoble, France*

*** Laboratoire M3M, Université de Technologie de Belfort Montbéliard, France*

Abstract: The paper deals with the issue of the technical control of concurrent engineering projects. A new approach for designing aided-design-control tools is proposed. It is not based on the successive representations of the product to be designed, but on a model of the processes, which guide technical decisions. These processes have been analysed by observation/action in an industrial project and thereafter modelled. Discrepancies due to technical management of projects can then be characterised, helping to define more precisely the needs of designers to obtain technical control of the design process. Requirements for aided-control of the design process have been proposed. Finally, a mock-up of such a tool has been implemented by structuring and capitalising distributed design processes.

Key words: product design control, design control, viewpoints

1. INTRODUCTION

Consumers economically pressure companies and oblige them to ever more reactive, to renew product families more quickly. Market globalisation gives rise to increasingly fierce, international competition. The new configuration forces companies to make their product development process progress to analyse operations and to bring best products to the market. First of all, companies answer the challenge by improving their production methods. Now, they must adapt their design methods. The objective of product development evolution is to improve product design by addressing all aspects of a product, at each moment of its life. Addressing the life cycle of a product requires clarification of the problems, the requirements the

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 483–492.

product has to achieve at each stage of its life cycle. Consequently, the design process must include the points of view of all the people involved in its life cycle -the so-called life cycle actors- early in the phase of product definition. The design process tends toward a concurrent engineering approach where the actors are involved in the very early stages. It can succeed only if it is coupled with a change in company organisation. The structures specifically built to perform projects existed before concurrent engineering emerged but need to migrate to support the new design processes [1]. This evolution of engineering projects goes with the creation of a new job: the project controller [2], [3], [4]. A project actor is a person involved in the design project at any moment of the life cycle of the product. A project actor can be a person or a team of persons.

2. CONTROLLING A DESIGN PROJECT

Company management is evolving and in this process, a concurrent design project becomes an independent feature able to produce and choose its own rules and own action process [5]. Inside these independent features, the control actor is the keeper of these rules and action processes. He leads all necessary operations during project studies, development and operations. A concurrent design is specific due to better integration of the different points of view during product definition. Integrating actors' knowledge and know-how has to be managed, decided and planned to obtain the best coherence of the process and achieve the global optimum of the product. Optimising the product does not only address the product's technical features. It must also include economic, time and cost constraints. Therefore, controlling a project consists in controlling the activity regarding the set of constraints to be complied with. Control activity addresses a lot of areas of expertise. Tools to aid control activity have been made available and support cross-domain activities; their add-ons are different depending on the features they represent.

3. THE ISSUE OF TECHNICAL CONTROL OF A DESIGN PROCESS

To be efficient and widely used, control-aid tools have to perform the needs specific to each project. How should the needs of a particular project be addressed? Tools in use can partially complete the task. Efficient control-aid tools are those which address the control activity through the domains common to projects. They are based on common references, and in fact the projects must adhere to the same market and budgetary rules. Managing time remains a common constraint, even when the scale depends on the project. As a consequence, it can be noted that the tools currently used in project control activity are based on those topics. When one addresses topics specific to the project so the product's technical content, supporting the control activity becomes more difficult. First of all, the difficulty comes from the heterogeneity of the technical features handled by project actors. From one project to another, the nature of the features varies a lot and even their existence or progression cannot be forecast. Moreover, a product under design continuously moves and is transformed. Those evolutions are very specific and are quite impossible to be planned when there is some innovation in the project. Progress comes from the confrontation of the points of view expressed by actors belonging to diverse disciplines, with diverse skills and competencies, operating in diverse logics of action. All these points of view which stem from the product life cycle activities have to be integrated in the global solution. The technical content of a project moves quickly and continuously; it is the main limit to the use of CAD systems as control activity support. We can therefore note that the objective of an integrated design project is to achieve a suitable system definition by confronting points of view and specifying the product structurally and physically. As we have seen above, many tools can be used to control projects by cost, quality and time constraints, but tools to help the project leader to anticipate the coherence and the integration of technical solutions proposed by the design actors do not yet exist.

Among the most difficult control tasks to be planned are the integration of the points of view and the way they are translated into product specifications. This is the main reason why aided-tool are not present in this activity. Because there is no planning, design actors find it hard to develop a structured method in this field. Consequently, the control actors continuously zig-zag with the progress of the technical content of design projects. Generally, they operate only in a reactive way. This means that the current state of the product must be published and then the discrepancies checked for the design process to be redirected down the right path. This reactive actors' position is closely linked to the nature of the tools used. They are based on a representation of the current state of the system to be designed. The representation has been built by capitalising on decisions making distributed through the actors' network. This system model gives a global view of the product progress to the control team but does not allow the design process to be anticipated and finely turned by integrating the points of view. Therefore, a control actor must consider the product progress by centralising the points of view of all the design actors, then translating them into technical focuses, and integrating them into the current solution in fine.

The points of view, the techniques, the methods, the objectives of the actors are extremely diverse and lead to conflicts between local solutions defined in a distributed way. Technically mastering a project does not only consist in settling conflicts; the pilot actor must help design experts to identify then clarify and solve conflicts as soon as possible. This is the essence of efficient control; they have not to avoid crises but to clarify them early when many ways are still available to settle them. To do so, a control actor must have tools to, on the one hand anticipate the local on-going solutions before final definition, and on the other hand track interactions among all the distributed on-going solutions.

Therefore, tools currently used to control design projects address the project technical content by monitoring the technical decision-making throughout the project. We propose a new way to design tools to aid the technical control of design projects. Regarding the role to be played by a control actor, it seems that he needs to know and master the distributed design processes that lead to decision-making on the technical content of the project. Our approach consists in proposing a tool based on the models of the design processes rather than on the successive product representations (Fig. 1).

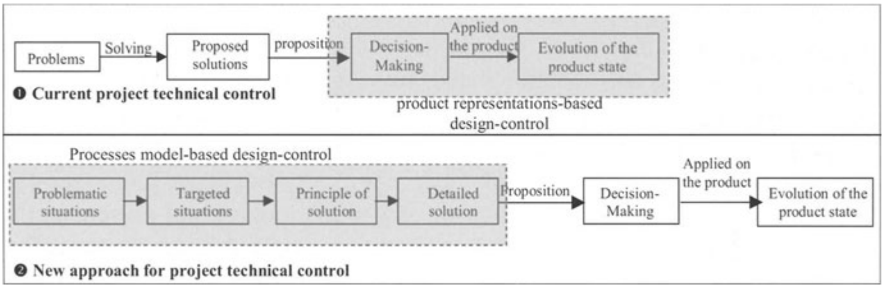


Figure -1. A new approach to instrument the technical control of projects

4. A DESIGN PROCESS MODEL

We are going to try to understand what the running approaches are and what are the features built by the control team when trying to co-ordinate technically and organisationally the design process. The model tries to reflect that. We stress the fact that the model has no vocation to be normative. It has not been built to describe a process that should be, but it results from an analysis of actual observation when designing. Our work in industry tends to propose two features closely linked and present in all distributing design activity: the problematic situation and the targeted situation.

4.1 The problematic situations

When the project started, the first actions taken by the men involved in the project were to define and make explicit the need to be satisfied from a functional point of view. First of all, we have got a need to be satisfied expressed by functions to be assured, and a set of physical principles chosen to achieve the functions. The main project consisted in progressively specify more clearly the system to be defined. In an integrated design process, the specifications involve the whole life cycle of the product. Therefore the actors can express some structural specifications, from the physical principles to the elementary component. Progress in an increasingly detailed definition of the system to be designed cannot be reduced to an aggregate of results and local solutions planned beforehand. At each stage of the project and in each domain, problematic situations occur all along the design process [6]. A problematic situation is a configuration of the project which does not enable the work to properly go on. It is a set of circumstances, events, parameters or features, which impact together to produce new situations the designers have not foreseen or never imagined. They are situations in which the established procedures cannot be used efficiently. By nature, a problematic situation addresses several points of the project and it is very rare that it is related to merely one area of expertise. A problematic situation generally arises from the interaction among several features depending on several project domains. So, the notion of a problematic situation covers the set of relationships among the numerous features of a project whose antagonisms expressed through problems are only the visible and formalisable part for actors involved in the project. When designers find a locked situation, they generally know that they are in a problematic situation because they feel it, even if they cannot express it totally and in a unique way. Contrary to a problem expressing the formalisation of both causes and observable consequences, a problematic situation formalises the visible consequences whose causes can be interpreted. These situations must necessarily be dealt with for the project to continue. But how can these situations be dealt with, how can corrective action be taken in to a configuration where consequences are known but where cause may vary according to the points of view and interpretations of each person and in fact are not accurately defined? Introducing problematic situations into the model of a process design allows tracking and understanding the orientations a process can take to maintain progress. Our stay in the company helped to highlight the fact that the control team does not seek at all costs to formalise the problematic situations in order to deal with them, but that it builds a specific feature which we will call the targeted situation.

4.2 The targeted situations

To answer problematic situations they face, the person involved in the design process build targeted situations. A targeted situation is more than a formal problematic situation. When designers try to define and analyse the situation they are in, the problem they try to extract generally becomes so constrained and complex that there is no longer a global solution to fit it. Therefore, the designers do not really formalise the situation to be achieved, which will enable successful continuation of the project. Unlike a problematic situation, a targeted situation is entirely delimited. Designers define explicitly the features of the project or product addressed by the search for improvement of the project configuration. They also define the domains where action will enable the new-targeted situation. Such work is typically the activity of project control; the control team must do it. A targeted situation is two-sided. It can be considered, on the one hand as a formal expression of the solution to a met problematic situation, and, on the other hand, as the construction of a newly identified problem. Indeed, a targeted situation is originally built as a theoretical solution that seems runnable and answers the problematic situation. Therefore, a targeted situation is not a detailed chosen solution. Only a few precise elements have been defined at this stage on the means to achieve the solution. In that sense, the targeted situation cannot be considered as a initial statement of a problem the designers must solve. In fact, the targeted situations are generally expressed by technical functions to be performed.

4.3 From problematic situations to targeted situations

A control team is in charge of planning and co-ordinating the activities of all the persons involved in the design process. Its quite exhaustive knowledge about the current state of the project and especially about the current state of the product, gives it the opportunity to efficiently co-ordinate and make coherent the whole specifications, requirements, constraints or solutions principles contributed by the points of view of the project actors. That is why the control team must support the building of a targeted situation, when a problematic situation occurs. The control team itself has the responsibility for that approach; the construction itself results globally from a co-operation activity. The control team cannot define alone all the impacts of a problematic situation. It needs to enrol knowledge and skills relative to all the domains involved in the situation. A pluri-disciplinary group must be built. Actors must co-operate to achieve a joint definition of the targeted configuration. First of all, co-operation leads to the definition of a joint objective, shared by all the actors. It also enables to build a frame for the actions. This framework defines the domain in which the group is allowed to act on the project. It can be formalised by some constraints, which define the limits of the solving approach, even if it does not really define the approach to achieve this. In addition to the construction of the targeted situation, the actors also build some criteria to assess the design. These criteria are defined independently of the problematic situation. They are a double vocation: they enable definition of a set of properties used to measure the performances of the solution to be chosen, and they are also used as tools deepening the definition of the targeted situation during processing. So, the actors jointly define criteria and targeted situations.

When defined, a targeted situation has to be distributed through the network of the project. This role is allotted to a particular actor who will be in charge of it and must achieve it. Choosing an actor is tightly linked to the content of the targeted situation and the expertise domain of the person (Fig.2).

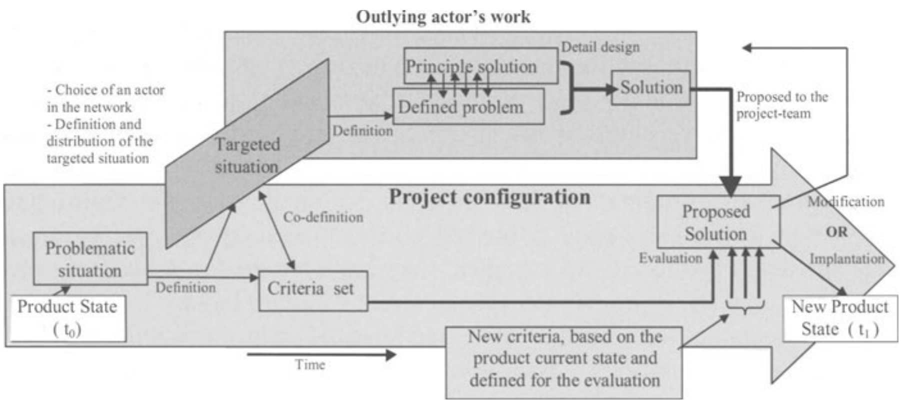


Figure -2. Distibuted design activity model

4.4 Work to achieve targeted situations

A targeted situation is accurately described by objectives to be achieved. It remains inaccurate on the working conditions to achieve this. The first action undertaken by an actor to perform a targeted situation is to transform this specific and very contextualised feature into a new problem, which can be solved by his domain mechanisms. He builds a problem able to support theories and techniques mastered by him. When the actor works to achieve a targeted situation, he goes into a logic (a rationale) different from the logic used by the control team. This actor situates himself outside the integrated running of the project and performs a mono-disciplinary work. So he can work in his original discipline logic with his original discipline criteria and assessment more or less far from the integrated work. When targeting a situation, an actor builds then expresses a problem on which he will apply his technical skills; the problem must be formulated in such a way that he can solve it. Even he does not know how to solve it at the moment, he knows that a solution exists. Then the actor tries to map out the formulated problem and the bank of solutions, possibly a bank of solution principles, to be found worldwide or specific to his expertise domain. He can choose the solution or principle best fitted to the problem specifications. If he succeeds, he details the solution or the principle by enriching it, adapting it to the specificities of the project and the product structure (Fig 2). When solving the target situation, the actor must simultaneously handle the problem he has formulated and the solution principle he has chosen. He simultaneously refines the solution to the problem and refines the initial problem [7]. Due to his outside and disciplinary situation in which the actor tries to achieve the targeted situation, it is frequent that the control team is blind to the solution principle chosen and detailed at any given moment. This information frequently reaches the control team only after the solution has been published and proposed to the group for integration into the global on-going solution (Fig. 2). In a project of that size, numbers of actors simultaneously work to achieve targeted situations. They are quite independent, and generally do not take into account current progress on the project and may not involve co-ordination. The different solutions proposed from these separate works are finally published and evaluated by the whole project team. When evaluated and accepted, they are integrated in a current global solution to make it progress to a new current solution. In such a process, the control team has to determine links, interferences or incompatibilities among all those local solutions, possibly existing among themselves or between a single individual and the global project. Such interferences are very difficult to forecast or anticipate because they are very close to the nature of the chosen solution, this choice being made by the actor outside the project when the control team is generally blind to the work.

This model of distributed design processes enables us to formalise the different levels where discrepancies occur due to the project control activity. That analysis of the discrepancies has then led to proposal for a tool to control a design project.

5. A NEW AIDED-TOOL FOR TECHNICALLY CONTROLLING THE PROCESS

From the overall considerations, a new aided-tool has been proposed to answer the needs highlighted above. The now tool has been built to help the control actors be more efficient in their task of technically mastering the projects. It must provide a description of the activity of all the actors involved. It must be a space of co-ordination in which actors may express their points of view. Moreover and of course, it must also be a platform for constructing joint items, which can be expressed in the actor's own language. So, the discipline knowledge and the discipline objectives may be confronted. The tool can be expressed following two axes according to their use. The first axis has been dedicated to the co-operation between the control team and the project actors. It is used to structure exchanges and actors capitalisation approaches when solving targeted situations. This part of the tool consists of several modules to characterise problematic situations. When the situation has been characterised, a new module aids a multi-disciplinary group (control team plus discipline experts) to define targeted situations. The targeted situation can therefore be transmitted to a specific actor in charge of achieving it. This actor can choose and characterise a solution principle. If he cannot, he may try to alter the targeted situation. When the solution principle has been characterised, the actor's work consists in detailing a solution, which will be evaluated. The process is supported by modules, which also add information into a database. The database is addressed not by product states but by information relative to on-going modifications on the product. The team actors may centralise the technical contents of the approach in progress in the database. It may confront the activity of separate actors running in parallel to track the possible conflicting consequences stemming from the local solutions in progress. From this information, the control team may anticipate the future product development and if necessary re-organise activities to be sure the distributed processes will converge. The control team acquires real capacity for early detection of interferences between local solutions and for alternative processing of the technical decision in an engineering project (Fig 3).

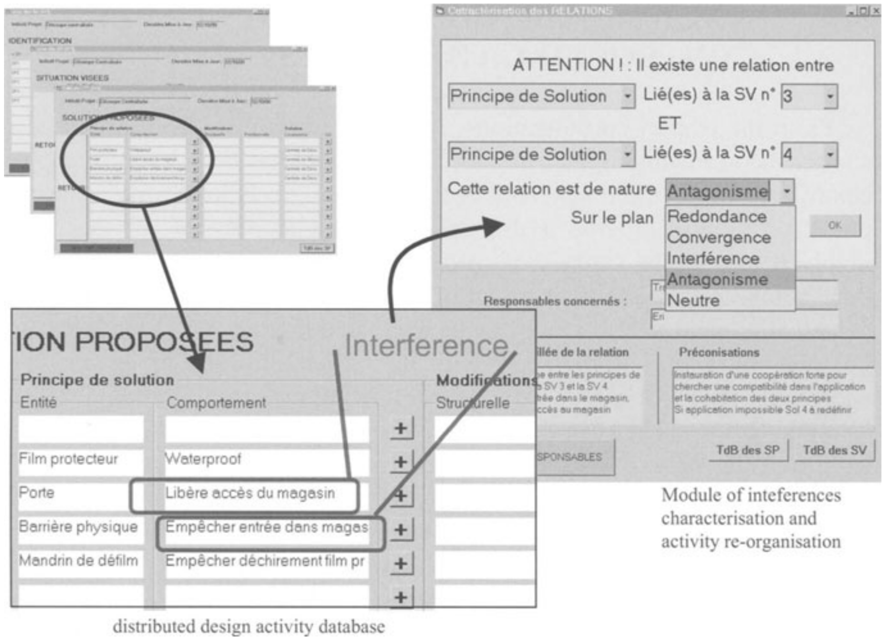


Figure -3. Technical solutions interferences anticipation

REFERENCES

[1] Afitep AFNOR « Dictionnaire de management de projet » 1992, AFNOR, Paris

[2] C. Midler, « L'auto qui n'existait pas, management de projet et transformation de l'entreprise » 1993, Inter-éditions, Paris

[3] C. Midler, « L'acteur-projet : situations, mission, moyens » 1993, in Ecosip, « Pilotage de projet et entreprises, diversité et convergences » édition Economica

[4] C. Fray, V. Giard, I. Stokes, « Le pilotage économique des projets », in Ecosip, « Pilotage de projet et entreprises, diversité et convergences » édition Economica

[5] G. de Terssac, B. Maggi « Autonomie et conception » 1996, in Coopération et Conception, sous la direction de G. De Terssac et E. Fridberg, édition Octares

[6] D. A. Shön « Le praticien reflexif, à la recherche du savoir caché dans l'agir professionnel » 1983 Collection formation des maîtres, Les éditions logiques, édition française 1994

[7] E. Blanco « L'émergence du produit dans la conception distribuée, vers de nouveaux modes de rationalisation dans la conception de systèmes mécanique » 1998, thèse de doctorat de l'Institut National Polytechnique de Grenoble

TOOLS FOR DYNAMIC SHARING OF COLLABORATIVE DESIGN INFORMATION

V. Riboulet*, P. Marin*, F. Noël*, B. Delinchant** and L. Gerbaud**
** Laboratoire Sols Solides Structures, Grenoble, France*
*** Laboratoire d'Électrotechnique de Grenoble, France*

Abstract: The development of Internet technologies supports the design of more efficient collaborative tools. This paper defines a collaborative environment based on theses technologies. It sets the conditions of use of this environment and points out the needs in information sharing. From this analysis we define tools whose functions complete currently available. Two new specific pieces of software illustrate the starting point of this environment development. The first tool, CoDVS, offers functions to archive data shared for the collaboration between actors. The XML format is used to achieve this goal while other functions give opportunities to update project data in a CVS fashion mode. The second tool, CoDISS, allows the dynamic definition of concepts shared by several actors, together with the connection of these data with the working parameters of each actor. Actors collaborate together by accessing this shared database that they can edit. They also automatically update their skill oriented models to reflect changes decided by their colleagues. Updates are automated with an API using the CORBA protocol.

Key words: Collaborative design, data sharing, Internet, Synchronous and Asynchronous collaboration, Archiving.

1. INTRODUCTION

Industrial Design is an activity where collaboration is fundamental. Indeed, many actors collaborate to achieve a common goal i.e. the definition of a product that can be used and may be sold. Each designer has its own knowledge and know-how, its own uses. If everyone involved speaks about the same object, the product, the representation that each actor uses will not necessarily be the same as the others.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 493–502.

In this context, collaboration must be organised to allow a better communication between actors. As shown in industrial cases, especially in large projects [1], physically co-located meetings are often expensive, impossible to manage, and may not be very efficient. This paper proposes an approach to develop collaborative work regardless of distance, where each actor is allowed to use the best representations for his own analysis, by being disconnected from a central data base or unique CAD software. To achieve this goal we must define new tools to help designers to share their knowledge and common data, while ensuring compatibility in order to enable definition of the product.

We propose a framework based on Internet technologies to let designers communicate together while they use their own tools. Our developments were considered in order to improve the collaboration between specialists from Mechanical design and Electrotechnical techniques. However we assume that the tools proposed here are drafts for more global solutions.

2. A FRAMEWORK FOR COLLABORATIVE DESIGN

2.1 Assumptions about the context

As said in the introduction, the challenge for software developers is to provide users with solutions that allow the best compromise between communication among themselves, and ergonomics for each of them in their daily activity. The commercial solutions commonly developed in the CAD context over several years are of two categories.

The first solution is to impose the use of the same CAD/CAM software on every actor involved in the design process, and to develop many modules in that software (geometry modelling, various simulations, different manufacturing/assembly processes, drawing and bill of material edition, etc…), in order to try to cover the needs of each actor. In this way, everybody in the project is working around the same geometric kernel model [2] [3], and there is no need for information translation, or if so, it is provided by the CAD module itself. But these modules are generally not very convenient for each actor. Craftsmen often prefer to use a simple and efficient specific piece of software designed for their trade.

The second category of solution is to let users work with their preferred specific tools, but to impose common PDM (Product Data Management) software, in order to ensure that each available piece of data is always unique and up to date (may be keeping a history of versions). But in this way, you have to develop data translation procedures between the different software used in the project. It appears in practice that these translation procedures are always perfectible because of the variety of configurations and concepts being used by one piece of software and not by others. Moreover actors miss comprehension of information because of the lack of context specification in these automatic exchanges.

2.2 Framework specifications

That is why we state positively that focusing on communication between software [4] and direct exchanges of predefined concepts cannot lead to an efficient communication between craftsmen or every actor in the design process. Therefore we turn towards communication between people who can have discussion within the project context, explain each to other their specific concepts and parameters, and then define and format dynamically shared entities representing this needed shared information. Let's say that this approach is taken from the concept of "intermediary object" proposed in [5]. In this way, every information exchange becomes clear for every actor concerned with it, without misunderstanding, and the translation procedure may be defined by each actor according to their own knowledge and practice. In this context, our aim is to develop a WWW-based collaboration system, involving together either existing communication software for direct and informal dialog between actors, existing specific trade software for every actor's technical purposes in the design project, and common modules designed to allow dynamic definition and sharing of customisable concepts.

2.3 Recent solutions for collaborative work

Our study of available technologies, in terms of tools that allow collaborative work over the Internet, may be summarised as follows:

A first category of software concerns classic tools like e-mail, chat, newsgroups, instant messaging (IM), shared whiteboard, file transfer (ftp), video-conferencing (VC), application sharing (AS) and common data base sharing (DB). Some of these provide synchronous data exchange (chat, IM, whiteboard, VC, AS), the others asynchronous (e-mail, newsgroups, ftp, DB). Some of them allow technical structured data to be shared or transferred (ftp, AS, DB), while the others are mostly dedicated to the exchange of non structured data or information. Some of them are useful for text information (chat, IM, newsgroups), and others may be used for graphic (draft sketches or precise geometric definition, 2D or 3D) representations of information. A few types of software may put together several tools in a collaborative environment. Let's cite Microsoft NetMeeting, HP Shared-X and SGI Inperson. SGI Inperson provides the ability to share 3D CAD models on the whiteboard, which is very useful in mechanical part design activity.

Table -1. An overview of commercial collaborative design tools

	Technology				**3D Geometry**				**Collaboration**						**Documents**			**Project management**		
Tools	Dedicated application	Via Browser	ASP	P2P	Proprietary engine	Visualisation	Manipulation	Annotation	Chat	Forum	White board	Private Messages	Video-conferencing	Inter-appl^{n} Data Sharing	Archiving	Versioning	Conflicts resolution	Dedicated module	Capitalisation	File access manag^{t}
One Space	x	x				x		x	x		x	x								x
Centric	x	x			x	x	x	x	x	x		x	x		x	x	x	x	x	x
Alibre	x		x		x	x	x		x	x		x			x	x	x	x		x
Reality Wave		x			x	x	x	x				x								
Groove		x		x		x	x	x	x			x				x	x	x	x	x
E-Vis		x				x		x	x	x	x	x			x	x	x	x		x
Informative Graphics		x				x				x					x					
Nexprise						x		x	x				x		x	x	x	x	x	x
Co	x			x										x						

But a new category of tools has appeared recently, tools that are specifically dedicated to collaborative technical-product design over the Internet. They plan to offer a team of engineers (including design, production or marketing offices, sub-contractors etc.), a way of working together on a technical activity, sharing data about the concerned product [6].

The operational aspects of commercially available tools have been classified in table 1. In the first column we focus on the technology used (Dedicated application *vs.* browser; application service provider *vs.* peer to peer architecture). The second column shows the possibilities of interaction with the geometry of the product: visualisation, manipulation, the possibility of adding annotations and whether a tool only allows use of its own dedicated 3D engine. The third column is about the way these tools support collaboration: chat, forum, blank screen, private message, video-conferencing, and the ability to share data between different applications. The fourth column focuses on the way the software deals with document management: ability to save, resolution of conflicting versions. The last column deals with project management: are these features available, is there a management of files access rights, do they support capitalisation ?

2.4 The context of our design experiments

Collaboration begins whenever two different skills are represented in a working meeting. For the purpose of our project, collaboration between Electro-technical engineering and Mechanical engineering is considered.

Electro-technical engineers are mainly concerned with electric current, magnetic fluxes, material properties, interface between parts. To analyse the part, they use analytical models, the finite element method, generic mathematical tools etc. Mechanical engineers are mostly concerned with the way to assemble parts, shapes to be produced, material properties, manufacturing processes, structure behaviour etc. and they mainly use CAD tools, the finite element method, manufacturing modules and process planning.

Both, Electro-technical and Mechanical engineers, work simultaneously and must share the results of their analysis. It is not possible to enumerate every concept that they need to share. It mainly depends on the project context and must be defined dynamically. Therefore one must provide the designers with a framework designed to manage synchronous meetings where dynamically defined concepts are to be shared.

2.5 A schematic view of Framework specifications

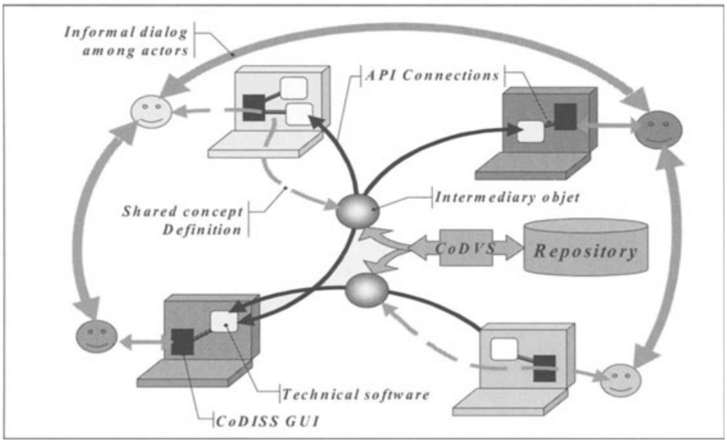


Figure -1. A view of information system

A schematic view of our collaboration system can be found in figure 1. Apart from tools dedicated to direct informal communication, like chat, IM, whiteboard, VC or AS, one must provide every person involved in the design project with a GUI (Graphical User Interface) that allows the construction of intermediary objects to be shared with their colleagues. Then connections with technical software are developed using each dedicated API (Application Programming Interface). These links are created on one hand between the shared objects and some features or parameters in the technical software, and on the other hand between the CoDISS GUI and the technical software. A system dedicated to store and manage shared objects has been also developed and is connected to a repository. Among the latest generation of distributed objects development tools [7], we chose to manage shared information with the CORBA protocol. A first mock-up of this kind of software connection has been proposed in [8].

3. SPECIFIC TOOLS

3.1 CoDVS : Collaborative Design Versioning System

CoDVS was defined to answer some collaborative data management issues during a design process and is based on the following requirements :

a) Manage dynamic data and information creation and storage,
b) Manage asynchronous access to data and information,
c) Manage data and information versioning,
d) Allow concurrent access and modifications of data and information.


```
  </Data17>
- <Data18>
    <DataName>cyl_head_thick</DataName>
    <Value>6</Value>
    <Type>double</Type>
    <ProjectName>DecEM</ProjectName>
    <OwnerName>PM</OwnerName>
    <Comments>Thickness of cylinder head. Probably magnetic role. Increased to impr
      mm ; Tolerance +/- 0.1</Comments>
  </Data18>
- <Data19>
    <DataName>Cyl wide</DataName>
```

```
- <Element path="Sauvegarde_CoDISS/Donnee00/Valeur" placement="1.1.1">
    <LastVersion Value="1.1" Date="1/9/2001 15:33" />
  - <OldVersions Value="1.0" Date="1/9/2001 15:32">
      <ModifyText Value="525.2" />
    </OldVersions>
  </Element>
```


Figure -2. Parts of XML file generated by CoDISS, and managed by CoDVS.

Data versioning with access management already exists in commercial products like PDM tools, but does not match our requirements for automatic interaction between co-operative tools. Collaborative design via the Internet needs to adopt a format compliant with internet protocols. A new text file standard has appeared recently, called *XML* (eXtended Markup Language) that allows the exchange of structured information over the Internet, and therefore is the most suitable format for our purpose. A sample of XML structured data is described in figure 2. The XML format allows a tree structured data organisation, with capability of dynamic structure modification. That's why it has been chosen for the CoDISS intermediary objects architecture. Furthermore the generic capabilities of the CoDVS tool may be used for any concurrent XML information versioning system.

The main principle of CoDVS is based on an existing tool called CVS (Concurrent Versioning System), that provides file or directory version management. But instead of directories and files, our tool considers branches and leaves of the XML tree. Each designer's personal data is stored through a client version (figure 2) of CoDVS in a repository located on the server. This repository consists of an XML file, containing the 'up to date' data and the whole history of modifications.

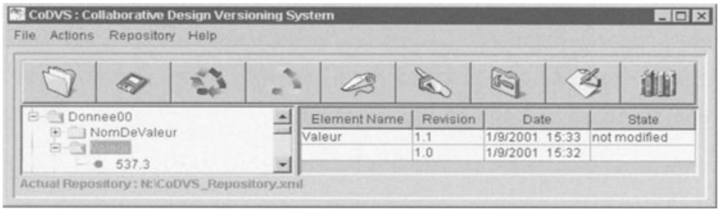


Figure -3. GUI of a CoDVS client

Let's have a look at an example that illustrates CoDVS functionalities. A mechanical engineer wants to share technical information with an electrician. Some data are only defined on the mechanical side, while others are defined on the electrical side, and some data have to be shared. Depending on the side where they are used, data can be 'read only' or 'read/write'. Let's assume that each designer have at their disposal an 'up to date' version of shared data on their client, taken from the repository. When a designer modifies a value that is used by another designer, the XML file is transferred to the server which compares data in the XML file with that in the repository to identify changes. The first phase of this comparison is achieved by an automatic analyse of likenesses between old data and new data to identify the corresponding leaves and branches of the XML trees. Due to the dynamical data structure, a probabilistic algorithm, using likeness criteria, is needed to estimate the corresponding tree nodes of the two XML files. After this node association, the comparison method inspects data modifications, creations or removals and stores this information in the repository. The other designer, notified of these changes, updates his data in his client. Among these changes, some data may have been used by the second actor in a 'read only' mode, but other data may have been modified simultaneously by both actors. This conflict is detected by CoDVS and must be managed by the second designer. Thanks to the repository versioning system, each data version of the XML file (see figure 2) can be retrieved and replaced in 'up to date' data.

3.2 CoDISS : Collaborative Data and Information Sharing System

The ability of CoDVS to provide a powerful and flexible way to manage data in a distributed project is completed with another tool, CoDISS, which is a special tool to dynamically share data, concepts, and more generally information. During a meeting (virtual or not) actors decide to check a given property of the product. After the meeting they will have to analyse shared properties with their own tools. The analysis result leads to a new version of the property. Any one and perhaps all the collaborators may request the current state of the product. CoDISS is principally made up of a GUI (see figures 1 and 4), API connections with several types of software, and the ability to share objects over the net.

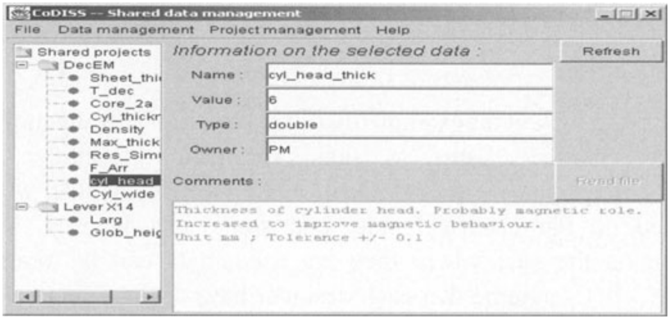


Figure -4. A view of the GUI of CoDISS, the application to manage shared objects

The GUI is dedicated to both structure definition of shared objects and management of the connections of shared parameters with the user specific software. It allows each actor to define any new object, by combining basic concepts in a specific way (numeric parameters, text fields, data files, geometric features, links or pointers to other objects…). The GUI must also provide a way to describe the relation between these concepts or parameters and the related concepts in the technical software. It is important to note that as the design process goes on, intermediate object definition as well as connections to technical software are managed dynamically by the actors.

Every new object provided by an actor in the design project is put on the net by the way of the CORBA protocol, so that other people can access it, and then connect it to their own software. Thus CORBA allows connection between two or more completely different types of software.

Having defined an intermediate object, or having collected it from the CORBA bus, an actor often needs to connect it to his personal software parameters. This can be done by the way of a specific interface, if possible developed using the API of this software. In some CAD software (like SolidWorks or Pro/Engineer), the API provides the user with access to an internal product data base as well as to nearly every internal function to modify the product model. So, a program module using this kind of API is able to perform any interaction between information available in CoDISS and the CAD software product model. This is the way we experimented to connect mechanical CAD parameters to some parameters of intermediary objects: the API adds a menu item in the CAD software, which opens a dialog window showing shared objects and asking the user to select the CAD parameters his wishes to refer to. When this connection has been made, the other actors working with CoDISS (or any other software connected by the collaborative CORBA bus) may refresh their GUI and see the new parameters extracted. Every one may then change the value of this shared parameter according to its specific constraints, but in this case the others are asked if they agree and whether they want to update their model. If yes, the modification is automatically propagated to their particular software parameter.

In the case of “Ansys-like” software, there is no actual API. The user interface is made of a command file where a series of instructions can be found. Let’s recall that Ansys is a mechanical structure simulation software. All information about the simulation is located in this command file: geometry description and parameters, material characteristics, boundary conditions, loads, element types, expected results …

Obviously, connections between CoDISS and some of these parameters must be made. And the only way to get or modify simulation data is to read/write in the command file. That is what we have done. Analysing the Ansys file gives the list of available Ansys parameters and the mechanical engineer may then connect them to CoDISS objects.

As shown in figure 2, the shared parameters from CoDISS are stored in XML files, compatible with CoDVS. Thus CoDISS and CoDVS are complementary tools to generate and manage shared information in the collaborative design process.

4. CONCLUSION

The aim of the set of software proposed in this chapter is to support dynamic creation and sharing of “intermediary objects”, linking them to skill oriented software, together with storing and managing versions of this shared information in a collaborative engineering design context.

Further studies will lead us to look at dynamic connection of new technical software, *i.e.* as the design process goes on. Another limit of CoDISS tool is that designers do not use the same parameters in their own software as their colleagues do in theirs. We mean that a shared concept, a length or a mass for example, should be associated with various parameters defining the concept in each piece of software. Thus the capability to translate the concept into a new one adapted to the use of a specific designer should be offered.

These new studies should lead to an efficient environment for collaborative design in a multi-actor context.

REFERENCES

[1] S.E. POLTROCK, G. ENGELBECK, "Requirements for a virtual collocation environment", *Information and Software Technology*, 1999, Elsevier Science Ltd.

[2] L. ROUCOULES, S. TICHKIEWITCH, “CODE: A Cooperative Design Environment - A new Generation of CAD Systems", in: *Concurrent engineering : research and applications,* vol:8 N° 4, pp.263-280.

[3] N. SHYAMSUNDAR, R. GADH, "Internet-based collaborative product design with assembly features and virtual design spaces", *Computer-Aided Design*, 2001, Elsevier Science Ltd.

[4] D. XUE, Y.XU, “Web-based distributed system and database modeling for concurent design", Proc. of DETC’01, ASME 2001 *Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Pittsburgh, Pennsylvania, Sept. 9-12, DETC2001/CIE-21287

[5] J-F. BOUJUT, E. BLANCO, “Intermediary objects as a means to foster co-operation in engineering design”, Proc. of COOP’2000 Conf. – Workshop 2 ; *The Role of Objects in Design Cooperation : Communication through Physical or Virtual Objects*, Mai 2000, Sophia Antipolis, pp 11-22.

[6] A. XIAO, H-J CHOI, R KULKANI, J K ALLEN, D ROSEN, F MISTREE, "A web based distributed product realization environment", Proc. of DETC’01, ASME 2001 *Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Pittsburgh, Pennsylvania, Sept. 9-12, DETC2001/CIE-21766.

[7] L.A. GUERRERO, D. A. FULLER, "A pattern system for the development of collaborative applications", *Information and Software Technology*, 2001, Elsevier Science Ltd.

[8] B. BELHABIB , F. WURTZ, J. BIGEON, E. ATIENZA, "An Integrated platform for Electrical Design", IEEE-IECON'99, San Jose, CA, USA, November 29th - December 3rd, 1999.

AN INTERNET BASED QUALITY TOOL FOR CUSTOMERS TO IMPROVE PRODUCT DEVELOPMENT

G. Dragoi* and D. Brissaud**

** Polytechnic University of Bucharest, Dept. of Mechatronic and Fine Mechanic, Romania.*

*** Laboratoire 3S, INPG, Grenoble, France*

Abstract: In today's competitive market, customer satisfaction is the critical factor for success. In addition, the concept of continuous improvement means lower costs and higher profits. The integrated system presented in this paper offers a comprehensive solution for turning customer complaints into an opportunity for success. The system manages relationships with customers, consumers, suppliers, distributors and prospects, tracks complaints, requests and activities resulting from these relationships as well as issues and corrective actions resulting from internal investigations, facilitates communication and provides a specific organization with consistent and easy access to information. Through a customizable workflow, an establishment of ownership, due date reporting and automatic escalations, each customer is guaranteed a timely and thorough response. Continuous improvements occur when problems have been identified, root causes determined and solutions found. The Internal Areas are user defined and can be based on departments, production lines, manufacturing facilities, etc. The Internal Areas are generally concerned with process improvement, product defects and quality audits. The system supports embedded keywords for automatic insertion of database information. This system is a client-oriented interface for integrated design quality systems. It has been implemented in a Romanian company.

Key words: Integrated design, cooperative tools, quality system, internet.

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 503–512.

1. INTRODUCTION

Due to the increasing global competition, manufacturing enterprises are forced to work more and more productively in order to survive in the world market. Common to competitive enterprises today, is the quest for designing products and processes to meet a large variety of customers' needs, in a short time, based on few resources. In the new economy [1] of the 21st Century, industrial society is penetrating and evolving in the existing markets of a consumer society with a variety of products intended to improve mobile communications. Many of these products are unique, some are derivatives of existing products, while others are an integration of a number of different products. At the same time, the market demands more and more customized products, resulting in an increasing number of products and variants. A critical aspect in the creation of these products is the understanding of an external user perception [1]. Quantification of this understanding servers as the decision base for the quality properties to be synthesized in the product [2]. Design is the creation of an understanding of a product based on our knowledge [3]. The real word, the model and the decision making are connected together. This is a projection of the real world, the data collected in the model forms the basics for the decision making, and the decisions will, when carried out, give an effect in the real world.

Today's and tomorrow's product development combines with different disciplines to a high degree, i.e. marketing, product design, manufacturing system design, manufacturing, sales, service and after sales [2]. The key is collaborative engineering, a tool for the extended enterprise based on an open information model. The global market increases the need for common and collaborative processes and sharing information seamlessly between companies involved in an extended enterprise [4]. The new requirements are much more focused on collaborative processes that handle product definition data [5]. Partners in the extended enterprise need to exchange legacy data and communicate with other systems outside their own secure corporate boundary [5]. The present approach is to globalize the collaborations on a worldwide scale. The collaboration at such a level is called the extended enterprise level. The objective is to give the opportunity to find technical and financial information but also to link customers to service companies through an e-market place. Companies work in collaborative processes within virtual organization to a large extent which means that information concerning product definition needs to be available, communicated and transferred in a non-homogeneous information system environment. At the same time, each company wants to have secure control of and maintain the integrity of their own data. The development of information systems [4] supporting this process is moving towards more open system architectures and by standardizing the information models describing the information a platform for communication is founded [6]. This activity is performed with different systems for different purposes, e.g. during product design, simulation is used to analyze and predict [3]. Product and process engineers can simulate the product model in order to analyze the effects caused by different processes. Production engineers use discrete event simulation to analyze product realization. Off-line programming is used to make collision tests, reachability tests and to generate machine code.

Today, the three elements of competitiveness are: Timeliness (Can the product be delivered to the market Just in Time consistent with customer's demand?), Quality (Will the product work properly as intended?), Affordability(Do customers wish to buy the product?).

1.1 Objectives of enterprises in the new economy

Concurrent engineering means designing for assembly, availability, cost, customer satisfaction, maintainability, manageability, manufacturability, operability, performance, quality, risk, safety, schedule, social acceptability, and all other attributes of the product. The most recent perception of concurrent engineering is that it includes concurrent design, which is the simultaneous design of the product and its evaluation, prototyping, testing, production, deployment, operation, support, evolution, retirement and management. How much more does it include? The literature is not clear on this point. Recursion could be used to define concurrent engineering. Even though concurrent engineering is not well defined, its use tends to reduce cycle time and cost when appropriately applied. The challenge of the future is to scientifically determine the definition of "appropriately applied." At the moment, each tool or method is responsible for a particular domain feature, which results in an improved quality from that domain's point of view. We propose to develop a system of integration that takes into account the advances of the methods previously seen, keeping flexibility of the activities of design and control of production systems. Using quality methodology [7] generates concepts to reduce negative effects and improve performance of existing product designs. An integrated design system rests on the four views of quality: product view, process view, service view and client view, and includes analytical tools to structure the problem and knowledge base tools to point to the direction of solution concepts. For forty years, quality methods have been explored and improved. The principle is to satisfy the customer who must have the greatest say to the detriment of the manufacturing processes; these methods are external-enterprise based. The integration of the different participants of the design project is realized through work in inter-disciplinary groups in which every participant brings his/her knowledge and constraints to build tables and criteria to evaluate the design solution. The limits of the methods come from the exclusive requirements of the customer in the definition of the production system. The internal aspects of the enterprise are disadvantaged deference to the external aspects due to the customer. Those two aspects have to be well balanced for an optimal system of production to satisfy both the customer and the enterprise.

How can we measure value? Given the above model, we must first identify the quality, which affects individual choices. Next we must understand and model its effects on individual choices as quality varies. The combined effect of many individual choices approximates the value distribution for a market characterized by the individuals and the quality used to define the individual choices. Let us note that individuals may make different choices with the same level of quality because the chosen quality could not include all causes of quality. The result is noisy, and hence it can be characterized by a probability distribution, even when an individual chooses. The constraints of each discipline and each professional are also important to a greater or lesser degree. Product definition, process definition, quality, production management, economical, sociological or organizational constraints are strongly coupled in the development of production systems and cannot be split into different rigid frontiers [3]. The primary focus of Quality Systems is to develop, maintain and audit company-wide systems [4]. This is to assure continuous improvement. This applies to the whole system, not only to production functions. We use many developmental tools to assure that our systems and processes are effective. In each domain concerned with project/product definition, manufacturing processes, production management, etc. many tools and methods have been developed to clarify and improve the engineering work. Those tools and methods have increased productivity and the reliability of products and processes. The informational standards of manufacturing enterprises are based on methods, concepts, tools and techniques using artificial intelligence, which permits us to adapt their activities and also their applications to continuous fluctuations of the economical environment. In fact, if we consider the enterprise not as a monolithic organization but as a set of treatment and information communication systems, we find a more dynamic organization necessary to adapt the enterprise to the constant variability of its economical environment.

2. A TOOL FOR CUSTOMER-FOCUSED QUALITY

With increasing product complexity, customers assign a higher responsibility to enterprises and to their and suppliers and distributors. Manufacturing across industry sectors and geographic boundaries is developing closer links with fewer suppliers. Today, that means many suppliers perform the complete design and manufacturing process of product components. In this case, the supplier designs a product component on the basis of the demands given by the customer. Also, therefore, a suitable description of the customer requirements is an important precondition for trouble-free co-operation between customer and company or suppliers. Society has changed from a closed market and closed manufacturing environment to an open one. It is no longer necessary to have centralized manufacturing facilities. The functions can be distributed. Design, manufacturing, production planning and marketing, part servicing etc. could be done in different places in a country, continent, anywhere on the planet, using network facilities for theexchange of information, services and goods (see figure 1). This model allows for the creation, management, sharing and reusing of electronic information about products and processes in a collaborative environment throughout a product's life cycle. Such a globalization leads to a cross cultural dialog between governments, corporations, societies and, most importantly, individuals [7]. It becomes necessary to model the globalization process and use the global communication networks for a high level of interactivity in cooperative problem solving in manufacturing. In this way, information and communication technologies are the main sources for maintaining global competitive advantage.

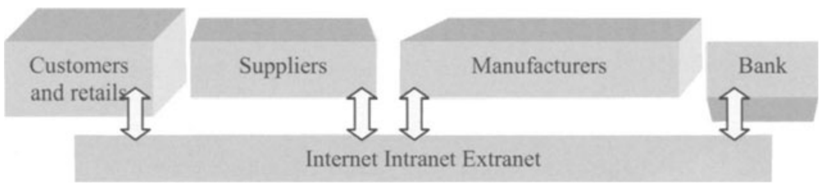


Figure –1. The exchange of Information Services and Goods through Networks

Today, enterprises operate as nodes in a network of suppliers, customers, engineers and other specialized service providers. Virtual or extended enterprises materialize by selecting skills and assets from different firms and synthesizing them into an apparently single business entity. In a real meaning, an e-business is any business that uses Internet or Internet technologies [8] to attract, retain and cultivate relationships to customers, streamline supply chains, manufacturing and procurement systems and to automate corporate processes to deliver the right products and services to customers quickly and cost-effectively, as well as to capture, explore, analyze, and automate corporate processes information on customers and company operations in order to provide better business decisions. We propose an Internet/Intranet/Extranet-based tool for customer focusing on quality [1]. Its structure can be seen in Figure 2.

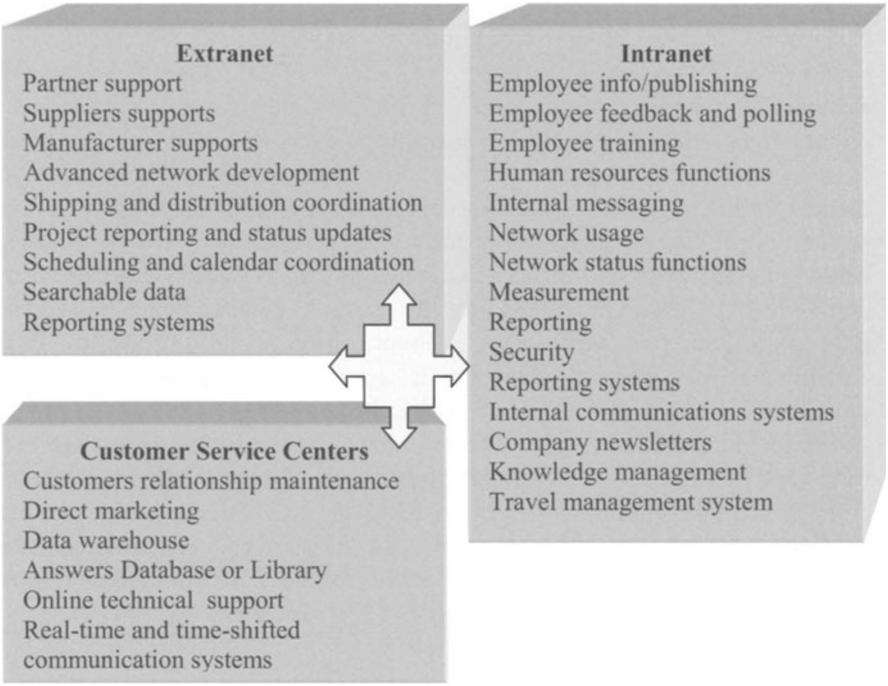


Figure –2. An Internet-Intranet based tool for customer focused quality

The target of the system is customer-focused quality [1]: it manages relationships with customers, consumers, suppliers and distributors; it tracks both complaints, requests and activities resulting from these relationships and corrective actions resulting from internal investigations; it facilitates communication and provides an entire organization with consistent easy access to information [9]. To develop such a tool, the first step consists in identifying customer segments: to select criteria to rank client segments and to rank the client segments. It improves previous designs, which had never been considered possible, and is able to satisfy new demands. It builds a database of customer segments. It allows a more aggressive attitude because of the possibilities offered by technical advances. This step measures the customer perception of the product relative to the competition. Data collected from customers is held in the database and is used as a reference for comparison [7]. The Data-Base helps to verify that the list of requirements is significant for the customer population, capture additional customer requirements, discover strengths and weakness of the product and identify weaknesses in competitors' products. Competitors who should be selected are those successfully offering a similar product or service or those recommended by the customers. Customers can also work jointly to integrate many of your quality programs, so thus improving the strength of your organization and the people working for it. The system has been designed to support and track several types of processes, including employee suggestion programmes, continuous improvement, quality improvement teams (or any related team process), and to monitor team effectiveness using surveys, customer ideas and vendor feedback. Through a customisable workflow establishment of ownership, due date reporting and automatic insertion each customer is guaranteed a timely and through response. Continuous improvements occur when problems have been identified, root causes determined and solutions found. The Data-Base offers customer complaints or request management, comprehensive customer profile with contact management, problem category, product defect and cause analysis, internal and external corrective action tracking, return authorization and field service tracking. The database features are based on full web enablement, letter and email integration, attachment of related documents, used defined queries, extensive reporting and charting, reminder, past due and escalation notices, high flexibility and scalability, etc. The application overview presents a user-friendly interface and is easily navigable. The menus can be accessed through the browser style buttons on the left side or can be accessed through the windows menu on the top. The menu can be customized to match local terminology and workflow. Within the Setup Menu, the infrastructure of the entire system has been defined. This is where terms can be re-named, views customized and all drop-down lists defined. The account types are user defined and may include: customers, consumers, suppliers, distributors and others. The database relationships are presented in figure 3. Each Account View is composed of contacts and concerns. Their content can be customized to reflect local terminology and work process. Internal areas are concerned with process improvement, product defects and quality audits and are user-defined and can be based on departments, production lines, etc.

Concerns begin with a problem definition tab (optional tabs include: reference, products, returns, repairs, action requests, documents, final resolution, etc.). The optional tabs include: activities, categories, comments, documents, equipment, etc. Action requests drive the problem resolution associated with each concern; types of actions can be defined, then tracked for timely completion. A corrective action request follows ISO guideline and expects a full investigations and comprehensive reply: problem statement, containment actions, findings, root causes analysis, preventive actions, permanent corrective actions, and follow-up verification. The system includes an automatic insertion of database information. A complete set of charts is on hand to visually represent problem areas, trends and improvements.

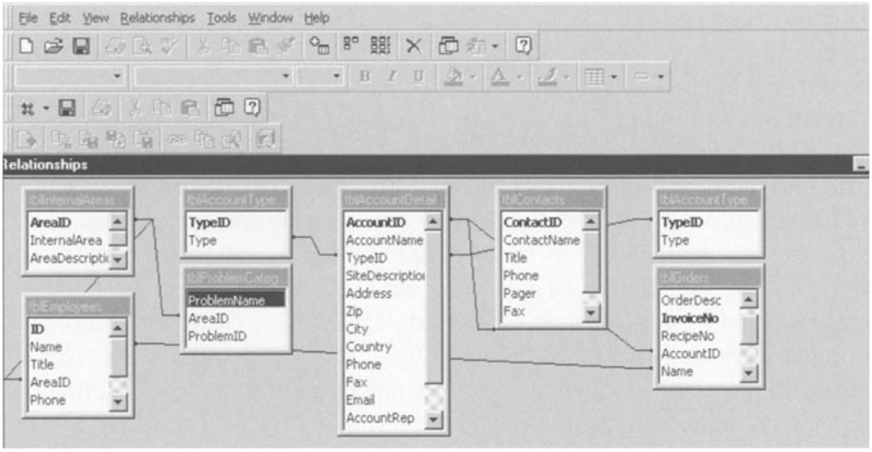


Figure –3. Data-Base relationships

The first step consists in defining who the customers are, and then establishing the customer's needs and demands by direct dialogue. This is the basis to follow for each session's work and it must be done without distorting the customer's desires. If this step seems trivial, let us consider the business model of today's large corporation compared which a small production company. For the large corporation, the development and production engineers and staff have little or no contact with the customer. The customer's original requirements have been filtered through the marketing department, before being deliverd to the development staff, and then to production. In many corporations, the development staff itself is a hierarchical structure. The designers of the product are far removed from the customer. There is a real need to ensure the customer's needs so that the "Voice of the Customer" is not misinterpreted or distorted. The voice of customer's segments and the context of use [7]: The customer's needs are sorted into required quality, reliability issues, solutions, functions, failure modes and safety. All this information must be implemented in databases. The voice of customer context table should add the system resources to the Who? What? Where? When? How? Information. In today's competitive market, best quality and lower costs to give customer satisfaction are critical success factors. For businesses, continuous improvement means lower costs and higher profits. This solution offers a comprehensive solution for turning customer complaints into an opportunity for success. The final verdict on the success of a company's efforts at meeting customer-demanded quality is long-term market acceptance of the resulting product. If the methodology has been properly and thoroughly applied, the statistical process control procedures chart the company's ability to meet customer-required quality. Out of control processes are then indicative of a loss in perceived quality. Quality loss means direct financial loss to the company resulting from providing sub-optimal products to the customer. Future market shares are reduced, and warranty claims mount. If customer quality requirements are fully deployed to production, cost can be minimized. Producing at the quality target value is considered free.

3. CONCLUSIONS

We propose a methodology to develop a system of integration that takes into account advanced quality methods and tools keeping the flexibility necessary for the activities of design and control of production systems. This model has been jointly developed at Polytechnic University of Bucharest and INPG at Grenoble. The project is supported by the Research Minister of Romania and is implemented in business. The model follows the general approach of Concurrent Engineering and considers manufacturability, cost and quality at the product design stage. The combined use of the tools to evaluate products provides a systematic approach for identifying where design quality can be improved at the conception stage (virtual product). Benefits of the system include shorter development time and reduced time to market, reduced engineering change, rapid comparison of alternative design and competition products, systematic component costing and lower component cost. The target idea is to create a standardized, open-systems environment and infrastructure that leverages common platforms to maximize speed and access to information, to allow flexibility and rapid development of specific end-user applications and streamline the integration of commercial and custom applications across the business. Supporting customer value driven steps are: open communication channels with customers, engagement in joint strategy and planning sessions with operational groups, movement to a client-based consultative relationship and the ability to listen to customers and be open to change to educate customers in technology in terminology they will understand, Join project teams, be an active member of reengineering projects. It could be possible to recruit IT resources with customer service skills, to decentralize IT , move resources out to operational departments or rotate IT resources into operational positions and measure customer service performance. Customer value driven steps leading to innovation are: research and track emerging technologies, subscribe to leading-edge technology journals and attend technology conferences on a regular basis, know the business needs and operational processes thoroughly, align technology capabilities with business needs and goals, attend business meetings and operational meetings on a regular basis to stay current on business needs and issues, and to migrate resource profiles to provide more technology and business analysts and fewer programmers. The strategy to create open and standardized platforms and infrastructure [7, 8] requires us to: invest in technology infrastructure to support robust networking and rapid access to information, design and develop an open-systems environment, establish standard protocols and interfaces that allow easy integration of systems and applications across the business, and create platforms and networks that allow flexible application development for end-users.

REFERENCES

[1] G. DRAGOI, D. BRISSAUD, An Internet–Intranet-based tool for Customer Focused Quality Fifth Conference on Management of Innovative Technologies, 11-12 *October 2001,* Piran, Slovenia.

[2] G. DRAGOI, D. BRISSAUD, Cooperative design system: an approach to multidisciplinary optimization, *The Romanian Review of mechanic of precision, optics & mechatronic*, Bucharest, Romania, 2000, nr.18/2000, pp. 1861-1865, ISSN 1220-6830.

[3] S. TICHKIEWITCH, B. RADULESCU, G. DRAGOI, Specific task for a co-operative design support tool, International CIRP Design Seminar Design in the New Economy, 6-8 June 2001, pp. 97-103.

[4] G. DRAGOI, *Intreprinderea convergenta. Informatica Industriala*, Editura PRINTECH Bucuresti, 2001, ISBN 973-652-293-8.

[5] G. DRAGOI, D. BRISSAUD, L'entreprise étendue : le système informationnel, une approche vers l'intégration et le travail collaboratif, in *Modélisation de la connaissance pour la conception et la fabrication intégrées*, Ed. Mirton Timisoara, 2000, pp. 175-202, , ISBN 973-585-216-0.

[6] C. WHINNEIN, R. INHOF, "*Corporate Information Factory",* John Wiley Computer Publ. 1998

[7] G. DRAGOI, D. BRISSAUD, A methodology for cost and quality optimization in a design system by linking quality methods, in *Integrated Design and Manufacturing in Mechanical Engineering'2000*, Kluwer Academic Press, 2002.

[8] G., GARDARIN, *Internet/Intranet et bases de données*, Edition Eyrolles, 1999.

[9] 3COM: Tehnical Paper, Converged networks-case studies, 500671-002-1999.

EMERGENCE OF A VIRTUAL HUMAN BEHAVIOUR IN A VR ENVIRONMENT

A Multi-Agent Approach

B. Maillé*, P. Chedmail** and E. Ramstein*

** Snecma Moteurs, Moissy Cramayel, France*

*** IRCCyN, Ecole Centrale de Nantes, France*

Abstract: Nowadays, time to market becoming shorter and shorter, physical mock up is being more and more often replaced by digital mock up. However physical mock up was very useful to the present day to check accessibility for example. So, it is now important to have tools available to simulate with digital mock up what physical mock up used to be used for. For example, to check accessibility, a tool that allows the simulation of the manipulation of an object by a manikin in a cluttered environment without collision is needed. In this paper, we propose a new way to create human behaviour of a manikin using a multi-agent approach. Our system is based on a former architecture already presented. It consists in a collision-free path planning system using co-operation between software and human operator abilities. The architecture for this co-operation is based on a multi-level tree architecture, which enables the architecture to be completed by adding as many agents as we need. The conflicts between agents are managed by acting on their period of activity. We demonstrate in this paper the fact that collaboration between many agents, each providing the manikin with an elementary behaviour, can provide the manikin with a coherent global virtual human behaviour.

Key words: multi-agent, virtual reality, virtual human, behaviour

1. INTRODUCTION

Path planners for mobile objects in cluttered environments have been developed in the framework of robotics during the past 20 years. With the rising use of digital mock up, virtual reality tools are increasingly used

G. Gogu et al. (eds.),
Recent Advantages in Integrated Design and Manufacturing in Mechanical Engineering, 513–522.

within the industrial framework, allowing accessibility or maintainability checks between mechanical parts without physical mock up. For this, it is necessary to simulate an object moving without collision. We couple software abilities with human operator capabilities in order to use a human's global view ability. Our architecture for this co-operation is based on multi-agent principles.

In the next two paragraphs, we successively describe the state of the art for accessibility checks and our multi-agent architecture.

In paragraph 4, we demonstrate the possibility of an emergence of global virtual human behaviour by using collaboration between different agents.

2. STATE OF THE ART FOR ACCESSIBILITY CHECKS

For assemblability, accessibility or maintainability checks two main methods are usually proposed: automatic path planning approach and simulation through robotic or virtual reality tools.

2.1 Automatic path planners

Collision-free trajectory research approaches are mainly based on two principles. Some methodologies require a global perception of the environment ([13], [15], [2]) while others consider the objects movements only in its close or local environment ([11], [1]). All these techniques are limited either by the computation cost or the existence of local minima as explained by Namgung [14].

For highly cluttered environments, it can be useful to be helped by human global view of the environment. Hwang [8] has settled for such an approach: the designer can define a sequence of sub-goals in the environment leading to a partition of the problem into several simpler path-planning problems. This approach is a sequential mix between figures 2 and 3. Kavraki and Latombe [10] propose a similar approach by placing sub-goals defined in a randomized manner. These methods confirm that human abilities or a stochastic approach greatly enhance path planners.

2.2 Accessibility with direct simulation

Another approach is based on robotic CAD systems that allow the designer to manipulate robots, mechanisms or even manikins. For this approach, the designer uses his or her global perception to achieve the task, but this manipulation is usually sequential and long. Classical virtual reality tools enhance the robotic on-line manipulation with the use of immersion devices. Some drawbacks are the lack of kinematic constraints and the fact that collisions between objects are not automatically avoided, although they are usually detected.

2.3 Conclusion

The high potential of automatic path planners and direct simulations has been pointed out. Nevertheless, the first algorithmic approaches are long and lead to local minima. The second approaches with on-line manipulations are too long with robotic CAD systems or require a high number of expensive devices for virtual reality immersion. In order to avoid these drawbacks, we intend to settle for a mixed approach of the presented methodologies using local algorithm abilities and global view ability of a human operator.

3. ACCESSIBILITY AND MAINTAINABILITY CHECK WITH A MULTI-AGENT SYSTEM IN VIRTUAL REALITY.

3.1 Considered multi-agent architecture

The above chapter points out clearly the local abilities of several path planners and the global abilities of human vision. We intend to manage simultaneously these local and global abilities. The main purpose is to know how to build on interaction between human and algorithms in order to have an efficient path planner [3]. For this, we have built a path planner based on multi-agent principles proposed by Ferber [7] with a human interactive integration between algorithmic processes rather than on a sequential integration as proposed by Hwang [8].

For our purpose, we have no direct communication between agents. This kind of architecture can be compared to "blackboard" systems described in [7], [6] and [9]. Each agent has a period of activity, which expresses priority between the agents and a main process collects elementary agents' contributions (normalized by a predetermined step) and updates the database.

Figure -1 is composed of different basic "blackboard" levels, which allows a partition dividing of our software construction, and enables easy addition of supplementary agents. Our multi-agent system is built with this kind of hierarchical architecture.

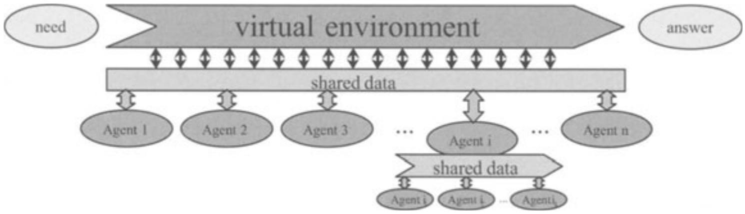


Figure -1. Blackboard principle with co-operating agents

The purpose of the following paragraphs is to recall our new CAD/CAM solutions based on a distributed approach using a virtual reality environment. The above chapters clearly point out the abilities of several local path planners and of the human global vision. Consequently, we try to manage the latter simultaneously, local and global, abilities. We build a coupled approach using multi-agent and distributed principles. These principles were defined by Chedmail and accurately described in [3].

3.2 Object manipulation.

For object manipulation we use former principles defined by Chedmail and Le Roy in [4]. Hereafter we summarise this approach for elementary agents' definition, these agents are represented in *Figure -2*.

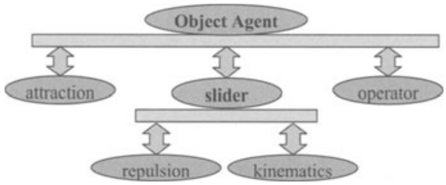


Figure -2. Object Agent distribution

The elementary agents that compose the **Object Agent**, used for object manipulation, are:

- Attraction agent: this agent affects the position and orientation of the manipulated object with elementary moves toward its target.

- Slider agent: this agent, composed of two elementary agents, acts in order to nullify the collisions between the manipulated object and the cluttered environment, by acting on the position and orientation of the object (**Repulsion agent**) and on its internal degrees of freedom (**Kinematics agent**).

- Human operator, operator agent: this agent allows the human operator to act on the trajectory in real time.

3.3 Manikin manipulation.

The multi-agent approach for our software architecture for manikin driving leads to a distribution at different multi-agent levels (*Figure -3*).

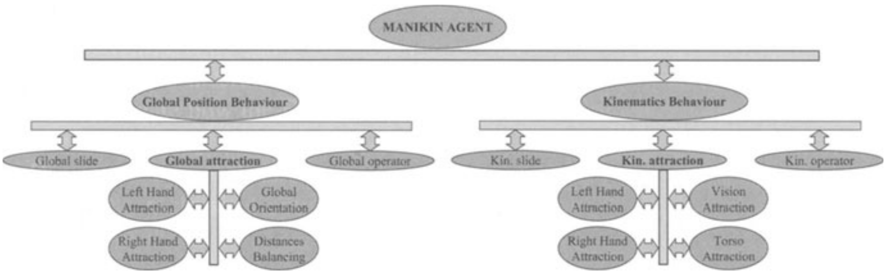


Figure -3. Manikin Agent distribution

The architecture for manikin manipulation is composed by many agents that can act either on the global position of the manikin or on its internal degrees of freedom. These agents are designed thanks to a functional distribution.

The agents used by the *manikin* agent are distributed in two main agents:

- The ***Global Position Behaviour*** agent manages the global position of the manikin. All the agents composing this one works in order to achieve their goal by acting on the global position of the manikin.
- The ***Kinematics Behaviour*** agent manages the internal degrees of freedom of the manikin. All the agents composing the latter work in order to achieve their goal by acting on the internal degrees of freedom of the manikin.

The goal of each agent is detailed hereunder:

- The ***Slide*** agents' goal is to make the manikin avoid collision and slide on the environment.
- The ***Operator*** agents' goal is to allow the action of the operator.
- The ***Hand attraction*** agents' goal is to make the hands reach their target.
- The ***vision attraction*** agent's goal is to make the manikin look at the manipulated object.
- The ***Torso Attraction*** agent's goal is to make the torso be oriented towards the manipulated object.
- The ***Global Orientation*** agent's goal is to make the manikin be oriented towards the manipulated object.
- The ***Distance Balancing*** agent's goal is to balance the distances between the hands and their respective target.

3.4 Conflict management

To give the object or the manikin coherent behaviour and to avoid collisions we have defined priorities between agents. This is also necessary because some agents can act on the same degrees of freedom. The relative priority of each agent is defined by its period of activity, and its contribution is required only at this period of activity. So, the greater the period of activity, the less important the agent is.

For example for an object: if the *slider* agent's period of activity is 1, the *attraction* agent's is 8, and the *operator* agent's is 4, then the *operator* agent allows the operator to act on the trajectory but cannot force the manipulated object to collide. By acting on these periods of activity, we modify the global behaviour of the manipulated object or manikin. Thus, the choice of the periods of activity is very important and depends on what behaviour we want to obtain for the manipulated object or manikin.

3.5 Global multi-agent architecture.

While gathering the different systems for object and manikin manipulation we obtain the multi-level diagram of our global software architecture, composed of several multi-agent systems that are able to resolve local constraints. This architecture enables the generation of a collision-free path for a manikin manipulating an object.

This multi-agent architecture allows the manipulation of as many objects and manikins as we want, one *Object Agent* being used for each object, and one *Manikin Agent* for each manikin.

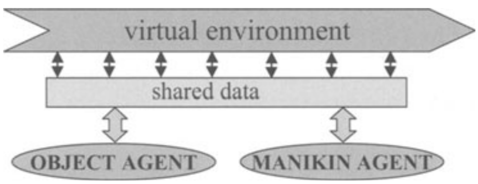


Figure -4. Architecture for one object and one manikin manipulation

4. BEHAVIOUR EMERGENCE

We present hereafter some experiments performed with different combinations of agents for reaching one or two targets in an environment with no obstacle. These examples demonstrate that the activation of different agents leads to different forms of simulated human behaviour.

4.1 One hand manipulation

In the first experiment (*Figure -5*), the hand and the eyes of the manikin are attracted towards a given target. In the second experiment (*Figure -6*), out of the former ones, we have activated global attraction on the right hand and on torso orientation. We insist on the fact that the camera orientation is constant and identical in these two experiments.

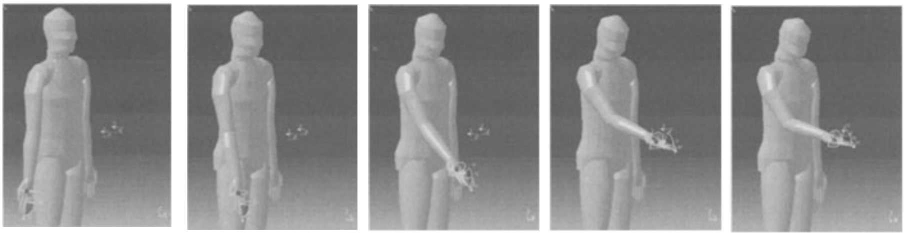


Figure -5. Manikin with kinematic attraction on right hand and vision

Figure -6. Manikin with kinematics attraction on right hand and vision, global attraction on right hand and global orientation attraction

One can note the difference between the postural attitudes of the manikin during and at the end of each experiment. The second seems more "natural" than the first one.

4.2 Two hand manipulation

In this second sequence of manipulations, both hands are attracted towards two different targets. The first experiment (*Figure -7*) takes into account only the kinematic hand attraction agents. The following experiments integrate respectively the global hand and torso orientation attraction agents (*Figure -8*), the distance balancing and vision attraction agents (*Figure -9*) and the kinematic torso attraction (*Figure -10*).

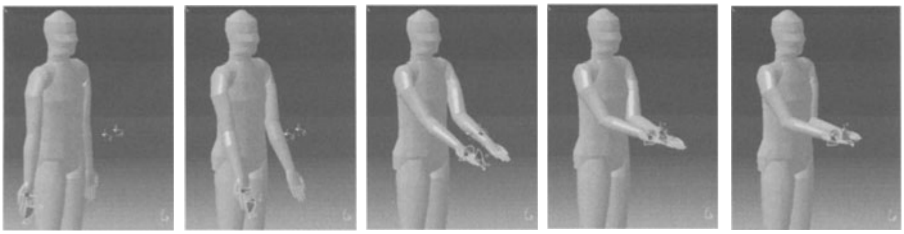


Figure -7. Manikin with kinematic attraction on both hands

Figure -8. Manikin with global and kinematic attraction on both hands, and global orientation attraction

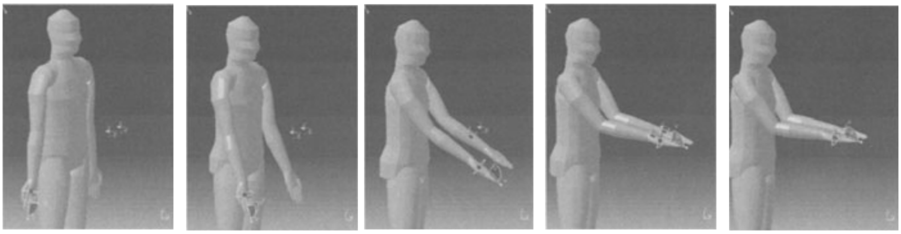


Figure -9. Manikin with global and kinematic attraction on both hands, global orientation attraction and distances balancing, and kinematics attraction for vision

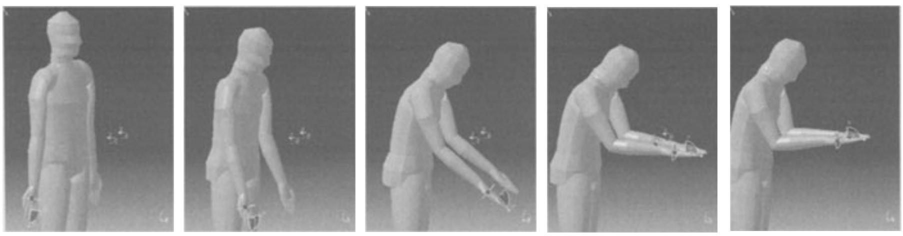


Figure -10. Manikin with global and kinematic attraction on both hands, global orientation attraction and distances balancing, and kinematic attraction for vision and torso

Figure -11 hereunder presents the top view of the final positions for the different experiments.

Figure -11. Top views of a) the initial position and of the final positions of b)*Figure -7*, c)*Figure -8*, d)*Figure -9*, e)*Figure -6*

4.3 Comments on these experiments

These experiments clearly demonstrate the modification of behaviour when we introduce more and more agents. The global behaviour of the manikin becomes more and more "natural".

Most agents modify the global behaviour (for example torso attraction or global attraction modify the position of the shoulders and head). They affect the contribution of nearly all agents. A few agents, such as vision attraction, only affect local behaviour. The activity of the latter kind of agent has little effect on the global behaviour, contrary to the first kind of agents, whose relative priorities have a great effect on the global behaviour and have to be correctly managed.

In other experiments, we have demonstrated that in a cluttered environment, our system is able to make the manikin avoid collisions and slide on objects. It is also possible to act on the trajectory of the body or of the hands during such path generation.

5. CONCLUSIONS AND FURTHER WORK

Our architecture enables object driving with a manikin. Moreover, we manage to couple heterogeneous constraints such as collision avoidance, human operator contribution and manikin behaviour in the same architecture.

A great advantage of this architecture is that it enables easy addition of supplementary agents. Future work will strongly lean on this ability: we intend to add some simple *ergonomics* agents to provide the manikin with an ergonomic behaviour and other agents to manage the legs and the hands of the manikin.

As a conclusion, we have demonstrated that the addition of complementary agents to manage elementary behaviour of a real human can lead to a coherent global behaviour for a virtual human, thus allowing easy and accurate validation of accessibility and maintainability of mechanical parts in an assembly.

ACKNOWLEDGEMENTS

We address our acknowledgements to the Dassault Systems Company, and particularly to Jean-Luc Cuinier, for his collaboration and technical support on this project.

REFERENCES

[1] Barraquand, J., and Latombe, J.C., 1991, "Robot Motion Planning : A Distributed Representation Approach", *The International Journal of Robotics research*, December 1991.

[2] Boissonnat, J.D., Faverjon, B., and Merlet, J.P., 1988, "Perception et planification", Traité des nouvelles technologies, Série Robotique. Hermès, Paris, France, 1988.

[3] Chedmail, P., Damay, T., and Rouchon, C., 1996, "Integrated design and artificial reality : accessibility and path planning", ASME Engineering System Design and Analysis (ESDA'96), vol. 8, pp. 149-156, ISBN 0-7918-1503-X, July 1-4, 1996, Montpellier.

[4] Chedmail, P. and Le Roy, C. "A Distributed Approach For Accessibility And Maintainability Check With A Manikin". DAC-8677, ASME Design Engineering Technical Conferences, Las vegas, Névada, USA, September 12-15, 1999.

[5] Chedmail, P. and Le Roy, C. "Multi-agent architecture for an interactive path-planner". Advances in computational multibody dynamics, IDMEC/IST, Lisbon, Portugal, September 20-23, 1999.

[6] Ferber, J., and Ghallab, M., 1988, "Problématique des univers multi-agents intelligents", Actes des journées nationales PRC-IA, 1988, pp.295-320.

[7] Ferber, J., 1995, "Les systèmes multi-agents, vers une intelligence collective", InterEditions, 1995.

[8] Hwang, Y. K., Cho, K. R., Lee, S., Park, S. M., and Kang, S., 1997, "Human Computer Cooperation in Interactive Motion Planning", ICAR'97, Monterey, CA, July 7-9, pp. 571-576, 1997.

[9] Jennings, N. R., Sycara, K. and Wooldridge, M., 1998, "Roadmap of agent research and development", Autonomous agents and multi-agent systems, Vol. 1, pp. 7-38, 1998.

[10] Kavraki, L. E., Latombe, J.-C., Motwani, R. and Raghavan, P. Randomized Query Processing in Robot Path Planning,1996.

[11] Khatib, O., 1986, "Real time obstacle avoidance for manipulators and mobile robots", International Journal Robotics Research, Vol. 5, no. 1, pp.90-98, 1986.

[12] Le Roy C., "Accessibilité et aptitude au montage, réalité virtuelle et approche multi-agent", thèse de doctorat de l'Université de Nantes, 5 janvier 2000.

[13] Lozano-Pérez, T., and Wesley, M. A., 1979, "An algorithm for planning collision-free paths among polyhedral obstacles", Comm. ACM, 22 (10), pp. 560-570, 1979.

[14] Namgung, I., and Duffy, J., 1997, "Two Dimensional Collision-Free Path Planning Using Linear Parametric Curve", Journal of Robotic Systems 14(9), pp. 659-673, 1997.

[15] Tournassoud, P., 1990, "Géométrie et intelligence artificielle pour les robots", Hermès, Paris, France, 1990.

AUTHORS INDEX