Shape and Topology
Optimisation for Manufactured Products

Frank Mill

Ph.D.
The University of Edinburgh
2003
ABSTRACT

The work described represents the candidate’s research in the area of engineering component representation, particularly in the context of constrained optimisation. Early papers describe a general approach to detail design that has since become commonplace, however, the advances that were hoped for in some downstream applications are proving more elusive. Whilst the early work describes constraint generation later work looks at different approaches to the shape and topology optimisation problems. The latest papers also include part of the candidate’s work in considering specific applications of parameterised manufacturing features.
DECLARATION

I declare that this critical review has been composed by myself and is all my own work except where otherwise stated. Furthermore I declare that none of the material contained in the review has been submitted in full or in part for the award of any other degree.

Signed,

Frank Mill

Edinburgh, September 2003
ACKNOWLEDGEMENTS

Many thanks are due to my co-workers on the research described in this thesis. All of the work that is described was undertaken as part of a team and therefore I have benefited from their discussions and knowledge of their work. I hope I have represented this accurately and referenced this diligently throughout the text.

I would particularly like to thank David Clifton, Jane Naish and Andrew Sherlock for carrying out their research work and contributing to open discussion on many of the areas covered in this thesis, thus blurring the distinction between supervisor and student.

I would also like to say a big thanks to my family for putting up with me during the very long time it has taken me to produce this document.
FOREWORD

The work presented in this critical review is related to the candidate's research in the field of part representations for manufactured products, particularly in design and tooling selection and optimisation.

The structure of the review is largely dictated by the Edinburgh University Postgraduate Study Programme Regulations, Part 3.11 Ph.D. by Research Publication. As a result the review consists of a section on each paper covering the aims, objectives, results and conclusions together with a critique looking at the paper with the benefit of hindsight. The candidate's role in each of the papers is described in the methodology or critique for each paper.

In all, eight papers are presented and these were chosen as a representation of the candidate's research work during the ten year period of the review. Several other papers were published which were not felt to be immediately relevant and some that were, were also written but published outside the timeframe for this document.

Some additional references are given at the end of the review where these are relevant to the candidate's on-going work but the majority of referenced material is to be found in the papers themselves.

References in the critique itself are given in a tagged Harvard style format. Each reference is represented in the text by a 6 digit alphanumeric tag which is also used to order the list of additional references at the end of the review. References to the papers submitted as part of this review are shown throughout the text in underlined format.
CONTENTS

Abstract ..........................................................................................................................2
Declaration .....................................................................................................................3
Acknowledgements ........................................................................................................4
Foreword ........................................................................................................................5
Contents .........................................................................................................................6
Glossary .........................................................................................................................9

1 Introduction ..........................................................................................................10

2 Feature Oriented Design ......................................................................................13
  
  2.1 Representation problems in Feature-based approaches to design and process planning (MILL93) .................................................................................................. 14
    2.1.1 Aim ......................................................................................................14
    2.1.2 Objectives ............................................................................................14
    2.1.3 Methodology ........................................................................................14
    2.1.4 Results and conclusions .......................................................................14
    2.1.5 Critique ................................................................................................15
  
  2.2 Design for Machining with a simultaneous-engineering workstation (MILL94) .................................................................................................................18
    2.2.1 Aim ......................................................................................................18
    2.2.2 Objectives ............................................................................................18
    2.2.3 Methodology ........................................................................................18
    2.2.4 Results and conclusions .......................................................................18
    2.2.5 Critique ................................................................................................19

  
  2.3 An Industry-based study of cutting process capability representation requirements for an integrated simultaneous engineering workstation (NAIS97) ..22
    2.3.1 Aim ......................................................................................................22
    2.3.2 Objectives ............................................................................................22
    2.3.3 Methodology ........................................................................................22
    2.3.4 Results and conclusions .......................................................................22
    2.3.5 Critique ................................................................................................23
2.4 Features as Autonomous Agents: An Alternative Paradigm for Concurrent Engineering (JACQ 97) ...............................................................25

2.4.1 Aim ......................................................................................................25
2.4.2 Objectives ..........................................................................................25
2.4.3 Methodology .....................................................................................25
2.4.4 Results and conclusions ...................................................................25
2.4.5 Critique ..............................................................................................26

2.5 Further comment and discussion ...........................................................28

3 Biological Analogies ..................................................................................37

3.1 A voxel-based representation for evolutionary shape optimisation (BARO 99) ....................................................................................................................38

3.1.1 Aim ...................................................................................................38
3.1.2 Objectives ..........................................................................................38
3.1.3 Methodology .....................................................................................38
3.1.4 Results and Conclusions ..................................................................39
3.1.5 Critique ..............................................................................................39

3.2 Biological analogies in manufacturing (MILLO 00) ....................................42

3.2.1 Aim ...................................................................................................42
3.2.2 Objectives ..........................................................................................42
3.2.3 Methodology .....................................................................................42
3.2.4 Results and conclusions ....................................................................42
3.2.5 Critique ..............................................................................................43

3.3 Further comment and discussion ...........................................................46

4 Machining Applications ..............................................................................51

4.1 Diverless weld inspection and repair using ECM/ACFM techniques (CLIF 00) ..........................................................52

4.1.1 Aim ...................................................................................................52
4.1.2 Objectives ..........................................................................................52
4.1.3 Methodology .....................................................................................52
4.1.4 Results and conclusions ....................................................................53
4.1.5 Critique ........................................................................................................... 53

4.2 A direct analytical solution to the tool design problem in electrochemical machining under steady state conditions (ALDE00) ............................................................... 55

4.2.1 Aim ................................................................................................................. 55

4.2.2 Objectives ........................................................................................................ 55

4.2.3 Methodology .................................................................................................... 55

4.2.4 Results and conclusions .............................................................................. 55

4.2.5 Critique ......................................................................................................... 56

4.3 Further comment and discussion ..................................................................... 57

5 Conclusions ............................................................................................................ 58

6 List of papers .......................................................................................................... 59

7 References .............................................................................................................. 60

Appendix A – MILL93 ......................................................................................... A-1

Appendix B – MILL94 ......................................................................................... B-1

Appendix C – NAIS97 ......................................................................................... C-1

Appendix D – JACQ97 ......................................................................................... D-1

Appendix E – BARO99 ......................................................................................... E-1

Appendix F – MILL00 ......................................................................................... F-1

Appendix G – CLIF00 ......................................................................................... G-1

Appendix H – ALDE00 ......................................................................................... H-1
GLOSSARY

Author - throughout the text the term author is used to describe the authors of the various sources of references, including the papers that are the subject of, this critical review.

CAD – Computer Aided Design
CAE – Computer Aided Engineering
CAM – Computer Aided Manufacture
CAPP – Computer Aided Process Planning

Candidate – throughout the text the term candidate is used to describe the author of this critical review.

CNC – Computer Numerical Control
CODL – Component Description Language
DFX – Design for X
DXF – Drawing Exchange Format
FMS – Flexible Manufacturing System
FOD – Feature Oriented Design
FODDS - Feature Oriented Detail Design System
FODL – Feature Oriented Design Language
IGES – Initial Graphics Exchange Scheme
IMS – Intelligent Manufacturing Systems
KBS – Knowledge Based System
PDM – Product Data Management
PLM – Product Lifecycle Modelling
PPDL – Process Planning Description Language
SEW – Simultaneous Engineering Workstation

STEP – Standard for the Exchange of Product Model Data
1 INTRODUCTION

The work described in this critical review relates to a ten year period during which many problems in the representation of engineered products have been investigated by the research community. The primary intent of the work described herein is in the representation of part designs in relation to the needs of manufacturing tooling. A secondary aspect of the work relates to the investigations in optimisation of part shapes and correspondingly to tool shapes.

Initial work by the candidate began in 1983 with an investigation into the possibility of employing CAPP for a small scale FMS. At that time the proposed solution offered by the candidate and many other researchers lay in the development of large knowledge based systems which would be programmed to 'know' how to machine single parts. It was already evident at that time that such systems would decompose parts into machining elements that would have machining methods associated with them, perhaps including programmed routines such as canned cycles on CNC machines (see e.g. MILL84).

Such an approach worked for very simple part manufacture, i.e. where part representations could be easily coded to allow retrieval of elemental or sub plans. However, there were a great many problems encountered when knowledge based CAPP systems were applied to almost any non trivial parts. Although a robust definition of part complexity was never made explicit the reasons for a part being considered so by researchers arose from a few simple considerations.

Firstly, the descriptions of the part elements themselves were difficult to obtain. The candidate's early work involved the use of extended boundary representations of features consisting in turn of groups of geometric primitives with their associated loops, edges and vertices (HUSB90). These high level features could also hold information on dimensional tolerances within features.

The second major problem arose, however, from the fact that the complexity in most planning environments was not due to the primary features themselves but in the fact that there were relationships between the features on a part. The simpler types of relation to represent were termed explicit relationships. The simplest of these were dimensional tolerances linking one or more features together. A more complex form of tolerance relationship was seen in the use of geometrical tolerances which could
relate large numbers of features together. The system devised by the candidate allowed for both types of these tolerances to be represented although only simple geometrical ones to be coded. An even greater level of complexity was needed to describe implicit relationships. These were impossible to define fully because they could not be exhaustively catalogued. Relationships of this type included such things as feature overlaps or thin walls between features.

Although the candidate’s work produced a language that could represent complex feature based descriptions of parts, this and others’ work showed up a myriad of problems that would have to be solved before these descriptions would have any practical use.

The first major problem that had to be addressed was how the feature based descriptions would be generated in the first place. Early research, such as that described above, made use of text descriptions which would be impossible to generate manually for anything other than the types of parts that were used as test pieces by the research community. It was clear that for CAPP research to make any progress part descriptions would have to be generated automatically from CAD systems. Two approaches to this problem became popular with researchers, FOD and Feature Recognition. The candidate’s own work and that of his research group was centred around the FOD approach and the first four papers, (MILL93, MILL94, NAIS97, JACQ97), offered as part of this critical review describe work done in identifying problems for a test product, developing a FOD interface and using features as handles for the storage of machining know how. These are discussed further in Chapter 2 of this critical review.

The second problem that needed addressed was how the feature descriptions would be used. Process and other design and planning refinement were time consuming tasks that were seen to be open to enormous benefits if they could be automated or partially automated. Early work by the candidate and others showed that these problems often exhibited enormous search spaces and that to solve these with knowledge bases would be computationally very expensive. The candidate proposed a CAPP system whereby a knowledge base would be used to trim the search space until the remaining sub-problem could be solved by optimisation methods (HUSB90). The methods researched included ones influenced by biological analogies and these found applications in wider areas of design and planning. The two papers discussed in
Chapter 3 of this review (BARO99, MILLO00) describe a sample of the work undertaken by the candidate and his research group in this area.

Chapter 4 discusses papers presented in the review (ALDE00, CLIF00) which look in some depth at specific machining problems which are related to tool shape design for complex feature shapes in the context of electro-chemical machining. Throughout the review, sections on each paper cover a description of each papers aim, objectives, methodology and main conclusions as required by postgraduate regulations and each section also carries a critique of the paper together with a description of the candidate’s role in it where this is not described in the methodology section.
2 FEATURE ORIENTED DESIGN

Following early work in feature based process planning and the problems that became evident as its result, a more complex means of plan development was proposed. Given that any new representation would have to deal with complex feature interactions, design systems would have to be able to represent these where they were made explicit by the designer or infer them using geometrical reasoning where these were implicit.

The work addresses the stage of representing and constraining the search space of designs in the context of machining environments.
2.1 REPRESENTATION PROBLEMS IN FEATURE-BASED APPROACHES TO DESIGN AND PROCESS PLANNING (MILL93)

2.1.1 Aim

The aim of the paper was to describe to the wider research community some of the problems in adopting simple feature-based approaches for part descriptions in a manufacturing context.

The paper was motivated by the candidate's view that much of the research of the time was dedicated to the study of simple components that had simple independent self-contained features. These were relatively easy to represent but did not allow any analysis of likely manufacturing problems nor did they help in addressing downstream manufacturing applications.

2.1.2 Objectives

The specific objectives of the paper were:

• to present a useful set of specific problems that could be used as a benchmark for researchers to use in assessing each other's work,

• to present problems that researchers could use in the design of their methods for representing components, and,

• to show the importance of feature interactions.

2.1.3 Methodology

A constraint on these objectives was that each problem presented would be one that had been found on a real component. The features presented were those collected by the candidate in discussions with industrialists and other researchers. These were then designed into a single composite component by the candidate.

2.1.4 Results and conclusions

'Feature oriented design and feature recognition are insufficient on their own for the complete integration of CAD and CAM'

'Distinction must be made between different feature views of products, e.g. design features and production features.'
‘Most difficulties in the production engineering of products are due to interactions rather than the features themselves’

‘The feature interactions must be satisfactorily modelled if true generative process planning is to take place.’

‘Alternative feature descriptions can be of use for CAPP. Their representation can be achieved in many cases. It is probably worthwhile dealing with those cases that can be analysed easily whilst recognising and flagging more complicated cases.’

2.1.5 Critique

As part of the work in this area the candidate carried out survey work, studying real component parts from a range of companies and industries. In an attempt to distil the common feature relationships that were found into a manageable problem definition, a test part was created that would help guide future development of his group’s feature oriented design system. The test component is described in MILL93.

In common with many papers in this field and from that time, a considerable section of the text is spent explaining the justification of the approach taken. The background of CAPP work through the HAPPI system (HUSB90) and through the context of the engineering firms who helped with the work meant that the candidates’ view was based around environments where detail design was the main design activity undertaken and manufacturing facilities were usually general in nature and predominately in house. Most such businesses operate in commercial markets rather than consumer goods. This helps to explain some of the reasoning behind the work described.

In commercial markets it is common to design items that have very few perhaps only one single realisation of a product. This often means that development costs are high per part and design projects are typically carried out in a highly cost conscious manner and this results in parts that exhibit some noticeable characteristics.

Firstly because there are considerable limits on the time taken to produce designs the process itself tends toward parts that typically have one function per feature. Interestingly this is very much at odds with ‘design’ in the biological world where multi-functioning ‘features’ are the norm. This will be discussed further in Chapter 3. Single features with single functions in a cost limited environment also mean that
designers desire and can re-use design information from other projects. This can mean designing with features that already have some information about how they were previously made and also encourages the use of standard features or even parts. Even with non standard features, typically there are limited manufacturing facilities available and so the range of ways of making a feature are in practice limited (see also CUTK91).

With such considerations in mind it is not surprising that the paper's definitions of design and manufacturing features were as they were. Furthermore the logic of the argument concerning FOD v Feature Recognition is only justifiable in the most constrained of manufacturing environments.

A further effect of the cost limitation and low production runs is that there is a tendency towards simple part geometries. This will be discussed further in future chapters of this critical review.

The major contribution of this paper to research in the area lay in the description of feature interactions that were of interest to those attempting to develop CAPP systems. Ten years later, at the time of writing this critical review it is interesting to consider this paper as a contribution to the literature that shows that automated CAPP systems may in fact be impossible in general environments rather than the paper simply being a list of areas that needed to be addressed. The paper describes problems with access to holes, nested features, simple geometric tolerances, crossed slots and holes, mechanical stability during machining and problems with small cut outs. It is also evident from a review of the test component that there are many potential interactions between the features that the researchers never considered but which could be major factors if the part was being planned in a real situation.

The paper ends with a brief discussion about how such interactions might be represented for communicating to a CAPP system. In particular, a discussion reflecting on the problem of combinatorial complexity ends the paper. Such complexity would become a major theme of much of the work that would subsequently be undertaken in both design and planning and presented in the papers that follow.

The candidate carried out all of the survey work and designed the component that is presented in the paper as well as making the first draft of the paper itself which was
published as an internal research report. Second drafts and the production of figures and useful further discussions were provided by the paper's co-authors.
2.2 DESIGN FOR MACHINING WITH A SIMULTANEOUS-ENGINEERING WORKSTATION (MILL94)

2.2.1 Aim

Using geometric reasoning algorithms at an early stage can help to impose useful constraints on the design engineer and therefore limit the size of the design search space that will be investigated. The primary purpose of this paper was to present to the research community the arguments for this approach and to give examples of techniques for implementing it.

2.2.2 Objectives

The specific objectives of the paper were:

- to present the argument that no matter what stage of design is under consideration, it is worth considering likely manufacturing consequences,
- to present a brief overview of how the Simultaneous Engineering Workstation (SEW) represented an implementation of the proposed general approach to detail design and to explain the architecture of the system, and,
- to explain specific examples of how the SEW system handled feature interactions.

2.2.3 Methodology

The paper described the individual elements of the SEW and how these interacted. The presentation of each module was supplemented by discussion explaining the approach taken. Subsequently, methods were described that could be applied to help constrain the possibilities arising in specific problem areas. These problems were more detailed than those presented in MILL93. The paper therefore represents how the type of problems presented there might be approached in a more specific way.

2.2.4 Results and conclusions

'Most automated DFM systems have been developed using KBS approaches with production rule systems.'

'Much of the reasoning that facilitates the provision of DFM information for metal cutting involves a considerable amount of 3D spatial reasoning.'
'There are considerable advantages to be gained by directly coupling CAD DFM modules to CAPP software where this is possible.'

'The more concurrent a system is, the greater the potential is for the direct coupling of software modules for CAD, CAPP, fixture planning and NC code generation, and this allows greater development of DFM facilities.'

2.2.5 Critique

The second paper, MILL94, follows on from MILL93 and describes early attempts developing a prototype FOD interface known as Feature Oriented Detail Design System (FODDS) (see also CASE94). The candidate developed the part description language further, referred to as COmponent Description Language (CODL). At this stage process planning was being regarded as a design task in itself. The architecture shown in the paper shows how a CODL script could be accessed by geometric reasoning algorithms and updated with information about implicit feature interactions, effectively turning them into explicit feature interactions.

This approach of enriching design descriptions was carried further in other work. The overall architecture of the SEW developed in this direction too. Instead of reading in CODL, enriching it, performing planning tasks and then subsequently writing out what is referred to in the paper as Process Plan Description Language (PPDL), the system evolved into one where a CODL description of the part could be accessed by any downstream function and enriched with whatever information was needed, be it general geometric information, tooling possibilities (described in depth in NAIS97) or, for example, clamping decisions. This could be done automatically where possible and manually where not. In modern thinking this fits well with a product data model approach. The method of using a language rather than a more structured and inflexible database approach is advantageous in that extensions to the language can be added at will and new information can be readily added in a format that allows efficient storage. In actual fact, the language could be implemented in a database that was fully normalised. The language used consisted of sets of triples made up of two objects and a single operator. This was originally developed partly out of convenience of implementing in the Prolog computer language and partly because it was thought at the time that a new range of fast storage devices would become available for storing such data (MILL86).
In fact the language could be readily implemented in any relational database system and would consist of a set of tables representing relations, each with two columns for the objects. This flexible data structure has been implemented in a number of languages (Prolog, Lisp, C, C++) and databases (Oracle, Q+A and Access) and has allowed the candidate's research group to investigate problems in FOD, process capability, general process planning, scheduling, metrology, clamping and fixturing and finally, cost estimating.

The original language was flexible, efficient and powerful but it was text based and had no graphical interface making it very cumbersome to add new components. The work presented in MILL94 shows how the research group linked CODL to a 3D solids modelling engine (ACIS). The user interface was designed so that features added would be written to CODL and used to drive the solids modeller. A number of algorithms were also written to check for potential problematic implicit feature interactions such as void detection, disjoint features, tool access blocks, feature intersections and feature proximity. These algorithms are expanded and explained in greater depth in SALM97.

It is clear from the paper that geometric reasoning is a part of the planning process and that even relatively simple checks are very difficult to automate. There is, however, an important dichotomy in the paper. The suggestion made earlier in this section of the critical review was that CODL could be enriched with additional algorithms as needed but the ordering of these is a major issue. Clearly the type of geometric reasoning algorithms discussed are required to be carried out before process planning takes place. This however means that checks such as access checks can only be carried out with generic tool set ups since the actual tools and machines have not yet been chosen.

This means that, theoretically at least, such an approach could not produce optimal process plans. This is because of the complexity of the whole design and planning process. The architecture shown in the document and implied from this discussion was thus very successful in enabling research of the area but would not suffice as an architecture for any real 'simultaneous' workstation that could claim to find optimal plans. This was and still is a general problem with research in this area and will be returned to later in this chapter.
The candidate carried out the early work that led to this paper. In particular the early feature sets and text based languages such as CODL and PPDL were developed and implemented in the Prolog language. The candidate subsequently overseen the further development of the proposed methods by managing and supervising the work of the co-authors who further developed the work area.
2.3 AN INDUSTRY-BASED STUDY OF CUTTING PROCESS CAPABILITY REPRESENTATION REQUIREMENTS FOR AN INTEGRATED SIMULTANEOUS ENGINEERING WORKSTATION (NAIS97)

2.3.1 Aim
The aim of this paper was to disseminate findings on the possibilities for implementing actual process capabilities from a machining environment into the SEW feature modelling system.

2.3.2 Objectives
The specific objectives of the paper were:

- to present the feature library that was implemented in SEW following an analysis of the design methods in use at Mandelli SpA, a manufacturer of large CNC machining centres, and VMA-NC GmbH, a maker of turned flexible high speed shaft couplings,
- to critically analyse the usefulness of the feature set in dealing with real manufacturing concerns, and,
- to suggest improvements that might be made to enhance the real world applicability of feature based design systems.

2.3.3 Methodology
Following the development of the SEW a study was undertaken to document the design feature in use at a collaborating company. The features identified were programmed into SEW. In parallel with this the manufacturing methods associated with each feature were studied in an effort to investigate the manufacturability of the feature set. The manufacturing methods associated with an individual feature, termed 'microcycles', included operation sequences, their alternatives and details of tooling. They also consisted of the logic that was used to justify the choice of a particular microcycle.

2.3.4 Results and conclusions
The study highlighted again the need to consider feature interactions. In particular strategies for meeting tolerance requirements were extremely important.
Rule based systems cannot represent all types of process knowledge used in process planning, so better structured and enhanced models are a prerequisite for progress in the integration of CAD and CAPP.

The development of process models and geometric reasoning algorithms should be driven by the demands of the applications that use them to support decisions making.

An industrial study used to evaluate the process capability models in a feature-oriented detail design system integrated with a computer automated process planning system led to the identification and development of models to enhance system performance.

Models of shape creation capabilities, models for collision checking and more detailed error models have been discussed as particular areas needing further development.

2.3.5 Critique

The paper reviewed in this section, NAIS97, is a further extension to the general SEW work. The paper is once again based around the planned general system architecture at the time which was in turn built around the CODL feature description language. Although the candidate had studied many industrial components in developing the early versions of CODL these were implemented as ‘general’ features which, it was hoped would be applicable to a wide range of companies. Previous work had concentrated on general models of industry mostly, so that, for example tool types and machines were of a generic nature. The work described in this paper explains the extension of the work into more a detailed single company environment.

The general process planning problem is complex in the extreme. It was becoming obvious at the time that more direct industrial input was needed firstly to test feature libraries and planning algorithms and secondly because domain knowledge was seen as vital in constraining the process planning to a more manageable but specific sub problem. The work described proves this and follows the candidate’s efforts to enlist the help of, and set up collaboration with, a number of companies. Several companies were used to provide components which were used as test pieces. These were reviewed by the research group and if needed CODL would be revised to accommodate the lessons learned, for example if a new feature was needed. In addition, one company in particular would be used as a test for the whole SEW
approach by providing test components, details of manufacturing methods and information on tooling and machines. The company chosen was an Italian machine tool company named Mandelli SpA. Much of the detailed information was gathered by Dr Jane Naish, then a research assistant on the project who spent a six month period at the company location and who later extended and detailed the work in her PhD thesis (NAIS98).
2.4 FEATURES AS AUTONOMOUS AGENTS: AN ALTERNATIVE PARADIGM FOR CONCURRENT ENGINEERING (JACQ97)

2.4.1 Aim
The aim of the paper was to present a novel model for the overall design process in a computer aided engineering environment. The paper seeks to promote the acceptance of an ‘active features as agents’ approach rather than existing models with passive part data structures.

2.4.2 Objectives
The specific objectives of the paper were:
- to show the design of a new data structure suitable for a computer aided design environment,
- to build sample applications based on the new paradigm, and,
- to discuss future directions for work of this type.

2.4.3 Methodology
Instead of the much published model of a passive product data structure being acted upon by modules for detail design, process planning and any other design activity, the paper proposes an active part model whereby each part feature has some limited autonomy to perform its own self design tasks. This is done by programming the feature agents to perform basic actions which are partially dependant on the information passed to them by knowledge modules for activities such as detail design, process planning or part programming.

The data structure itself was developed from the passive SEW data structure developed and presented by the candidate (MILL94). The paper describes the approach and the mode of operation of a demonstration system and describes the steps involved in a simple example.

2.4.4 Results and conclusions
The potential for the use of autonomous features in computer aided engineering was demonstrated and the behaviour of a prototype system shown.
The multi agent system provides an efficient model for the architecture and procedures for design activities in a computer environment and compares favourably with previous iterative design-evaluation-redesign methods.

The model also readily allows for the inclusion of many constraint types on individual features.

A higher level of interactivity in planning processes should be enabled by architectures such as those characterised by autonomous features.

2.4.5 Critique

Autonomous agent based design assistance has become a major area of research in CAE (e.g. see BALA96, FROS96 and WUND96) and this paper represents an early contribution to this. The general approach in using agents is a radical departure from previous architectures for CAE systems that are based around applications that are used sequentially, for example, concept design, detail design, process planning and part programming. Some feedback can be accommodated with existing models but it usually achieved through iteration whereby earlier applications are revisited and changes made as a result of new information that has become available due to some downstream activity. Agents allow for a much greater degree of concurrency than alternative methods. Although designing by features is now widely accepted by the research community and the marketplace, these features are normally passive entities. Endowing features with the means to perform rough checks on such things as thin walls or tooling availability can reduce the need for downstream work and reduce iteration. The features used in the work of the authors use the definitions and overall structure of the candidate’s Simultaneous Engineering Workstation data structures and languages, e.g. CODL and PPDL (described in MILL94).

The paper gives a simple example and shows how autonomous agents can interact and opens many issues related to the control of these. Features that can act in an autonomous way can perform useful functions but the model presented assumes that overall system control remains with the designer. Two major issues arise from the work described in the paper. Firstly there is a danger that agents can become locked in loops whereby an action is taken that triggers a secondary action which in turn reverses the result of the first action. These types of control issue were developed and explored in detail by Jaquel (JACQ00)
The second issue is related more to decision making in the context of the information made available to the author where this relates to the use of resources. Work by the candidate to extend the agent approach to areas such as process planning emphasise this problem. Autonomous feature agents may be able to find resources such as machines or cutting tools that can make them but they cannot easily choose which is the most efficient one to use because of the interactions with other features. In order to progress this work it is necessary to extend the role of the agent to make decisions about cost. This must be done by developing a market in which features compete for resources and as such brings the work into the area of economic agents and biological ones. This is explored further in later papers, particularly in the following chapter which looks at the candidate's work in biological analogies and consideration of the overall size of the search space for this class of problem.
2.5 FURTHER COMMENT AND DISCUSSION

Successfully modelling products, simulating their existence and manipulating the corresponding data associated with them is a complex problem (CHAN90). Present Product Data Management systems are largely concerned with the storage and retrieval of design and manufacture related documents across computer networks. Typically, these consist of separate files made up from solids models, drawings, analysis results, manufacturing information from tooling and quality assurance databases and MRP schedules in addition to business related documents and database files containing information on such things as costing. The wide scope of study that comes under the banner of product modelling means that many areas of manufacture are impinged upon. This is of course partly the reason for the fundamental importance of the activity, since so many aspects of a firm's operations rely heavily on useful product models. The wide scope of activities related to product modelling also mean that models might relate to different stages and times in manufacture. Trends in quality assurance and marketing and the results of legislation have widened the extent of product modelling to include areas of pre-manufacturing (e.g. material/part sources) and post-manufacturing data that are related to a product. Current interest in environmental issues relating to decommissioning and re-use/recycling have also led to the desire to model the whole Product Life Cycle (PLC).

The trend toward storing and manipulating data at different levels of a hierarchy presents problems too. It is becoming increasingly common for templates of products or parametric models from part families to be used and this in turn raises considerations of the trade off between data storage minimisation and ease or speed of manipulation.

Complex hierarchies of parts and assemblies requiring data access across networks with multiple departments' differing needs are further complicated by the need to cope with engineering document release constraints. Further, engineering changes often need to be satisfactorily tracked for any given product.

The complexity of the issues in considering product modelling make it impossible to prescribe a simple single future direction in which development should take place. Clearly consideration should be given to demand pull factors from the marketplace as well as considering the possibilities that present technologies make available. The needs of the market are likely to change too as product modelling technologies
mature. Users are likely to require ever larger participation networks with sharing of information becoming a major issue along with the need for data collection throughout the PLC.

Just as the joint approach of developing de facto and formal standards has been important in the past it is likely to be so in the future. CAD data transfer standards like STEP alongside e.g. DXF operate side by side as markets dictate.

The following discussion considers the current state of product representation methods for the phases of 'engineering', 'production', 'in service' and 'end of use'.

Perhaps the most important aspect of product modelling at the design and engineering phase is concerned with representing part geometries and topologies. Traditionally 2D representations have been prevalent through drawings. Attempts to create standards for such drawings have been widespread but the most widely used standards that emerged have been de facto ones such as DXF. The high levels of market concentration in CAD may explain this. A few companies owned a very large share of the market and demand naturally made their representations the most commonly used.

Imposed formal standards such as IGES have played an important part too, however, especially as 3D file transfer became more widespread. 3D representations have been problematic when, in addition to wireframe data, information relating to solid geometry is also needed.

Some years ago the DJINN group (GEOM95) developed a common language for solid modelling systems based on the concept of a standard geometry language. This could be implemented in any computer language and could be bound above any solids modeller including both set-theoretic and boundary representation types. Driving such modellers with universal geometry would allow data to be passed between systems at least at this high level.

Assembly models are also increasingly used in CAD systems and several offer a means to pass assembly descriptions through part models and constraints between different design modules for applications in animation, kinematics and solids analysis (see e.g. MEDL00). The STEP standard too incorporates the need for assembly information.

Design analysis work is needed to check the likely operation of products so that they meet design specifications or may be used as part of an improvement/optimisation activity. Recent improvements in analysis tools and their ability to read part
geometries from solids models and assemblies are dramatically increasing the user base of automated analysis tools.

Finite element tools in desktop packages are becoming widely used for tasks such as structural analysis. These systems incorporate the ability to read geometry from midrange CAD software through solids models. Commonly the software is tightly integrated so that transfer is not necessary. Automatic meshing and the incorporation of extensive domain knowledge to a given type of problem means that certain types of analysis can be carried out largely automatically. These tools are useful as design checks although there is a danger that engineers inexperienced in F.E. analysis can use them and therefore risk misreading simulation results or performing erroneous analysis. In optimisation, tighter integration means that results can easily be linked to CAD representations, however, researchers continue to work on the problems of tightly integrating meshing and search procedures for a wide variety of problem types. More general search methods using zero order methods are also becoming more common in the research community because they are robust, although they are also frequently inefficient. Kinetics and kinematics analysis tools are also popular for analysing mechanisms and these too integrate tightly with the current breed of CAD system. At present they are capable of reading assembly descriptions and interpreting these in terms of their constraints on movement. Current systems are specialised, dealing with rigid bodies, and to date it is not possible to carry out joint movement/deflection analysis.

One approach of current interest to the research community that might improve the ability to integrate different representations and analysis tools may be to use agent based intelligent representations of components. These systems can be hierarchical. Unlike other agent-based work in manufacturing, the agents are numerous and alike or identical. The approach has been compared to modelling cells as insects.

Because manufacturing processes are inherently imprecise, artefacts vary in form and material properties and thus in performance. In order to control this variability designers must specify constraints on the part models. Although constraints may be applied to any source of variation, the most common ones of concern in terms of modelling are dimensional tolerances. In the past 30 years a great deal has been achieved in terms of modelling part geometries in 3D so it is perhaps surprising that
the application of geometric variation analysis tools for product modelling have advanced little.

Traditional dimensional limit tolerances were in use from around the late 1800s and are still in use today even although the problems associated with them were widely apparent in the 1940s when the rise of geometrical tolerancing began. Although geometric tolerances were clearly useful they were often badly applied in the past and their lack of systemisation meant that they did not serve the purpose of standardisation adequately. The problems with geometric tolerancing were put up with until modern technology made the desire for the 'mathematisation' of geometric tolerancing schemes. Although in widespread industrial use the relevant standards are elaborate and often interpreted differently in different organisations. These standards tend to specify rules for local interactions but clear guidance on methods for extensive and complex stack ups are not adequately defined. In the 1970s CMMs became available without sound methods for their use. In particular the measurements made by those using these machines were often irreconcilable with what traditional two point manual measurements gave. This led to what was known as the 'Metrology Crisis' of the 1980s and progress to its solution has proceeded slowly.

An additional problem with the use of traditional dimensional limit tolerances also became of interest to engineers in the 1960s. The assignment of worst case tolerances for assemblies with tolerance build up problems was objected to and a substantial number of engineers turned to statistical tolerancing techniques. These methods are growing in use in industry today. They are usually based on normal distributions and are typically used with root sum of square stack ups. One major difficulty with statistical methods in dimensional tolerancing is that many companies have tended develop their own methods of use of these and there are, as yet, no accepted standards that cover current practice (although there is ISO activity in this area). Another approach to this problem involved departure from assemblies of interchangeable parts. Some firms returned to the method of matching individual parts so that a sub-assembly would be within an overall tolerance and could be interchanged with another complete sub-assembly although individual parts were not interchangeable.

More recently similar problems emerged with the use of geometrical tolerances but the application of statistical methods is more complex with these and whilst there is
undoubted industrial interest in them, there are no established standards and there are major open issues (again ISO is active).

Future work in tolerancing is likely to proceed on several different approaches, as Voelcker (VOEL97) has described.

The first approach can be described as the maintenance and 'mathematisation' of current standards and is aimed at improving current methods through making them computationally more robust. The second front involves extending current standards to make them more widely applicable. This includes correcting major deficiencies in e.g. the concept of datum systems (SURE94) as well as work in statistical tolerancing (SRIN97). A third, and much more ambitious area of current activity, is that which seeks to rationalise comprehensively a wide range of tolerancing and metrology applications. If successful tolerancing schemes that were consistent with modem metrology methods in dimensional and surface measurement would result and these could also conform to wide ranging CAD standards such as STEP.

In parallel with these activities work is also likely to proceed on better basic understandings of the theories involved in tolerancing.

While designers cope with the many problems of designing components that function in terms of many criteria, they also often face the difficult task of considering how components are to be manufactured. Indeed manufacturability is often thought of as equal in importance to functionality. Considering some factor other than straight functionality may involve designing for manufacture, maintenance, or disposal for example, and is commonly referred to as Design for X (DFX). In the case of manufacture this may include subdivision into categories such as Design for Machining, Design for Handling, Design for Assembly, Design for Fabrication, Design for Moulding or any other aspect of the production process. Because each activity requires consideration of different aspects of a component or product, each also often requires different representations.

Designers are also requested to make up different drawings for the different manufacturing functions and these may include part drawings, assembly drawings, fabrication drawings, fixture drawings, layouts etc.

Part geometries on their own may not be sufficient for modelling manufacturing problems and may be augmented with information on material, dates required or other attributes. It is also possible to use different views of part geometry and topology
through manufacturing process oriented representations. These might take the form of parameterised designs or designs coded in terms of 'features'.

Feature oriented descriptions of components can be created through feature based design systems or through the recognition of features from solids models. Feature based design systems are becoming common in the marketplace but each system supports its own feature set. Moreover, any given design can be modelled in multiple ways on these systems and it is therefore not possible to tie particular manufacturing methods to features in an extensive way. Thus most modellers use features as a means of high-level geometry creation without a link to manufacture. Increasingly modellers are incorporating design environments that are associated with particular types of manufacture, e.g. moulded plastic or sheet metal parts.

Manufacturing problems frequently involve reasoning about orderings of cuts, treatment of areas of components, clamping and holding parts, assembling and machining partially assembled states and doing all of these in any order. The complexities of reasoning about time in realistic worlds rather than simply stating orderings, specifying dates or estimating the duration of single tasks, are very difficult to deal with. Allen (ALLE91) discusses at some length the difficulties and the many ways to reason about time. Whilst researchers in manufacturing have understood the importance of temporal representations (PART89) (BHAS99), they have worked in a narrow area and tended to use simple temporal reasoning. A more useful product model for wide CAPP use should have a rich temporal reasoning capability. It is not sufficient to use constructs like 'before' and 'after' on their own. Additional useful constructs like 'during' 'meets' or 'overlaps' are useful too.

The requirements for product modelling largely represent the needs to describe a product's manufacture. If complex product descriptions can be created and readily interchanged between systems then it makes possible re-use of this data at almost every stage of production. Robust geometric and feature enhanced models could make the development and verification of NC code more powerful. The IMS project STEP-NC (BROU98) had the aim of making improvements to the CAD/NC interface by developing a STEP based standard for inclusion in NC data models. Such a standard would improve this interface by allowing the use of CAD data without conversions (although IGES based systems are also currently in use). This would make possible a more satisfactory way of transferring complex models such as those...
containing splines. It would also enable two way transfer of data across the interface. The use of full CAD models at the shopfloor would also mean that changes could be made to part programs where this is desirable. There appears to be considerable commercial interest from vendors and potential users of STEP-NC around the world.

The realisation of feature oriented models complete with temporal information could potentially revolutionise shopfloor documentation. On-line pictures and configurations of partially manufactured products could be called up for reference at will providing a means of constantly comparing planned with actual. Vastly improved instructions for shopfloor work would be possible.

Perhaps the most obvious area where design and production need to share a common model is that of quality assurance and metrology. The improvement of tolerance methodologies would make possible tolerance descriptions that related directly to their means of measurement on modern metrology equipment. Importantly the use of large product models allows for more customised data to be carried with individual products. Thus custom Bills of Materials for individual products that are often created from templates are becoming increasingly common. Moreover, it also becoming more easy to record manufacturing data, including metrology or other test results, that are relevant to a single product. This helps not only at the manufacturing stage but also makes for an improved product model being released with the product into its service phase.

In the service phase of a product it is useful to have access to a product's design and manufacturing data. This can be used to help in obvious areas like fault analysis and maintenance. It is also important as a tool for tracing the source of faults and this is increasingly becoming a QA requirement especially in safety critical environments.

In addition to the flow of data from design and production into the service phase it is also becoming more common for a flow in the opposite direction to be used. Data from the service life of a product can be added to a product model. This is useful as a feedback tool for marketing and design functions. Further analysis of performance against design and manufacturing processes allow for on-going product improvement and the generation of data useful in the design of other products.

An area that is currently of interest to the research community is that of service life forecasting. This involves developing models of a product's behaviour in its planned environments. Thus information may be generated that predicts the nature of the
product under certain conditions of loading or chemical environment. Clearly all of
the data relating to the service phase of a product has an important temporal element
to it.

At the end of a product's life some of the data generated at any of the previous stages
of design, production and service may still be required if decommissioning or
recycling are important issues. The chemical make up of products is clearly important
in determining how to dispose of them. Where recycling is possible and desirable then
nearly all past data on a product may be useful. The part design, manufacturing
methods and service history of a product or sub-product may all be important in
enabling recycling decisions to be reached. Also, if a product is to re-enter service it
may be used directly as a replacement part (e.g. aeroplane or car spares), as a discrete
component in a regeneration process (e.g. remoulded tyres) or as part of an input to a
batch or continuous process (e.g. scrap metal or recycled plastic)

These options and the data needed to be carried with components make the
requirements for product modelling in the end of usage phase complex. With open
system recycling (e.g. glass bottles) there is no need to carry data from one phase to
an other but the main interest of the research community lies with closed loop
systems. Here, a sub-component of a product may be recycled and the product model
information must travel with it. In the case of a reconditioned sub-assembly, e.g., this
means that the new product's model will become more complex than a simple single
life-cycle view of it. In the case of parts that are used in a batch or continuous process
(plastic parts such as those found commonly in cars) a part may originate from a
process, serve life as a discrete component, be returned to a batch process and
subsequently become an input to new discrete components. This raises the spectre of a
potentially recursive product view.

As product models become ever richer and cover more of the total life cycle their
longevity becomes of increasing concern. It is not always possible to update old
drawings and other information but it may be a requirement to be able to view these if
need be. In the case of printed material this is straightforward but increasingly old
data is held in old systems. Thus there is growing concern that comprehensive product
models may not simply need data but also the means to process it by way of
embedded packages or even operating systems.
Currently, most of the considerations that were presented in the papers in this section and in this critical discussion form major parts of the modern STEP standard (ISO 10303). This standard is evolving to include sections or proposed sections on metrology and tolerances, feature sets, tooling, relationships, assemblies and further manufacturing data including some basic temporal representations and viewers for STEP compliant models. Extensions to the standard are coded using the EXPRESS language and an application programming interface to this has been designed (Standard Data Access Interface or SDAI) which has binding for several modern programming environments.
3 BIOLOGICAL ANALOGIES

The candidate started investigating genetic algorithms in 1985. Initially it was thought that these might be used to find a near optimal solution to the process planning problem. Later, other applications areas became apparent such as the problem of finding near optima for a range of other design based problems, e.g. the best shape for a wing or turbine blade. Since 1985 a great deal of research has been carried out in many fields of interest but the majority of this has been aimed at using genetic algorithms as optimisation tools.

Other biological analogies have also been investigated using techniques such as genetic programming, neural nets and rule based heuristics. This chapter provides a critique of the candidate's work in these areas and describes the partial success of the application of algorithms inspired by nature as well as considering in depth some of the on-going problems in using these for common manufacturing planning problems.
3.1 A VOXEL-BASED REPRESENTATION FOR EVOLUTIONARY SHAPE OPTIMISATION (BARO99)

3.1.1 Aim
The purpose of the work undertaken by the collaboration of authors was to attempt to develop both a representation of shape (and topology) and operators that could manipulate these to allow the application of a genetic algorithm. The aim of the paper was to disseminate the authors’ findings to the joint Artificial Intelligence and Engineering research communities.

3.1.2 Objectives
The specific objectives of the paper were:

- to develop a cellular representation suitable for describing 3D components without the need for unnecessary explicit constraints (e.g. those imposed by predefined topology),
- to develop individual operators that would allow the use of a genetic algorithm, for example, a crossover operator that would allow analogies of chiasmata to be used, and,
- to show the efficient operation of the developed representation and operators.

3.1.3 Methodology
Two engineering design problems were considered. Their shape and topology requirements were studied and a suitable spatial occupancy method was chosen. Because efficiency of memory usage was not considered to be a problem, a complete cellular representation was chosen using equal sized rectangular cells that were sufficiently small to achieve the desired level of granularity. The representation that was implemented allowed 2D profiles to be developed that could subsequently be, e.g. spun or extruded to give the full 3D shape.

Common operators for such representations were studied and where necessary these were improved upon to give suitable speed and robustness. Finally the operation of the complete method was tested on chosen test problems. The instances of part designs were evaluated using analytical models in the case of a bending beam and numerical finite element methods for an annulus problem.
3.1.4 Results and Conclusions

'Voxels were found to be a viable representation for shape optimisation with an evolutionary algorithm in 2D problems.

'a number of difficulties inherent with this (cellular) representation were addressed, primarily by use of specific genetic algorithm operators that utilized domain knowledge held about the problems'

'On the annulus design problem, the direct use of the voxels as the finite element mesh was found to be inadequate, and a convolution mask based solution to this issue was devised.'

'the flexibility of the voxel representation, along with the genetic algorithm's exploitation of a much expanded search space uncovered deficiencies in the specification used for the annulus design problem, leading to unwanted "overhangs" in the solution obtained.'

'it should be noted that genetic algorithm optimisers can easily be modified to be used as interactive optimisation systems........an engineer's practical experience and knowledge of the problem domain (could be used) to direct key choices in the optimisation process'

3.1.5 Critique

The paper showed a viable means of carrying out design optimisation on two sample problems namely beam and annulus applications. The combination of a general shape representation based on voxels and a robust optimisation technique using a genetic algorithm is very useful in that it applies the minimum possible constraints on solutions to this class of problems. Theoretically the success of this approach could lead to a class of design optimisers that would only need specification of the application of loadings to a part and would result in an ideal minimum weight shape being produced for the designer within the CAD environment. The limitations at the time of writing the paper however were considerable and it has taken some time for partial solutions to the problems arising to be overcome.

The paper addressed difficulties with genetic operators because little work had been done previously using cellular part representations. It was shown, however, that suitable operators could be devised and this has been shown to be the case by the
research community in general who have continued to work on cellular descriptions using a variety of newly developed operators (ANNIO1, BELA00, CAPE03, KITA00 and MICH92) as well as further development of mesh free methods (CHAN01 and GRIN02).

Problems arising from the general efficiency of genetic algorithms and of finding suitable search parameters continue however, and this is not only true of this application area but also of those discussed in the previous chapter and of those presented in the next paper ‘Biological analogies in manufacturing’ This problem will be discussed in more depth at the conclusion of the chapter.

Another problem arose from the fact that genetic algorithms successfully search the entire available space of solutions and in turn rapidly exploit any ambiguities in the original problem specification. This shows the importance of producing accurate robust problem specifications for use with these methods and this is in itself a major difficulty in the application of genetic algorithms to such problems. Where possible the use of domain knowledge on the part of the designer should be used to guide the search, perhaps interactively, thus avoiding the investigation of fruitless areas of the search space as specified.

Perhaps the major problem with the approach described in the paper lay in the representation itself. Voxels are geometrically simple shapes and produce simplistic part representations for analysis purposes. They do not serve as useful elements to be used in most finite element analysis applications. One way round this may be to produce a separate mesh at each iteration of the analysis although this would be inefficient. Alternatively, rather than using simple voxels, more complex element agents could be employed. Thus, as features were used to be self activating agents in (JACQ97), then finite element agents could be used to represent parts. These can be built into rich data structures that hold a great deal of information about the parts and can be readily processed with analysis applications and have self firing action plans that allow them to adhere to morphogenesis rules and be subject to evolutionary algorithms. This approach has formed much of the work of Sherlock (SHER03).

The problem that this paper is largely based around was originally given to the candidate following collaborative research work between the candidate and company. The candidate supervised the work of Sherlock who has since devised more suitable approaches for voxel based representations that are suitable for processing with
genetic based optimisation methods. The paper itself is the result of a further collaboration between members of the University’s Department of Mechanical Engineering and the Department of Artificial Intelligence who were keen to try to develop the representations that had been used beforehand.
3.2 BIOLOGICAL ANALOGIES IN MANUFACTURING (MILLOO)

3.2.1 Aim
Following work by the authors using biological analogies for design and manufacturing optimisation problems, the candidate was invited to give a presentation at a meeting of the Intelligent Manufacturing Systems Working Group, a European Commission funded network whose role is to study advances associated with the worldwide IMS project. Subsequently the authors were asked to submit a paper for publication in the Computers in Industry journal.

3.2.2 Objectives
The principal objective of the paper was to disseminate to engineering researchers current thinking by workers in the rapidly emerging and wide ranging field of biological approaches to manufacturing.

3.2.3 Methodology
Given the requirement for speed of dissemination the authors relied little on published work in the field and expressed instead their findings from their own work and from first hand knowledge of the work of other important groups in this area.

3.2.4 Results and conclusions
It is an extremely challenging task to model the entire life-cycle of a product from market research, through design, planning, sourcing and manufacture, and decommissioning.

Biological Systems exhibit characteristics of complex systems. They show evolutionary, adaptive, self organising and emergent behaviour.

There is currently strong interest in using knowledge of biological systems such as ant colonies as an analogy of the complex systems that are evident in manufacturing environments.

Although such analogies provide useful insights into processes such as design and manufacture, there are also fundamental differences in the way products can be modelled and developed.
3.2.5 Critique

MILLOO was written as a result of a presentation given by the candidate to the “Intelligent Manufacturing Systems Working Group” EC Network of Excellence (IMS-WG). The candidate was asked to review current thinking as of 1999 in the area of biological analogies in manufacturing. This followed earlier work by the authors in the area of genetic algorithms and agent based applications in design and process planning. The paper itself is a brief critique of what was considered to be the most important work undertaken by the research community as a whole and sought to give manufacturing engineers an overview as it was relevant to the membership of IMS-WG.

Initially the paper presents a comparison between what are termed biological design and product evolution. Whilst it is commonly accepted that industrial products evolve in the sense that they can change character gradually from one generation to the next, the authors sought to study the mechanisms driving this constant change so that they could better understand the phenomena.

It appears that there are two extremes of design implied in the paper. Firstly a blind, highly iterative method characterised in the natural world, and secondly a forward planning and modelling based intelligent method characterised by engineering design. Whilst the first is costly in terms of the thousands or millions of iterations that are often required to produce complex designs, the other is costly in the respect that it requires a model of the environment in which the design will function in order to predict performance. Current reality, however, is not always at these extremes. Natural evolution has not produced simple blueprints of designs but flexible developmental means of characterising features and designs. Many, if not all biological designs, have flexibility built into them. The firing mechanisms associated with genes can make it possible to use triggering mechanisms so that phenotypical features, for example sex, can be determined by environmental factors (RIDL03). Of more interest to the engineer perhaps, are the less dramatic reactive rules that are used in biological systems.

The growth of bones and trees are two examples where form is determined in part by environmental factors. The algorithm described in the paper is based on the growth phenomenon whereby extra material is grown in highly stressed areas. Such models are easily implemented with the use of suitable part representations. One such
representation is based on the use of the ant colony analogy. In this agent based approach, all agents are identically programmed and can receive signals from the wider as well as their immediate environment. Sherlock (SHER03) has implemented such models and investigated their use in some depth and it appears that such models offer a robust means of representing parts for a wide range of optimisation and analysis tasks including heuristic rule based, evolutionary programming such as genetic algorithms, and, for example, finite element analysis (see also e.g. GRAH01 and PAPA02.

An interesting aspect of biological design when compared to engineering design and which is not discussed in MILLO0 lies in the differences that are apparent in the end product. Biological forms are often far more complex than manufactured ones. It is sometimes argued that this is due to customer tastes but is likely to be largely because complex curved shapes are both difficult to compute and often very hard to manufacture. The complex process of biological development is based around gene expression through the production of proteins which in turn work in chemical environments to form shapes (RAFF96).

A second major difference is that while man made designs usually consist of a collection of features which perform individual functions, biology commonly makes use of multifunctional features. A simple example is hair, which in humans plays multiples roles, protector from the sun, impact protector, heat insulation and aesthetic feature. Multifunctional features in the man made world are rare however, and are usually considered to be the product of particularly clever design. It would appear that this difference may be due to the different ways that designs are evaluated. In the biological world designs are evaluated and selected for their fitness only at the level of the complete being which in turn dictates how successful that being is at surviving and breeding. With man made designs considerable evaluation is carried out on the drawing board by designers thinking in terms of individual functions. A clever designer might take more that one view of a single feature but this is a difficult task and because the evolution process rarely get many iterations through which to operate, richly multifunctional features are rare. The use of computer tools such as genetic algorithms could, in theory, enhance a designer's ability to produce more flexibly featured products but the process is still constrained by the limitations of current
multipoint design methods and multi-objective evaluation functions. Progress in these areas is somewhat slow.
3.3 FURTHER COMMENT AND DISCUSSION

MILLOO represents the candidate's views of particular aspects of evolutionary computing in engineering design and some of the work discussed previously is referenced in the text. The paper does not review general work, particularly in the field of genetic algorithm based shape optimisation. There have been many hundreds of papers published since the mid 1980s in this area, most presenting what is claimed to be an improved method of finding optimal shapes for a particular application. The improvements claimed usually fall into types. Firstly it may be claimed that the genetic algorithm used is more efficient than existing methods that are used and comparisons of run times are usually given for the genetic algorithm with other methods such as gradient based search, heuristic rule based methods or other zero order methods, for example simulated annealing. More recently hybrid methods have been adopted where gradient methods or heuristic rule based searches are combined with genetic algorithms. These hybrids are usually presented as improvements on past used single techniques.

Unfortunately there is little that can be generalised from such work. More commonly in fact the opposite is true, using many of these methods can often result in worse performance in finding optima than using more general methods. This includes the work described on wing design. Nearly all of the work reported a particular hybrid or genetic algorithm set up together with a description of detailed parameters such as the population size used, the mutation rate, fecundity or any of a long list of those available. This causes the problem that the researchers have often spent a great many weeks, months and sometimes years finding the right parameter set up in order to carry out a search in perhaps a few seconds or minutes less than other methods. This could be worthwhile if the parameters could then be applied to a more general set of problems but it highly unlikely that this is often the case. The reasons lie in what has become known as the 'No Free Lunch' theorem and is described in WOLP95.

Consider a set of problems, perhaps 'wing optimisation' problems S, as shown in Figure 3.2.4.1.
Now assume that the average most efficient time taken to find an optimal solution to this set of problems is given by $T$. A subset of these problems $s_1$, perhaps wing shapes for minimising drag and maximising lift in a particular speed range, may be defined. Suppose this takes an average most efficient time of $t_1$.

Since $s_1 \subset S$, $t_1 < T$

and $s_2 \subset S$, $s_2 = S - s_1$

then $t_2 > T$.

In other words, using the set of parameters that will give favourable times for $s_1$ will give worse results than average for $s_2$.

A second claim that is made for the use of genetic algorithms lies in the fact that they are relatively robust algorithms and this is the main reason that they are used in so many of the candidate's areas of interest. Genetic algorithms are good general purpose algorithms that can be applied to messy discontinuous problems and can readily be applied to a wide range of part representation models. Any search can be speeded up if suitable domain knowledge can be used effectively as this can cut down the area that is needed to be searched.
The candidate's first experience in studying genetic algorithms lay in work investigating applications in process planning. It was felt that the search spaces in this area were likely to be very large and that genetic algorithms might be used to find near optimal or satisficing solutions if suitable knowledge based systems could represent domain knowledge that could in turn cut down the size of the search space to a manageable size. Unfortunately the search space size estimation problem was commonly not estimated, a problem when subject to constraints is NP Hard.

Halevi (HALE80) published an expression for the size of the search space for a simple process planning problem, that of sequencing machining operations and choosing machines to perform these. Halevi's equation for $P$, the size of the planning search space assumed $M$ machine choices for each of $N$ features on a component that is to be manufactured, giving

$$P(N) = M^N \times N!$$

Using the same approach, this expression can be extended to include $Q$, the number of measuring configuration choices for each feature, $S$, the number of set-up choices possible for each feature, and $T$, the number of cutting tool choices for each feature. Hence,

$$P(N) = M^N \times Q^N \times S^N \times T^N \times \frac{N!}{2^I}$$

where $N$ is the total number of features i.e. inclusive of intermediate machining features and inspection features and $I$ is the total number of intermediate features. The expression assumes that in the worst case, the number of machines is the same for every feature as are the number of measurement configurations, set-ups and cutting tools. In the worst case this is true because these numbers are at the maximum possible in each case i.e. all machines, tools, set ups and measuring choices are available for each feature. The expression does not, however, allow the substitution of real values for individual features and is therefore rewritten as

$$P(N) = \frac{N!}{2^I} \times \prod_{i=1}^{N} M_i \times Q_i \times S_i \times T_i$$

where $i$ is an individual feature.
Extending this approach we can estimate the number of assembly possibilities by estimating the number of assembly orderings and the ways W and holding setups H that might be required for each assembly operation. Thus the search space A for an assembly of D discrete parts can be estimated as:

\[ A(D) = D! \prod_{j=1}^{D} W_j * H_j \]

Considering the combined component manufacturing and assembly problem as a whole yields:

\[ A(D) = D! \prod_{j=1}^{D} W_j * H_j * P_j * \left( \prod_{i=1}^{N} M_i * Q_i * S_i * T_i \right) \]

assuming the simplest and most common case where cutting processes are not performed on subassemblies.

Substitution of even the most modest values for the terms in this factorial equation leads to search spaces that quickly exceed the estimated number of atoms in the Universe and exhaustive search is not an option.

In reality however several factors influence the real space that is required to be searched in order to achieve a satisficing solution to the problem. Firstly there are many constraints that limit the outcomes. Simple independent anteriority constraints can, e.g., half the size of the search space. Constraints on the numbers of available manufacturing process choices reduce the search space for each discrete component linearly and similarly constraints on the number of assembly methods linearly reduce the overall search space. Further, although the theoretical worst case situation assumes that each of the factors are independent, in practice they are not. Thus, the choice of a particular cutting tool, e.g., will clearly also constrain the machine choices that are available. In practice, the engineer commonly applies constraints based on his feel for what is reasonable and thus limits his search space to the size that he is able to investigate. Such solutions are satisficing rather than optimal.

One further problem associated with the type of problem is that when anteriority constraints are applied, estimating the actual size of the search space, rather than the best case, becomes in general, intractable. How this can be managed is beyond the
scope of this review; however, the problem of determining if any solutions exist at all can be shown to be linear.
4 MACHINING APPLICATIONS

In this section two papers relating to Electro Chemical Machining (ECM) are presented. Again the theme that involved the candidate was shape design. In both cases the problem was the design of tools that were required to cut specified shapes.

In 'Diverless weld inspection and repair using ecm/acfm techniques' (CLIF00) the tool was designed to be a general shape that could accommodate a range of cut shapes by controlling tool movement orthogonal to the cut surface. Conversely in 'A direct analytical solution to the tool design problem in electrochemical machining under steady state conditions' (ALDE00), the tool movement was vertical in all cases and complex tool shapes were required.
4.1 DIVERLESS WELD INSPECTION AND REPAIR USING ECM/ACFM TECHNIQUES (CLIF00)

4.1.1 Aim
The aim of the work described in the paper was to develop an alternative means of crack removal and repair from subsea structures that would not depend on the use of divers. Traditionally cracks are removed by grinding, a process that is usually carried out by divers and is dangerous, depth limited, results in the region of the crack being destroyed and leaves considerable residual stresses.

4.1.2 Objectives
The specific objectives of the work reported in the paper were:

- to develop an ECM machining model for crack removal,
- to develop a suitable control system to achieve cutting of the required geometry,
- to integrate the above with an alternating current field measurement (ACFM) system for crack detection, and,
- to carry out preliminary deployment tank trials for proof of concept of operation from a remotely operated vehicle.

4.1.3 Methodology
Two sample tools were manufactured that were representative of the tool types that would likely be used in practical situations. For these, estimates were made of the likely tool deflections and these were a function of the electrolyte flow rate through the tool. The equilibrium machining gap should not fall below this level for fear that any interruption in the electrolyte flow would allow the tool to recover from the deflected state and touch the workpiece.

A simple model of the cutting dynamics of the ECM process was used to calculate system parameters. A test carriage was then designed and built in order to test the system in a test tank. The system allowed an ACFM probe and the ECM cutting head to be mounted on a single carriage that would be driven along a cam in order to achieve the desired cutting shape.
Trials were carried out using the carriage arrangement in a 10m deep fresh water test tank.

4.1.4 Results and conclusions

The development of ECM parameter models, together with data from machining trials, has enabled the sampling operation to be accurately parameterised and modelled.

The defect sampling/removal operation can be carried out at tool feed rates of up to 3mm/min, however, this would only be achieved at an increased voltage of 35V and this was not attainable with the equipment used. The parameterisation allows the effect of system variables to be predicted and therefore the various trade offs between cutting speed, flow velocities, voltage, tool size and overcut dimensions.

The ECM defect removal/sampling system was shown to be effective in a submersed environment under remote control working with an ACFM probe.

4.1.5 Critique

Much of the work undertaken as part of the project that is described in CLIFF00 was related to the prediction of sample shape from tool shape, path and feedrate. The initial tool shape chosen was a simple U-shape as this was easily manufactured and allowed satisfactory delivery of electrolyte through the channel that was machined into the front of the tool.

The model of the cutting process used predicted that a sweep of the tool shape with an allowance for an overcut could be used together with a further allowance to ensure the crack was completely contained within the sample by a reasonable margin of 1mm. This allowed the required shape of cut to be predicted and the guide curve for the tool could be computed. This was machined onto a cam and the tool carriage ran along this when cutting.

The use of the cam was successful, however, it is planned that tool carriage should be put under direct computer control so that the system would not need new cams to be fitted and the system could react to ad hoc cutting requirements in the event of, say an additional cut being necessary.

There were some minor problems with this model. Firstly the sample was slightly narrower and shorter than predicted by an amount dependant on the feedrate. This
was due to the simplicity of the cutting process model which did not include machining occurring from regions of the tool above the sample. Secondly the surface characteristics of the cut side of the sample were wavy (typical wavelengths of 1mm). This may have been caused by the control system reacting to inclusions in the sample material.

Since the paper was published considerable advancements have been made in the modelling of the ECM process and this is a major element in ALDE00 which is described in the next section. Despite the view taken when writing CLIF00, it now seems that it might be possible to use lower concentrations of NaCl and it may be possible to machine with seawater as an electrolyte.

One major problem with the approach taken in this project lies in evaluation of the environmental impact of the use of the process in subsea areas.

Considerable development of the work presented in the paper has been undertaken by Clifton and is described in detail in CLIFF01 and CLIFF02.

The candidate's role in the work described involved the overall management and supervision of the project as well advising on the tool and cam shape parts of the work and the modelling of these.
4.2 A DIRECT ANALYTICAL SOLUTION TO THE TOOL DESIGN PROBLEM IN ELECTROCHEMICAL MACHINING UNDER STEADY STATE CONDITIONS (ALDE00)

4.2.1 Aim
The overall aim of the project that this paper relates to was to find a convenient method of cutting down the number of iterations in designing ECM tools for complex shapes.

4.2.2 Objectives
The major objectives were:

- to consider first a 2D profile shape of the desired workpiece using a Fourier transform representation,
- to model the cutting process, and,
- to use an inverse method to specify tool shape.

4.2.3 Methodology
A sample workpiece was chosen as a test shape. The workpiece was then represented by using a 2D Fourier series. Although methods such as spline curves would have been more convenient, particularly when incorporating tool shapes in CAD models was needed, the Fourier series could not only model the shape but lent itself readily to the inclusion in the conformal mapping that resulted from the field model of the metal removal process.

An experimental validation was then carried out for differing machining gaps in order to assess the feasibility of the method.

4.2.4 Results and conclusions

The theory used in the paper uses an ideal model of the cutting process in ECM where a uniform gap conductivity is assumed and only electric field is considered.

The measured results show that the theory used is valid in that close agreement is reached between theoretical and actual surfaces.
The successful process model allows insight into the tool design process, in particular in being able to assess limiting cases where the theoretical tool geometry cannot be realised.

4.2.5 Critique

The paper represents an important advance in the process of designing tool shapes for a particular application, in this case electrochemical machining. The method used is both fast and proven to be accurate, however, as with the voxel design methods tried in BAR099, the method is not yet ready to incorporate into an automated design environment and to be integrated into a CAD system. The part and tool representation used was suitable for integrating with the analytical method used for characterising the ECM process but with complex surfaces would require a high number of harmonics to be used and this would counteract the advantage gained from using an analytical method. The system would not readily cope with sharp edges in the tool or workpiece although these are not always physically possible with the ECM process at least for workpiece realisation.

The candidate originally proposed this work and together with Clifton developed the idea and managed the on-going project that is investigating this area through the collaboration with Alder and the supervision of McLennan (MCLE02) who has further developed the system so that it is more suitable to work with modern CAD systems.
4.3 FURTHER COMMENT AND DISCUSSION

The papers CLIFF00 and ALDE00 are both concerned with shaping cuts in metals. Both were built upon complex process models that were developed by members of the ECM group at Edinburgh University. In the research described in CLIFF00 the shape design was achieved by manipulating parameters relating to the tool path whereas in the work described in ALDE00 the tool shape was varied.

The project work on the defect sampler resulted in successful trials of the prototype equipment. Further development of the sampler, for example using NC controlled tool path generation will await further developments in the market. The machine has several application areas in addition to subsea ones but development is likely only to take place as part of a commercial venture and is not now considered to be further interest to the research community.

The tool design problem using an analytical approach is likely to spawn further research however. Unlike the methods used in the optimisation problems discussed in BAR099, MILL00 and CLIF00, where forward generate and test iteration was employed, the tool design discussed in ALDE00 used an inverse optimisation method. The method relies on having a model of the cutting process that can readily be used in reverse and this favours analytical rather than numerical methods. The favoured CAD representations, e.g. B-Spline or full NURBS, are currently being investigated by McLennan (MCLE02) and this should make the method easier to use with parametric feature based modern CAD systems such as SolidWorks or SolidEdge.

Throughout the period of research described in the critical review the research interests of the candidate have narrowed. At the outset interest was in general feature descriptions which were only loosely related to general manufacturing processes. Subsequent work saw a concentration on specific manufacturing facilities with general capabilities and now the work in this section is related to a specific machining process. All of this work has been in the context of optimisation methods.

The candidate’s effort is likely to continue in specific areas and current interests exist in bottle, pipework and architectural ornament design.
5 CONCLUSIONS

1 Feature oriented design has become part of mainstream 3D modelling, however, it tends to be restricted to generic process sets rather than specific user defined manufacturing facilities. The ongoing lack of standard feature sets means that feature information cannot readily be transferred from system to system.

2 Feature Recognition is an important tool in transferring or building feature enhanced part models and is likely to become increasingly important as the technology becomes more powerful and network transfers and searches become increasingly popular.

3 Design for manufacture data and knowledge bases are becoming more widely used and this is likely to continue with web based applications.

4 Increasing modularity through specialist packages and plug-ins is likely to lead to greater user customisation and will also help achieve better concurrency.

5 Feature based paradigms are leading to improved methods of constraining part designs and therefore to more robust optimisation techniques in practice.

6 Biological analogies are becoming more widespread in many aspects of design as a result of advances in part modelling, computation and modern manufacturing processes.

7 Much of the published work on particular optimisation strategies, for example genetic algorithms, cannot be generalised and is of very restricted interest to general research communities.

8 The size of the search spaces that have been investigated are typically very large (NP Hard) and unfortunately the secondary problem of accurately estimating the size of these is also NP Hard.
6 LIST OF PAPERS

1 MILL93 – (Reproduced in Appendix A)

2 MILL94 – (Reproduced in Appendix B)

3 NAIS97 – (Reproduced in Appendix C)

4 JACQ97 – (Reproduced in Appendix D)

5 BARO99 – (Reproduced in Appendix E)

6 MILL00 – (Reproduced in Appendix F)

7 CLIF00 – (Reproduced in Appendix G)

8 ALDE00 – (Reproduced in Appendix H)
7 REFERENCES

[ALDE00]

[ALLE91]

[ANNI01]

[BALA96]

[BARO99]

[BELA00]
[BHAS99]

[BROU98]

[CAPE03]

[CASE94]

[CHAN01]

[CHAN90]

[CLIF00]
[CLIF01]

[CLIF02]

[CUTK91]

[FROS96]

[GEOM95]

[GRAH01]

[GRIN02]
[HALE80]

[HUSB90]

[JACQ00]

[JACQ97]

[KITA00]

[MCLE02]

[MEDL00]
[MICH92]

[MILL00]

[MILL84]

[MILL93]

[MILL94]

[NAIS97]

[NAIS98]
[PAPA02]

[RAFF96]

[RIDL03]

[SALM97]

[SHER03]

[SRAIN97]

[SURE94]
[VOEL97]

[WOLP95]

[WUND96]
Representation problems in feature-based approaches to design and process planning

F. G. MILL, J. C. SALMON and A. G. PEDLEY

Abstract. A major difficulty in the development of advanced and integrated CAD and CAPP systems lies in the difficulty of representing components. Here some of the problems that are often encountered are discussed and a composite component is used to demonstrate them. Possible approaches for solving some problems are proposed. The conclusion is reached that neither simple feature oriented design nor feature recognition methods alone will fulfill the requirements of advanced systems. Satisfactory modelling of the interactions between features in a component is a prerequisite to progress.

1. Introduction

1.1. Feature oriented detail design system (FODDS)

Current work at Edinburgh is involved in developing a feature oriented detail design system. The objective of this system is to take a feature-based description of a prismatic mechanical component, as produced by a human designer, and produce an enriched description of that component containing all the information required by a CAPP system. The CAPP system in question is a development of HAPPI (Husbands et al. 1990) and not only produces valid process plans, but also performs extensive optimization by using genetic algorithms. In this respect it is unlike Purdue University's quick turnaround cell (QTC) (Turner and Anderson 1988) and other existing CAPP systems (Murray and Miller 1989) for which optimization is not a major goal. This work is also being extended into the domain of job shop scheduling (Husbands 1993).

This requirement for information necessary for optimization has highlighted a number of problems particularly concerned with the interaction of features. The problems are first examined then our approach to them is outlined.

1.2. Features

For the purposes of discussion definitions are given for the feature terminology used in this paper. First, a design feature is defined as a discrete part of a component that fulfills a function in the component's final use. Examples are bearing surfaces, ventilation openings, lubrication grooves, etc.; the list might be endless. However, it might be that in an individual company only a finite number of features are actually used or at least a finite number of identifiable features make up a sufficiently large proportion of the total that it is worth a company's time to adopt a standard feature set.

On occasion several functional features are combined into one design feature. This assumes that multiple actual functions are combined deliberately by the design engineer using his/her expertise.

A production engineering feature is defined as a discrete part of a component that requires shaping in some way and may be made from a previously existing feature. In most cases there is a one-to-many relationship between design features and production engineering features, e.g. a bolt hole might map to a hole with a chamfer and a thread.

Within the single area of production the feature view may be broken down further depending on basic process considerations, for example features for casting or features for machining. In a machining environment individual features cannot be considered in isolation since the cutting strategy consideration often has to take account of the relationship between two or more features. Therefore a feature interaction is defined as the relationship between specific production engineering features. This relationship may be due to proximity, overlap, geometric tolerance or other considerations that are important to the production engineer. A feature has a fixed number of parameters, e.g. diameter and depth possibly, and a feature interaction may describe the relationship between a variable number of features. Clearly a feature interac-
tion has no topology and must be thought of as an entity quite different to design or production features.

The discussion of our approach requires an explanation of primitive design features. These features are deliberately simple, whilst allowing components of complexity to be designed. The primitive design features are shown in Figure 1 and are negative features, i.e. they are all created by material removal. They are also design features that translate simply to the manufacturing domain. A current limitation on design features is that they must have no concavities, this limitation is to simplify some of the interaction analysis software at present.

Feature interactions are the cause of some of the most serious problems in the development of generative CAPP systems and there are problems in specifying them in feature oriented design systems. A distinction should be made between feature interactions that are explicit and those that are implicit. Some feature interactions may be stated explicitly by the designer, e.g. geometric tolerances, but others, such as proximity or obstruction, must be detected. This is not easy since it is difficult to give a universal definition to such implicit interactions. Considerable research has to be done in the area of implicit feature interaction recognition as distinct from feature recognition generally, but the work already done in the latter area will be of help. The use of new more powerful modellers may also help to make progress in this field.

1.3. Feature recognition versus feature oriented design

Research into the use of features as a means of part representation has allowed work to be done on two aspects of representation generation, namely feature recognition and feature oriented design.

Feature recognition consists of software tools that are used to look for form features in representations that are normally created on solids modellers (Jared 1989, Henderson 1984, Srinivasan and Liu 1984, Joshi and Chang 1988, Staley and Anderson 1983). Feature oriented design techniques on the other hand present the designer with a library of features to use to create a model of the part in mind.

Feature recognition suffers from apparent drawbacks when first compared with feature oriented design. First, if a designer were to be working on a system with feature recognition tools, he/she would have to constrain the design in order that it could fit with the feature definitions in existence. In such circumstances it is easy to imagine how a designer might err either by generating something that cannot be recognized by the system or by over-compensating and under-utilizing the options available.

Consider the following truism. In order to recognize a feature one must first have a definition of that feature. If such a definition exists then there is no reason why the designer could not have made use of it in the first instance. It is desirable to give the designer tools that are at as high a level as is possible. Feature oriented design offers a way of achieving this aim. Thus the major advantage of feature oriented design is that it can present the designer with a set of high-level tools that are familiar to his/her natural way of working. Because the feature representation is generated simultaneously with the design, it also has the advantage that it allows the designer to be advised of some manufacturing considerations without actually being constrained by them.

Some early attempts at feature oriented design presented the designer with a set of production engineering features that implied the simultaneous generation of process plans with the design activity (Inui 1988, Case and Acar 1989). This is undesirable because of the differences between design features and production engineering features and because of the feature interactions, which may not be explicit as each feature is added to a design.

2. The Edinburgh composite component

The following is a description of a hypothetical component, which is meant to be used as a test piece for feature oriented design systems, feature recognition systems and advanced CAPP systems. It is not meant to be real and therefore its function is irrelevant, nor is it expected that researchers would be able to put the component straight
through their experimental CAPP system. The hope is that research workers in this area will find the component's complexity of interest and so it will generate worthwhile discussion.

First, individual aspects of the component are described and the reason for their inclusion in the composite component explained. Subsequently the component is presented in its complete form. This reveals further complexities worth consideration. Finally, our approach to some of the aspects first described is given.

2.1. Countersunk holes

There are two points of interest that arise from the example in Figure 2. This shows two lugs having counterbored holes facing in opposite directions. First, how the central holes should be represented (as one or two holes) is a problem and the question of whether the representation would be general (e.g. what would happen if the holes were longer or further apart?) arises. It is also worth considering how the counterbores should be machined.

2.2. Nested slots

In Figure 3, two slots are machined with slot A being cut in the base of the other. Of interest here is how these parts should be represented. It is difficult to define a feature description for a slot that allows slot A to be included. However, if this is done it might then be more complex to determine how to make it than would be the case with a simple slot definition. Also of interest is what anteriority constraints might be generated for these two features and how generally applicable rules for their generation are to all such problems of this class.

2.3. Parallelism

The Faces A, B and C shown in Figure 4 can all be easily cut when considered in isolation, but it becomes
more problematic to consider the machining implications posed when the geometric tolerances describing a close parallel relationship between A and B as well as A and C are taken into account. For example, account must be taken of how these relations are represented and, if they can be, what machining strategies would be used to meet the tolerance. Again there are problems of generality and consideration should also be given as to what would happen if these tolerances were tightened or loosened. In the real world it is often necessary for a process planner to move datum surfaces in order to solve such problems because the initial data make fixturing very difficult. This has profound implications for CAPP research.

2.4. Crossed slots

Figure 5 shows what might appear as two crossed slots, but these features might be modelled in a number of ways, all of which would have implications for the choice of manufacturing strategy. Many questions arise from consideration of features like this, for example, how many features are there? It is also worth comparing any discussion on this problem with that generated for the first example, the countersunk holes.

2.4.1. Crossed holes. There are again similarities between the problem of machining the area shown in Figure 6 and some of those already presented. The important thing about this example is that the holes do not have intersecting centre lines. As a result consideration should be given to the various possible machining strategies and their anteriority constraints.

2.5. Stability

Figure 7 shows three simple features, A, B and C. The difficulty here is in deciding in what order the features should be machined and then considering what would happen if C was longer, or the base of A or B was made longer, or if C was lowered so that its centre line was brought close to the base plane of A and B. Fixturirig and clamping should be considered.

2.6. Cut out

Figure 8 shows a simple pocket with a cut out. The difficulty here is not only representing this area, but it is also intended that the radius of the cut out is sufficiently small to rule out milling as a possible machining method.
2.7. The composite component

The difficulties presented by the interaction of relatively simple features have been outlined by means of some examples. Figure 9 shows a composite component made up of the previously presented examples, each of which display some clearly definable problems. When put together, however, many new interactions become apparent. It is beyond the scope of this paper to discuss them due to their number, but it is an interesting exercise to attempt to define an optimal process plan for the whole part.

The development of this part has taken place as a result of work on the CAPP system HAPPI (Husbands et al. 1990). This system cannot satisfactorily plan the part represented in this paper. It deals with the nested slots (Figure 3) in such a manner that a rigid anteriority rule enforces the upper slot to be cut before the lower one. The problem of stability (Figure 7) can also be handled, but there are many more complex stability
problems that would easily defeat HAPPI. The composite component is useful to consider in planning future versions of HAPPI and it is hoped that other researchers also find it of use.

3. An approach to these problems

Having considered the composite component above, and taking into consideration the requirement of FODDS to deliver information suitable for subsequent optimization, we in particular examine alternative representations as a means for dealing with physical interaction of features.

3.1. Crossed slots

Imagine the designer has designed a crossed slot as shown in Figure 10a and first introduced in Section 2.4. This situation will now be analysed to discover alternative representations. First, the two slots are intersected to discover if there is a physical inter-relationship between the two bodies. Intersecting any two primitive features cannot result in more than one body of intersection. This is a consequence of the primitive features having no concavities.

The two slots will have a cuboid as the result of the intersection (Figure 10b). If the cuboid is then subtracted from each of the two input slots, each of the slots will fall into two lumps, giving a total of four lumps (Figure 10c). Some simple examination of the lumps reveals that each lump can be encompassed by a slot giving the four slots in Figure 10d. For any shape of slots, whenever there are two intersecting slots, the resulting three or four slots will always occupy all and only the volume of the original slots.

3.2. Crossed holes

Another example of interest is that of crossed holes first introduced in Section 2.4.1. Consider the narrow hole and the wide hole in Figure 11. The narrow hole is perpendicular to the wide hole in the $x$-$y$ plane and their axes are offset in the $z$-plane (Figure 11a).

Intersecting the two holes and subtracting the intersection (Figure 11b) from the two original holes results in the narrow hole falling into two lumps (Figure 11c). The wider hole, however, remains unaffected as it is still one lump. Examination of the lumps and the volume of intersection reveals that each new narrow hole has a minimum and maximum depth (Figure 11d). Thus the left narrow hole is either $a$ or $a + \epsilon$ and the right is $b$ or $b + \epsilon$. Also note, however, that the alternative representation is of two narrow blind holes, whereas the initial representation was of one narrow through hole.

3.3. Avoidance of combinatorial explosion

For the simple examples of the crossed slots and holes above, the set of features resulting from alternative representations is only two or three times as big as the set of design features. These additional features only arise due to intersecting design features. However, consider the design of a fixture table consisting of a grid of crossed slots, as shown in Figure 12. If this is considered as $m$ vertical slots and $n$ horizontal slots, then each vertical slot has $2^n$ possible representations, i.e. each intersection can be considered as a break in the slot or not. Thus, the
Distinction must be made between different feature views of products, e.g. design features and production features.

Most difficulties in the production engineering of products are due to interactions rather than the features themselves.

The feature interactions must be satisfactorily modelled if true generative process planning is to take place.

Alternative feature descriptions can be of use for CAPP. Their representation can be achieved in many cases. It is probably worthwhile dealing with those cases that can be analysed easily and recognizing and flagging more complicated cases.

Alternative feature descriptions can be of use for CAPP. Their representation can be achieved in many cases. It is probably worthwhile dealing with those cases that can be analysed easily and recognizing and flagging more complicated cases.

3.4. Guidelines

In order to clarify certain aspects of the process by which alternative representations are generated, the following guidelines have been established for the current work:

(1) Features may be broken down, but not merged (owing to recognition difficulties).

(2) Any feature may only be split into more features of the same type.

(3) Analysis is only performed on pairs of intersecting features.

(4) Any resulting feature must occupy at least the lump from which it was generated.

4. Conclusions

(1) Feature oriented design and feature recognition are insufficient on their own for the complete integration of CAD and CAM.

(2) Distinction must be made between different feature views of products, e.g. design features and production features.

(3) Most difficulties in the production engineering of products are due to interactions rather than the features themselves.

(4) The feature interactions must be satisfactorily modelled if true generative process planning is to take place.

(5) Alternative feature descriptions can be of use for CAPP. Their representation can be achieved in many cases. It is probably worthwhile dealing with those cases that can be analysed easily and recognizing and flagging more complicated cases.

References


Design for machining with a simultaneous-engineering workstation

F G Mill, J C Naish and J C Salmon

During the design of a product many nonfunctional factors may be considered such as the ability to maintain, machine and assemble a product. Those interested in finding ways to support the designer in these tasks have frequently turned to investigations of how expert systems can give advice. Although they are sometimes useful, these approaches are often limited. This is because AI technology is not yet sufficiently flexible and is difficult to maintain, and because consistency problems can arise owing to subtle differences between true-life technology and its representation in a knowledge base. It is the authors' view that many of the most difficult problems facing the designer are spatial ones, and that help could be improved by further geometric-reasoning research. The paper describes the attempts of the authors to address some of these problems.

Keywords: concurrent engineering, feature-oriented design, computer-automated process planning

Many different activities are covered under the heading of design for manufacture (DFM), and before a description of the work undertaken by the authors in this area is given, it is sensible to outline the context of the work. In common with topics such as feature-oriented design, the term 'design for manufacture' means many different things to different people. Discussion of these points is clarified if the terms being used and the envisaged context of the technique's application are clearly defined. Because design itself is an ill defined term it is difficult to describe different aspects of DFM, but some broad categories are the following:

- **DFM where manufacturing methods cannot be known in any detail, possibly, for example, at the conceptual design stage:** At very best, all that may be achieved in this context is the provision of some rules stating good practice, e.g. the designer may seek 'simplicity' in his/her product.

- **DFM where general manufacturing methods may not be known:** In this context the designer may be working at a detailed level either on assemblies or single parts and may not yet know whether the part is going to be made in-house or not. He/she may be given guidance on what kind of parts may be made in-house, however, and these may affect his/her decisions.

- **DFM where the possible production and assembly processes may be known:** Again, rules may be given to the designer, but these can be very detailed and may be in the form of handbooks concerned with, for example, design for assembly¹,² or design for casting. They may also give advice on design-representation methods, e.g. how to use a particular coordinate system or geometric-tolerance methods to facilitate CNC coding. Most research on automated DFM and design for assembly (DFA) has been aimed at putting such guidelines in the form of databases or knowledge bases providing highly context-sensitive advice, normally by coding knowledge in the area in the form of an expert system to assist the designer³.

- **DFM where very detailed manufacturing methods may be investigated:** In this situation, the designer will be involved in the detailed design of components for which he/she is likely to know the manufacturing processes to be used and also the nature of the facilities for producing the part (probably because they are in-house). In the authors' experience, most designers have said that they would know whether they are designing a part which is to be cast or not, and that they may also be able to access the manufacturing department if need be to check on details of how the part will be made (though this is not always the case in practice).

The context of the work described in the following sections is aimed very much at the designer working in the last of the situations described above, and it is felt to be appropriate to the authors' work in CAD and CAPP,
which has been focused on the aerospace, defence and machine-tool industries.

Early work in CAPP led the authors to the conclusion that many of the difficult problems facing designers and process planners were to do with spatial reasoning about machining, where questions regarding tool access and clamping for example would need to be answered. Most of these problems are due to interactions between part features. A second major area of concern is that of ensuring that the designer can be given accurate and timely information at any level of detail that may be required, e.g. about speeds, feeds, times, costs, machine availability and scheduling priorities.

To try to automate these two activities, the authors have been developing algorithms to assist in spatial reasoning for manufacturing problems, and to provide access for the designer to process planning and NC code-generation functions.

It is the authors' view that the most successful DFM method lies in the use of simultaneous engineering where it is appropriate. What follows is a description of the work carried out in this area which includes contributions from several projects and funding bodies. These efforts are brought together in that they have made contributions to the prototyping of what has been called the Simultaneous Engineering Workstation (SEW).

**OVERVIEW OF SIMULTANEOUS ENGINEERING WORKSTATION**

The purpose of the Simultaneous Engineering Workstation (SEW) is to provide an engineer with the tools needed to populate a product model. Thus it is a system for entering designs, investigating cutting and clamping methods, and generating and proving process plans and NC code. This is achieved from one interface, and, because the architecture allows the process to be iterative, a degree of concurrent or simultaneous engineering is possible. The general architecture of the SEW is shown in Figure 1, and the current major modules are described below.

**Man–machine interface**

The man–machine interface (MMI) is the user interface, and it gives a common look and feel to all the system modules. Most of the interaction with the user takes place with features as the objects containing information. When an engineer uses the system, he/she can build up detail designs by using the features the system offers. The engineer can view the design and ask the system to show potential problems such as feature intersections, incorrect feature positioning (causing voids or features in free space), feature proximity (thin walls) and feature obstructions. The engineer subsequently builds up a process plan by using the planning space generator and optimizer. Finally, NC code is generated and proven. If problems are detected that would cause a module to fail, the engineer must return to an earlier module and make changes to the design, plan or code.

The MMI provides a consistent interface for the various modules which have been implemented, and it also provides the same benefits for new modules which are currently planned. Examples include a module to allow the input of features with positive volumes and one which offers the user the chance to develop fixturing plans in parallel with cutting plans. It is also hoped that, in future, it will be possible to add third-party modules to the system to fulfill functions such as feature recognition.

**Feature-oriented detail-design system**

The feature-oriented detail-design system (FODDS) is accessed through the MMI, and it allows the part designs to be entered using workpiece and feature descriptions. The workpiece may be made up of a B-rep model, or it may be built up using features. The main feature description is limited to features with positive volumes. It includes B-rep data with feature descriptions and dimensional and geometric-tolerance information.

The component descriptions are stored and accessed by the other modules using the COmponent Description Language (CODL), a compact text-based language developed by the authors.

**Geometric reasoning**

The geometric-reasoning (GR) module carries out investigations of access, intersection and proximity problems. It takes its input from FODDS, and adds to it descriptions of problems that may be encountered in machining the component. The type of problem detected are described later in the paper in the section on inferring knowledge for design for machining.

**Plan space generator**

The plan space generator (PSG) reads its part descriptions from a file containing a CODL description of a part which has been generated by FODDS and enhanced by the GR module. It uses a knowledge-based system, and its role is to generate every machining method that could feasibly be used for each feature on a part. It does this by accessing a database containing all the possible machining strategies and their constraints that might be applied to a particular feature. The knowledge base is used to deal with problems of feature interaction. These may be due to the problems detected by the GR module or be those due to geometric tolerances. An example is where two features are tied together with a concentricity tolerance. This requires that a machining strategy is
employed to make sure the tolerance is met. The possible machining methods to be used must obviously be a subset of those generated for the features in isolation, and so the effect of the knowledge base is to cut down the set of possible methods used so that those causing problems are discarded. The knowledge base can also generate ordering constraints to ensure that problems are avoided if necessary.

The output of the PSG is the result of an exhaustive search-and-discard procedure. It provides a network of all the possible machining methods that can be employed to meet the part specifications along with anteriority constraints. It gives details of processes, machines, and cutting tools that are needed for each operation as well as speeds and feeds which are read from a machinability database.

Optimizer

The optimizer (OPT) system generates near optimal process plans using genetic algorithms. It takes its input from the PSG, and this is in the form of a file containing a normally very large search space. OPT then tries to find a number of legal plans, i.e. those that meet the constraints determined by the PSG. These candidate plans are encoded in strings in a form that is analogous to that of genetic coding, and this allows the application of genetic operators to them. The optimization consists of a process of breeding using plans to generate offspring, and a selection process decides the fitness of each, and determines their survival by means of a process akin to that of Darwinian evolution. This means that the plans are generally improved upon through successive generations, and, after some time in the simulated environment, where low cost equates with fitness, plans emerge that are substantially cheaper than their forefathers. The total cost can be calculated from individual elements, and it may for example include costs associated with machining, movement between machines, setups or elapsed times of cutting. This method is similar to that used by Vancza.

NC generator

The NC generator module is used to generate NC code automatically. It reads in process-planning information using the Process Planning Description Language (PPDL) as well as bodies of tools, machines, fixtures and clamps, and it outputs NC code. If need be, the engineer can see simulations of cutting.

CODL, process-capability representations and Process Planning Description Language

These are not modules but methods of storing information about the part, blank and tooling. The role of CODL has been described above, and so we shall consider the process-capability representations (PCRs) and the PPDL.

Part of the PCR is implemented in a relational database, and contains technical and geometric information about tooling. The database contains tables with information on standard tooling data, e.g. about what drills are available and what their flute lengths, diameters and ID numbers are. Information on machine tools and fixturing equipment may also be stored. In addition, B-rep models of tools, machine tools and fixturing are also held. These models may be used to check access during the design process if the designer wishes to investigate the use of a particular tool or machine. The information can also be used by the NC system to build models of setups so that NC code can be automatically written without generating tool/fixture or tool/machine collisions.

The PPDL is the language which we use to represent process plans. It is output by the NC system along with the NC code itself.

INFERRING KNOWLEDGE FOR DESIGN FOR MACHINING

When producing process plans from a feature-based design, it quickly becomes apparent that producing an individual feature plan for any manufacturing feature in isolation is a comparatively simple exercise. The difficulties most often arise when features interact with each other or the blank or finished component in some way. We initially class these interactions into two types, explicit and implicit feature interactions.

Explicit feature interactions (EFIs) are those interactions that are both required and specified by the designer. These are expressed as relational tolerances between features or the blank, and they may include concentricity and parallelism.

Implicit feature interactions (IFIs) represent knowledge about the component that was not made explicit by the designer but that is required when considering methods of manufacture. Some of these IFIs may be known to the designer but not explicitly stated, whereas others may be things of which the designer is unaware but should be (two features that accidentally intersect may dramatically alter functionality), or of which the designer is unaware and largely unconcerned about as they represent process-planning problems. It is these implicit feature interactions and their recognition that we shall discuss in greater detail here.

Implicit feature interactions

We recognize these IFIs using geometric techniques that we term implicit-feature-interaction recognition algorithms (IFIRAS).

IFIRAS are employed at a preplanning stage. That is, no details as to tool, machine or setup orientation are known when the IFIs are recognized. However, many of the IFIRAS employed do make assumptions (e.g. 2D machining).

We perform IFIRAS to check for a number of distinct problems, namely

- void detection,
- feature presence,
- access-problem detection,
- feature intersection,
- feature proximity.
Void detection
Void detection is largely a design-validation check. It is performed as the finished body is being generated from the blank and negative features. Our chosen domain using traditional machining techniques does not allow us to machine voids in a component. Voids can be easily detected by counting the number of shells in the B-rep model as the features are entered into a component. The designer is merely warned that a void exists, as a feature that is later added to a component may link shells and so remove the problem. Void detection is a useful validation technique, but it can also be regarded as a primitive form of access-problem detection. A feature producing a void cannot be accessed from any direction.

One further complication of void checking is that the initial blank could, in theory, contain a void. A feature added to the workpiece that was 'machined' into the surface of this void would not change the number of voids, and hence it would bypass the checking, but it would still be unmachinable. Blanks containing voids are, however, a rarity.

Feature presence
The feature-presence check is another design-validation check, and it is almost the converse of void detection. Here, we check that every feature at least partly intersects with the blank to ensure that the feature is likely to change the geometry of the blank. Those features not present in the blank are discarded as no machining needs to be performed.

A more rigorous version of feature-presence checking has been proposed whereby any feature that is contained completely within another feature can also be discarded. This has its drawbacks, however. At the moment, we are able to deliberately hold multiple representations of feature combinations, for instance a hole split by a slot may be represented both by a single hole passing though the slot and by two holes each terminating at the edges of the slot (see Figure 2). The authors have encountered just such a case, where, in order to machine to a required concentricity tolerance on a pair of holes separated by a previously machined, wide slot, the holes were first drilled individually, and finally reamed in one operation. Thus both representations of features must be available simultaneously on the one component.

Additionally, we foresee a time when we would like to allow the design of features with temporal constraints. Consider a situation where a lug is on a blank to allow locating or clamping for further machining operations (say the smooth milling of a large surface area). The lug is then machined away before completion of the component. The lug has only a transient presence, but it must be retained until the surface area has been machined.

Access-problem detection
In the FODDS system, the designer inserts volumetric subtractive features into the workpiece in order to model the finished component. Features may be nested, that is, the accessibility (for machining) of a particular feature may be dependent on other features having been machined prior to it. These ordering or anteriority problems are searched for by producing 'access bodies' for each feature. These access bodies depend on both the feature geometry and anticipated machining methods. Typically, these access bodies are produced by projecting the top surface of the feature away from the feature. The resultant access body represents a notional tool of semi-infinite length (from the feature out to infinity). In addition, we grow our access bodies by a small amount sideways so as to allow for tool waver on approach (see Figure 3).

These access bodies are then used to check for access problems of three types:

- access problem with the blank,
- access problem with the finished component,
- access problem with another feature.

If an access body does not intersect with the blank, then we know that there is no access problem for that access vector with that feature, and the additional tests on that feature and access vector can be foregone.

If an access body intersects with the finished component, then we know there is a potential problem...
with machining that feature using that access vector. At the present time, without further reasoning, such as trying particular tools and machines, we assume that the feature is unmachineable from that approach vector.

The access body for a feature is then tested for intersection with all other features (see Figure 4). The resulting information is used as a first pass of the algorithm for producing anteriority constraints between features. Thus, given no information to the contrary, features whose access bodies intersect with other features have a constraint placed on them so that they are machined later than the feature they intersect with.

An additional class of access problem concerns the tool-clamping device. To identify such circumstances, for each feature, a virtual tool-clamping device is produced (e.g. a chuck for a drill). This body represents a worst-case scenario for machining a feature, as the maximum size of clamping device is assumed. It is generated by offsetting the edge of the top face of the feature by a distance R, where R is the radius of a worst-case tool clamp. If the tool to be used were known, then the offset could be R-r, where r is the radius of the tool; however, normally, the tool size is not known.

Collisions between the virtual tool clamp and the workpiece indicate that a feature may be difficult to produce with a short tool. Further reasoning can then be undertaken to determine where it is safe for the tool clamp to move to (see Figure 5).

**Feature intersection**

Feature intersections almost always indicate areas of process-planning interest. We discover all the cases where features intersect with any other feature. These cases having been discovered, reasoning must be performed to evaluate the intersection and take appropriate action.

From the intersection, we can:

- identify potential machining problems,
- derive alternative representations,
- add information regarding the required and optional volumes of a feature.

The intersection in Figure 6 of the two crossed holes illustrates all of these factors. There is a potential machining problem in that, if the wide hole is machined first and then the narrow hole is drilled as one feature, the drill is likely to bend and break on the inside surface of the wide hole. This can be resolved in two ways (in this case), by either machining the narrow hole first in one operation, or recognizing that the narrow hole can be represented by two shorter narrow holes drilled from either end of the blank. This may require an additional concentricity tolerance between the two new narrow holes. When it is represented in this way, we see that the intersecting portion of the two original holes need not be part of two new narrow holes, as the intersection volume will certainly be removed when the wide hole is machined. Thus the intersection volume is an optional volume for the narrow hole(s).

**Feature proximity**

The proximity of features is an important consideration when analysing the likelihood of cutting forces distorting the workpiece. Feature-feature proximity problems can be detected by growing all the features in all directions. If we are looking for wall thicknesses of less than a thickness t, we grow all the features by a distance t/2 and intersect all the 'grown features' with each other (see Colour Plate 1) (Vandenbrande\(^8\) proposes this technique without implementing it).

Two major questions now arise. How can thin walls between a feature and the workpiece be detected?
To select tools and machines according to geometric constraints, a representation of each available tool must be held in the database. Cutkosky\textsuperscript{9} and Chang\textsuperscript{10} discuss the limitations of holding only 'process-level' knowledge, and identify the need for 'tool-level' representations to be implemented in future. Detailed descriptions allow specific machine–tool combinations to be offered as solutions rather than just identifying a process type. Such a description would be used for further geometric reasoning if an IFIRA identified a potential problem such as a tool collision. A checking function as part of the tool representation would generate a solid model of a possible 'worst-case' tool to 'try out' on the component solid model as directed by the process planner.

For example, the hole-access problem shown in Figure 5 restricts the choice of drills to those of length greater than $H$. The existence of such an access problem will previously have been identified by an IFIRA and indicated in the CODL file. However, to select tools for this feature, further geometric information about the position and size of the collision body is needed. In this case, a binary search by length among the set of possible tools, previously defined by explicit restrictions, is used to identify the lower limit on the tool length $L_{\text{min}}$ imposed by this constraint. This drill is tested by generating the drill body and the spindle body of the chosen machine (or a 'worst-case' virtual body if a machine has not yet been selected) in the solid modeller. These bodies are positioned in the component model, and an intersection IFIRA is used to check for collision.

The solid-model representations of tools and machines are used again later in the NC-code verification module. As the spatial-reasoning needs for process selection are identified, new algorithms will be added and tool and machine descriptions extended to enable decisions to be made for process selection in response to the additional information in the enriched component description. The solid-model representations of tools and machines are used again later in the NC-code verification module.

Cutting parameters used by Mandelli are stored in a database and are used for cost calculations for each machining operation.

### Process-capability representation

The representation of process capabilities available to the CAPP system is being extended in response to a study of process-planning methods at Mandelli SpA, a machine-tool manufacturer. The modelling of the geometric constraints of processes and more detailed representations of individual tools and machines were identified as significant aims.

The planning space generator requires a set of all the possible manufacturing solutions for each feature (or related features) in the component, each with an associated cost. All the available tools and machines are represented in a technological database, and a set of machine-tool-setup combinations is selected according to feature type, tolerances and implicit feature interactions. The combined (cumulative) errors of machine and tool are taken into account in selecting processes to achieve tolerance restrictions.

### Process planning and NC-code generation

Process planning can be performed in two distinct ways. During design, the designer may ask the process-planning system to plan particular features so that the designer can investigate alternative design solutions. Specific tools or machines can be imposed on the process planner if required for individual features. The rest of the component can then be planned automatically.

The automatic process planner (HAPPI)\textsuperscript{8} takes a feature-based description of a component and looks for a set of possible machine-tool-setup combinations for each feature from the database. Using these, the planning space generator generates a search space of possible process plans, each with an associated cost. These costs include machining costs, tool-change costs and set-up costs. Machining times and hence costs can be accurately produced from the CL-data file produced by the STRATA-based NC-code generation package. It is hoped it will also be possible to verify the feasibility of solutions both interactively at design time and during automated process planning.
CONCLUSIONS

- Most automated DFM systems have been developed using KBS approaches with production-rule systems.
- Much of the reasoning that facilitates the provision of DFM information for metal cutting involves a considerable amount of 3D spatial reasoning.
- There are considerable advantages to be gained by directly coupling CAD DFM modules to CAPP software where this is possible.
- The more concurrent a system is, the greater the potential is for the direct coupling of software modules for CAD, CAPP fixture planning, and NC-code generation, and this allows for the greater development of DFM facilities.

ACKNOWLEDGEMENTS

The authors would like to thank the ACME directorate, UK, the BRITE/EURAM initiative and the Royal Society of Edinburgh, UK, for funding the work. Considerable thanks are also due to the bodies who have collaborated directly on the research. These include the University of Sussex, UK, Hannover University, Germany, GEC Ferranti, UK, Mandelli SpA, Italy, Pafec, UK, Spatial Technology, UK, and Strässle GmbH, Germany.

REFERENCES


Frank Mill graduated in 1983 from Napier College, UK, with a BSc in technology with industrial studies. After working at Coventry Polytechnic, UK, as both a research technician and research associate, he took up a lecturing post in the Department of Mechanical Engineering, University of Edinburgh, UK, in 1988. His research interests include computer-aided process planning, feature-oriented design, concurrent engineering, and genetic algorithms. He is currently a senior lecturer and the director of the Manufacturing Planning Group.

Jane Naish received a BEng from the Department of Mechanical Engineering, University of Edinburgh, UK, in 1991, and she has since worked on feature-oriented design and process capability representation. Recent research has concentrated on PCR and geometric reasoning for machining, with a study being carried out at the machine-tools manufacturer Mandelli, Italy.

Jonathan Salmon graduated from Imperial College, London, UK, in 1988 with a BSc in physics, and he then joined the Department of Artificial Intelligence, University of Edinburgh, UK, where he gained an MSc in information technology. After working in industry, he joined the Manufacturing Planning Group in 1990 as a research associate to work on feature-oriented design for process planning. His research interests include geometric reasoning and solid modelling.
An industry-based study of cutting process capability representation requirements for an integrated simultaneous engineering workstation

JANE C. NAISH¹, FRANK G. MILL² and JONATHAN C. SALMON²

Division of Integrated Systems, School of Engineering, University of Abertay Dundee, Kydd Building, Bell Street, Dundee DD1 1HG, Scotland, UK

Manufacturing Planning Group, Department of Mechanical Engineering, The University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JL, Scotland, UK

Received June 1995 and accepted December 1995

The drive to integrate process planning with design systems to achieve concurrent engineering places new demands on the knowledge used in automated process selection. This paper highlights the current fragmented and unstructured nature of cutting process capability representation. Advances made in developing complementary models needed to analyse spatial process constraints are described in the context of an industrial case study and a simultaneous engineering workstation (SEW). The SEW uses a library of volumetric removal features to describe the shape-producing capabilities of processes. Attempting to model and plan sample parts on site showed up some problems in mapping from features used in design to the manufacturing processes, identified implicit interactions between features that needed to be recognized and showed a need for more detailed process error modeling. To overcome these limitations, geometric reasoning algorithms and complementary tool and machine models have been developed.

1. Introduction

Process capability modeling has in the past covered a wide variety of activities. Analytic methods of predicting the behavior of the cutting process and the tool have long been used to predict cutting operation requirements as defined by power, tool life and surface roughness constraints. These are, and will continue to be, important areas of research. Faced with estimates of the power requirements, cutting speeds, and tool life, the engineer can turn to data about processes, machines, and cutting tools to help him select suitable cutting conditions. Large companies often provide data in the form of handbooks or in databases that describe the general capabilities of processes in terms of tolerances. Supplier's data on size, power, or error may also be supplied in these forms and will be augmented by user's data, usually stored in the head of the production engineer.

In attempting to partly automate operational planning the restrictions of the methods mentioned have become important. First, the available tools are fragmented and the lack of structure of an approach makes them difficult to automate (and even to teach). Secondly, they are often insufficient on their own; more complex models are often needed to properly predict, for example, the actual tolerance achievable.

Many researchers have made advances in modeling the shape creation process with notional geometries [1-6]. Although they are useful, these methods need complementary models to analyse spatial considerations if further progress in the automation of operations planning is to come about. Process capability representation (PCR) models developed so far for computer-aided process planning (CAPP) have been reviewed by Cutkosky et al. [7] and Chang [8]. Chang [8] identifies the scope of process capability knowledge as encompassing shape-producing capabilities, dimension, tolerance, and surface properties capability, geometric constraints (such as tool collision with the workpiece), technological constraints (due to power consumption and cutting force), and economic constraints. The discussion of future needs by Cutkosky [7] concludes that for geometric, kinematic, dynamic, and deformation analyses to be possible there is a need for more detail to be included in models. These papers indicate that future process capability models should make available a fuller breadth and depth of knowledge than previously implemented.

In most CAPP research, the shape-producing capabilities of a process are modeled with volumetric features such as slots, holes, and pockets. Manufacturing information about how to make these features is encoded in a knowledge base. The level of detail varies from general
Two ways of encoding manufacturing knowledge about features are identified by Shah [3]:

1. Associate each machining process with a list of features that it can produce, then for a given feature; find a candidate process;
2. Associate each feature with processes (and sequences of processes) then match the capabilities with the feature specification for each instance of the feature in the component.

The QTC knowledge base [4] uses geometric reasoning to translate design features to manufacturing features. Termed feature refinement, this method adds information needed for manufacturing planning, such as feature classification, relationships, and feasible approach directions, to the feature description. Manufacturing methods are modeled by using a development of the frame-based method by first choosing a set of feasible processes, then applying tool selection rules that synthesize tool database queries.

Problems can arise in using rule bases to select processes. Sometimes rules conflict, and meta-rules must be introduced to decide which rule can override others. The simultaneous engineering workstation (SEW) process capability rules do not attempt to find a single manufacturing solution for a single feature but supply the optimizer with a set of valid solutions for each given feature or related features. A typical conflict in trying to write process planning rules is that one rule may give as the ‘best’ process plan one where the cheapest machine is always used in preference to the others, but another rule may always choose processes that avoid a change in setup. Neither of these approaches is wrong and it is hard to decide which should override the other. In the Edinburgh CAPP system, HAPPI, conflict is resolved by using the genetic algorithm approach, which will include the best parts of each plan produced by the process planning rules in the optimal plan that it generates [9]. This system is described more fully in the next section as it is integrated with a Feature-oriented detail design system (FODDS) within the SEW developed at Edinburgh [10].

Process model and geometric reasoning algorithm development should be guided by the knowledge required by the applications that they support. Integration of the FODDS and CAPP systems introduced the need for enhanced knowledge modeling and this need was investigated by testing a prototype integrated system in a manufacturing company. Typical components were used to study design methods and planning decision-making and the system’s ability to model the component designs and manufacturing processes were assessed. The following section gives a brief overview of SEW. Process capability issues arising from the industrial study are then discussed. The discussion concentrates on problems in the use of features to model shape-producing capabilities and on the development of these models to enable collision checking and more detailed error analysis. Constraints such as tool access obstruction and feature relationships such as parallelism are not yet considered by most current CAPP systems, although they have been written about.

To allow reasoning about these constraints in SEW the process capability models have been extended in response to this study.

2. Overview of the SEW

The purpose of the (SEW) is to provide an engineer with the tools needed to populate a product model. Thus it is a system for entering designs, investigating cutting and clamping methods, and generating and proving process plans and NC code. The general architecture of the SEW is shown in Fig. 1. The current major modules are described below, and then the modules that involve process capability knowledge are described in more detail.

The man–machine interface (MMI) gives the engineer a common look and feel to all the system modules. Most of the interaction with the user takes place by using features as the objects containing information. When an engineer uses the system he can build up detailed designs by using the features that the system offers him. He can view his design and the system to show him potential problems such as feature intersections, incorrect feature positioning (causing voids or features in free space), feature proximity (thin walls), or feature obstructions. The engineer subsequently builds up a process plan by using the plan space generator (PSG) and optimizer (OPT). Finally NC code is generated and proven. If problems are detected that would cause a module to fail, the engineer must return to an earlier module and make changes to his design, plan or code.

2.1. FODDS

FODDS allows the part designs to be entered by using workpiece and feature descriptions and is accessed through the MMI. The blank may be made up of a b-rep model or a limited set of primitives. The main feature description of the component is limited to features with negative volumes. It includes b-rep data, feature descriptions, and dimensional and geometric tolerance information.

The component descriptions are stored and accessed by the other modules by using the component description language (CODL), a compact text-based language developed by the authors. This is generated automatically from the feature-enhanced solid model.
Cutting process requirements for integrated workstation

Fig. 1. General architecture of simultaneous engineering workstation (SEW).

2.2. Geometric reasoning (GR)
The GR module carries out investigations of access, intersection and proximity problems. It takes its input from FODDS and adds to it descriptions of problems that may be encountered in machining the component. The types of problem detected are discussed in Section 3.

2.3. Plan Space Generator (PSG)
The PSG reads its part descriptions from a file containing a CODL description of a part that has been generated by FODDS and enhanced by the GR module. It uses a knowledge-based system and its role is to generate every machining method that could feasibly be used for each feature on a part. It does this by accessing a knowledge base containing all the possible machining strategies and their constraints that might be applied to a particular feature. The knowledge base is used to deal with problems of feature interaction. These may be due to the problems detected by the GR module or those due to geometric tolerances. An example is where two features are tied together with a concentricity tolerance. This requires that a machining strategy is employed to make sure that the tolerance is met. The possible machining methods to be used must obviously be a subset of those generated for the features in isolation, so the effect of the knowledge base is to cut down the set of possible methods. Those that cannot meet the constraints are discarded. The knowledge base can also generate ordering constraints if necessary.

The output of the PSG is the result of an exhaustive search-and-discard procedure. It provides a network of all the possible machining methods that can be employed to meet the part specifications also with anteriority constraints. It gives details of processes, machines and cutting tools that are needed for each operation as well as speeds and feed rates, which are read from a machinability database.

2.4. Optimiser (OPT)
The optimizer (OPT) system generates near-optimal process plans by using genetic algorithms (GAs) [11]. It takes its input from the PSG in the form of a file containing a normally very large search space. The search space is made up of the set of all the feasible manufacturing solutions, described in terms of tool–machine–setup combinations and an associated cost for each feature on the component. The OPT then tries to find a number of legal plans, i.e., plans that meet the constraints determined by the PSG. The optimization consists of a process of breeding by using plans to generate offspring; a selection process decides the fitness of each and determines their survival in a means akin to Darwinian evolution. This means that the plans are generally improved upon through successive generations and after some time in the simulated environment, where low cost equates to fitness, plans emerge that are substantially cheaper than their ancestors. Total cost can be calculated from individual elements and may include costs associated with machining, movement between machines, setups, or elapsed times of cutting, for example.

This approach was first suggested by Husbands et al. [9] and a similar method is used by Vancza and Markus [12]. One of its attractions is that conflicts such as those
between heuristics, for example 'minimum number of setups' or 'cheapest machine', are effortlessly resolved in an optimum. This is because the GAs are entirely cost-driven.

2.5. PCR database

The declarative knowledge part of the PCR is implemented in a relational database and contains technical and geometric information about tooling. The database contains tables with information on standard tooling data (e.g., available drills and associated flute lengths, diameters, and IDs). Information on machine tools and fixturing equipment may also be stored. Sufficient information to generate b-rep models of tools, machine tools, and fixturing is also held. Rules used to select sets of feasible tools and machines are implemented in the PSG and can access the GR. These rules are the procedural knowledge part of the PCR.

In the relational database details of every tool instance are stored in a tool table (Table 1). The shape-producing capability is represented by associating features with tool types that can make the feature (Table 2). Each tool can then be linked to the features it can make by its tool type ID. Employing a tool type ID in the representation allows a combination of the two representation methods identified by Shah and Hsiao [3]. Primarily it is the second approach that is used, where each feature has machines and tools associated with it. The representation is flexible enough so that a tool added to the database can be associated with features it can make by using its tool type ID.

For each tool, values such as tool length, diameter, end angle, and machining errors are stored. These values are used by the rules for selecting possible tools to make a feature. In addition these values can be used by the GR module to provide parameters to a function to generate a solid body of the tool in ACIS if extra geometric information about the tool's capability is needed.

### Table 1. Tool instances in tool table

<table>
<thead>
<tr>
<th>Tool type ID</th>
<th>Tool instance</th>
<th>Length</th>
<th>Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drill</td>
<td>d1</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>Drill</td>
<td>d2</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>Drill</td>
<td>d3</td>
<td>70</td>
<td>20</td>
</tr>
<tr>
<td>Mill</td>
<td>m1</td>
<td>50</td>
<td>10</td>
</tr>
</tbody>
</table>

### Table 2. Associate tool type ID with feature

<table>
<thead>
<tr>
<th>Feature</th>
<th>Tool type ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hole</td>
<td>Drill</td>
</tr>
<tr>
<td>Hole</td>
<td>Mill</td>
</tr>
</tbody>
</table>

Similarly, machine tools are modeled in terms of the features they can make and the tool types they can use. Technological information is stored for use by the rules implemented in the PSG.

Each potential manufacturing solution for a particular feature consists of a series of processes defined as a 'microcycle'. In the database there is a microcycle table for each feature. This table consists of five process columns filled with the tool type IDs that can perform that process. Further columns hold general process capability information for each microcycle so that manufacturing methods identified as unsuitable can be discarded without considering actual tool or machine instances. Examples of microcycles for holes are shown in Table 3.

### 3. Issues arising from assessment of SEW software at industrial site

A study was performed at the Italian machine tool manufacturer Mandelli SpA to confirm that the system could be used to model and plan typical parts in the company and to identify what manufacturing knowledge would need to be available to produce feasible process plans. Mandelli supplied three sample components that contained design and manufacturing problems that typically occur in machine tool components.

**3.1. Representation of sample components in terms of features**

The SEW feature set was based initially on a general feature set of slots, holes and pockets and was developed in collaboration with Marconi. More company-specific features were added for modeling the Mandelli components and turned components from YMCA-NC, which manufactures flexible couplings.

The feature set was tested by studying the three sample components and trying to describe them in terms of the features available in FODDS. The following discussions illustrate problems that can arise in the mapping of features used in design to the corresponding processes that can make them. These problems are discussed in the

### Table 3. Examples of microcycles for holes

<table>
<thead>
<tr>
<th>Microcycle name</th>
<th>Preparatory process</th>
<th>Volume removal</th>
<th>Semi-finishing</th>
<th>Finishing</th>
<th>Super finishing</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Spot drill</td>
<td>Drill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>Spot drill</td>
<td>Drill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>Drill</td>
<td>Rough ream</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>Drill</td>
<td>Bore</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>Spot drill</td>
<td>Drill</td>
<td>Bore</td>
<td>Bore</td>
<td>Burnish</td>
</tr>
</tbody>
</table>

C-5
context of tool and machine tool selection with the search-and-discard method described in Section 2.3.

3.1.1. The parallelepiped feature

This feature was introduced into the feature set to represent a surface to be milled that is considered common enough on Mandelli components to justify an additional feature. The component shown in Fig. 2 is a problem from a version of a Mandelli gearbox housing. A representation problem typical on components in a casting environment is that of a surface that is to be milled. Such a surface is the top surface on the Mandelli component. In a stand-alone design package the designer would indicate this milled surface by attaching a surface finish value to that surface. For an integrated system, however, this surface must be linked to manufacturing processes and therefore a feature must be used to describe it.

Given the initial feature set, the obvious solution is to represent the milled surface by a pocket feature, which already exists in the feature library (see Fig. 3). However, the pocket feature implies that some or all of the milled surface (the base of the pocket) has an adjacent side wall that is also milled. The use of a pocket to represent a plane milled surface, as in this case, implies that the designer requires a curved wall in the pocket shape; this would be complex to mill. In fact all the designer requires is that the base of the pocket he put in his design gets made.

This argument led to the inclusion of a feature called parallelepiped in the feature set. A parallelepiped is a pocket feature where only the shape of the base surface is important. The set of solutions for the pocket feature could be seen as a subset of the solutions for making the parallelepiped. Because only the base surface is generated for the parallelepiped the tool can move outside the feature volume without risk of collision with any walls. Fig. 4 shows how the CAPP system would ‘see’ the shape it should generate as a plane plus four walls (and rounded corners), this results in tool motion being unnecessarily restricted and would result in an overly complex toolpath and a restricted choice of possible tools. A parallelepiped used in the model as shown in Fig. 5 would allow a much larger set of processes to be considered, many of which would be more appropriate.

It is desirable for the system to map from a pocket to a parallelepiped if the designer used a pocket feature in the way shown in Fig. 4. The mapping algorithm checks for the existence of walls around the edges of the pocket base and maps to a parallelepiped if none are found.
3.1.2. The ring feature

Initially the ring feature was included in the feature library for modeling turned parts at VMA-NC. A simplified version of a typical VMA-NC component is shown in Fig. 6, where the ring represents volume removed from the cylinder, typically by turning. Ring-shaped volumes are also used to model Mandelli components, for example Fig. 7, where the ring feature is like a circular slot.

These two possible uses of the ring in design initiated a change in the way in which tool access and motion were modeled for the processes able to manufacture the ring. Both turning and milling processes can remove a volume of this shape. Which of these processes is appropriate for a given feature depends on the tool access directions available, on the size and shape of the general component, on the fixturing available, and on the methods used for other features. The GR provides a check for tool access directions and provides information in the CODL description on which of these is available. The CAPP system can then use this information to map to turning or milling operations.

3.1.3. Mapping between ring, hole and groove features

Another mapping problem arose in the design of a sample Mandelli component. In this component a hole exists in the casting that the designer uses as a blank. The designer wishes the hole to have a higher surface finish value than the cast surface of the blank and therefore requires a finishing operation to be performed. As discussed for the parallelepiped, a design-only system could allow the designer to attach a surface finish value to the surface in the model. In this system, features are used as part of process capability representation for CAPP, so a feature must be used. The volume of material to be removed is shown (thickness t exaggerated) in Fig. 8; the difficulty arises in deciding which feature should be used to represent it.
Generally, the ring is used when the perceived manufacturing access direction is parallel to the axis of the ring, whereas the groove is used when the access direction is perpendicular to the axis of the groove. To the designer, however, the features are differentiated by the surfaces the designer wishes to create (e.g., for a groove we wish to create two parallel and one cylindrical surface).

The volume to be removed looks like a ring feature (Fig. 7) or a groove feature (Fig. 9). It should be modeled as a hole feature, however, because in this case there is no material left 'inside' the ring forming a boss, and the groove implies that three surfaces are to be generated as shown in Fig. 9. Cases 1–4 explain the tool selection consequences of using these features.

Case 1
Using the groove (Fig. 9) would mean that machining methods would be found to generate a groove in the cylindrical wall of the hole in the casting. In this case the upper and lower surfaces perpendicular to the cylindrical surface in the groove feature would indicate that a milling cutter could not be used to remove material from the cylindrical surface. For this reason the use of a groove to model the volume would be inappropriate.

Case 2
If a ring were used to model this volume, milling cutters of diameter less than or equal to \( t \), the width of material to be removed, would be chosen by the rule base. The width \( t \) is the difference between the outer and inner diameters of the ring. In Case 2, \( t \) is the depth of finishing cut required to achieve the high surface finish specified by the designer. This would typically be 1 mm or less. If the system tried to find cutting tools of diameter less than or equal to \( t \), no suitable tools would be found to manufacture the ring.

Case 3
Even if the designer were aware of this problem and made \( t \) larger so that some tools would be selected, the implied presence of an inner boss would still restrict unnecessarily the tool set considered. Milling cutters with smaller diameters are less rigid. If a strict straightness tolerance had been specified for this feature some or all of the tools would be rejected because they would deflect outside the tolerance zone for the feature.

Case 4
In case 4 the hole feature is used for the design and a large set of possible tools are selected initially. Any tolerance constraints are then applied to reduce the set; in the case of a strict linear tolerance the tools of smaller diameter would be removed from the set.

A designer may not wish to concern himself with the manufacturing implications of using one of the ring, the groove, or the hole in such a case, when using any feature results in the same shape. To free the designer in this way but still produce valid plans some geometric reasoning is needed in mapping from the ring to the hole, or groove to hole, so that all possible manufacturing solutions are considered. The GR algorithm checks for the existence of a boss inside the ring and for access directions. If no material is found then all hole manufacturing processes can be considered.

Unfortunately using a hole in this way introduces other problems. A high material removal rate (MRR) process is not required because the hole already exists in the blank. For this reason the preparatory process (e.g., spot drilling) and the volume removal process (e.g., milling) of each microcycle can be disregarded. This leaves the problem that the selection of tools for semi-finishing and finishing usually depend on the size of the tool used in the previous operation. If these roughing operations are discarded, we need to use an 'intermediate feature' representation. A feature between, say, roughing and semifinishing in a microcycle is termed an intermediate feature [9]. The representation of intermediate features ensures that valid tools are selected for finishing a hole that exists in a blank. The dimensions, surface finish, and tolerances of the hole must be used in the representation of the intermediate feature.
3.1.4. Implicit and explicit interactions between features

Considering how a feature may be machined in isolation is a relatively straightforward exercise. When a feature is related explicitly or implicitly to other features, this relationship influences the possible manufacturing solutions. Feature relationships are discussed further in Mill et al. [13]. An explicit relationship, stated deliberately by the designer, might be a parallel tolerance between two parallelepipeds; an implicit relationship might be the intersection of two holes.

We consider an example from one of the Mandelli sample components (but typical of cast components generally). Two walls are modeled by using parallelepipeds with a tight parallelism tolerance between them. The following describes rules for tool selection for related features generally.

Given two related features, A and B, on a component, the rules search for a set of tools to make each feature as though the relationship did not exist. \( T(A) \) is defined as the set of tools that can make feature A, and \( M(A) \) as the set of tools that can fit machines that can perform the necessary transforms for A. The intersection \( T(A) \cap M(A) \) is the set of tools that can be used to make feature A. In the same way the set of tools \( T(B) \cap M(B) \) is found for feature B.

The existence of the relationship between the features means that rules must be defined for finding sets of tools that can satisfy the relationship. The following rules have been defined.

**Rule 1. Avoid change of tool or machine for machining of features A and B.** This gives the set of feasible tools \( S_1 \):

\[
S_1 = T(A) \cap M(A) \cap T(B) \cap M(B).
\]

**Rule 2. Avoid a change of machine but allow a change of tool.** This gives the set of feasible tools \( S_2 \):

\[
S_2 = (M(A) \cap M(B)) \cap (T(A) \cup T(B)).
\]

**Rule 3. Use the same tool on different machines.** This gives the tool set \( S_3 \):

\[
S_3 = (M(A) \cup M(B)) \cap (T(A) \cap T(B)).
\]

The value of the tolerance set will dictate which of these rules is applied. Rule 1 is applied for the strictest tolerance relationships and is thus likely to result in a small set of feasible tools. Rule 3 is very unlikely ever to be needed, so it is not implemented.

It is likely that a tight relational tolerance would require machining of the two features in the same setup. This rule would be applied to the tool sets already found by rule 1 or rule 2 and a subset of tools found. A difficulty arises in taking the setups into account because at this stage the fixtures have not been chosen.

Geometric constraints and error models are discussed in the next section. The rules that use this information operate on the sets of feasible tools in the same way as described above so that the set is gradually reduced by the introduction of more constraints.

3.1.5. Recognizing implicit feature interactions

The GR module of the FODDS system is able to find implicit relationships that influence process selection for a component model; among these are:

- feature proximity,
- feature intersection, and
- tool access problems.

A full description of these can be found in [10]. These algorithms serve to identify potential machining problems in the part but do not provide sufficient information to enable decisions to be made about solutions to such problems. The task at the collaborating companies was to study typical components to look for implicit feature interactions and potential tool access problems, and discuss with process planning engineers the various manufacturing solutions that would be used to avoid machining problems. These discussions helped to define what further geometric information would need to be added to the process capability representation to allow such decisions to be made. Some examples of these discussions are detailed below.

**Tool obstruction** (Fig. 10). In this case the obstruction does not prevent access but the distance that a boring bar would have to reach would cause excessive vibrations. A rough reamer is rigid enough to overcome the problem.

![Fig. 10. Implicit feature interactions: tool obstruction.](image)
Cutting process requirements for integrated workstation

3.1.6. Process error model enhancement

The following models are being developed to augment the existing PCR’s error models so that decisions can be based on tool and machine level error models, rather than on the less accurate general errors associated with process types.

The tool profile models described here are used in high-level general tool access checks, before the actual tool and machine to make the feature have been chosen. At a more detailed level of analysis they can be used in the generation of tool instance models and error models for tool level selection. A situation where such extra information may be required is described in Section 3.1.4 above. Here, a ‘cut-off’ tool must be found to try out for tool access on a component as a limit to qualify a subset of tools. At this level of detail the shape information held is used in forming cumulative error models also described below.

The set of features implemented in the SEW can all be produced by a three-axis milling machine. This set supports the geometry of the vast majority of components in the domain of 2½-dimensional (2½D) components. Sungurtekin and Voelker [14] list the trajectories used in three-axis machining as:

1. linear in $z$ (the spindle direction),
2. linear in the $xy$ plane,
3. linear in space (simultaneous movement of the three axes),
4. circular in the $xy$ plane, and
5. circular in the $xz$ and $yz$ planes.

We first consider only features using trajectory types 1, 2, and 4. The reasons for this are now outlined.

If we take a theoretical look at 2½D machining techniques, we see that many have a rotating tool that is then driven along some path. The tool generally follows some approach and retract path at either end of its machining path. For 2½D objects, the path followed during machining is either parallel or perpendicular to the axis of rotation.

This generally leads to components composed only of analytic surfaces and hence in the more robust and reliable area of the domain of current solid modellers.

The exception to this is the machining of freeform surfaces, an area beyond the scope of this work. Any rotating tool can thus be represented by two profiles that when swept about the tool’s rotational axis give volumes
and surfaces of interest (see Fig. 13). The first profile represents the volume of the entire tool, and is used to check the access of the tool to the component. The second profile represents the cutting surface of the tool. To be consistent with Sungurtekin and Voelker’s nomenclature we will call the first profile the total tool profile and the second the operational tool profile.

Let us assume that we have some complete cutter path consisting of an approach path, a cutting path, and a departure path. These paths all consist of straight line segments or curve segments. In our restricted case the curves are all circular arcs. The tool does not change orientation during the motion. This is all consistent with a partial process plan for a 2½D feature to be machined on a three-axis NC milling machine.

There is a class of cutters such as t-slot cutters and dovetail cutters that have a limited access to the feature they are required to cut.

Boring bars present problems when trying to develop a basis for representing them with the tool profile model. They are classified as a type of traditional cutter for which the cutter profile can change dynamically during the cut. Fly cutters are another example of this class of tool.

The problem with the boring bar arises in the different access bodies that exist depending on the movement of the tool stem. The access body of a feature is the solid volume through which the tool passes during the manufacture of that feature. For a boring bar the access body for the insert and retract phase is equivalent to the tool profile swept around the vertical axis; however, their ‘access body’ during the machining phase requires the tool profile to be moved such that the tool tip is at the diameter of the required ring.

All manufacturing features must be machinable by at least one of the tools we define; however, we may be able to manufacture a feature in a number of different ways, provided that we ignore access problems. Only when we come to manufacture a feature can we identify which tool classes are actually able to perform the task; this is by constructing an access body based on the feature and the associated tool classes. Even this level of checking is based on the ‘perfect’ tool for the features parameters, regardless of the actual tools available. The tool access can also depend on the stage in the plan at which a feature is made. Once again the geometric reasoning here is preplanning and is merely to perform preliminary checks regarding our ability to manufacture the feature. As more detailed information becomes available about potential tools these checks can be performed again.

By using these tool access bodies, tool access problems on a component are found feature by feature for both the blank and the finished component. A comparison of tool access bodies would lead to one of the following situations:

1. tool has access to machine feature on both blank and finished component;
2. tool has access to machine feature on the finished component but not on the blank; and
3. tool has no access to machine feature on either the blank or finished component.

The first outcome means that the feature can be made at any stage in the plan. The second outcome means that once some other features have been machined, the tool can access the feature in question. This situation would trigger further analysis to define anteriority constraints on the plan. The third outcome would require a change of design to be made because the tool cannot access the feature at any stage in manufacture.

The following model is being developed to augment the existing PCR’s error models so that decisions can be based on tool and machine level error models, rather than on the less accurate general errors associated with process types. During the study of production rules at Mandelli SpA, engineers emphasized that machining errors represented in process capability models must be built so that the total cumulative errors resulting from both the machine error and tool error must be considered. For example, the linearity error of a tool combined with the position error of a machine would give a much more useful indication of process capability than considering the tool in isolation or using a general tolerance for that type of process.

As previously discussed, a tool’s or a machine’s shape-producing capability is represented in terms of the features they can make. When tolerances of these features are considered it is advantageous to generate the actual machining volumes so that a machining error volume can also be generated. Tool and machine models include a geometric description in the form of an ACIS solid body. This representation is being extended to hold tool motion information so that a volumetric error zone can be generated for a given specific tool–machine combination. This cumulative error volume would be used to test against the tolerance zone for that feature, to see whether the error zone ‘fits’ inside the tolerance zone.

Tolerance representation methodologies are being developed in parallel to enable such geometric tests to be conducted. The test uses a difference Boolean in the ACIS
solid modeller to find any volume belonging to the error volume that is outside the tolerance volume.

Fig. 14 shows an error zone generated for a drilling tool on a particular machining center.

Machining error analysis can be done analytically or by solids modeling, depending on the level of detail required for decision making. It would be too time-consuming to perform this test for all possible tool–machine combinations available to make each feature. For this reason meta-rules are needed to decide when the cumulative error volume needs to be generated for tool and machine selection. Typically a meta-rule would try to select a 'worse case', but still valid, combination to test. This would establish a cutoff for other tool–machine combinations in the set of feasible combinations.

4. Conclusions

Process capability modelling encompasses a wide range of activities. Rule-based systems cannot represent all types of process knowledge used in process planning, so better structured and enhanced models are a prerequisite for progress in the integration of CAD and CAPP. The development of process models and geometric reasoning algorithms should be driven by the demands of the applications that use them to support decision making. An industrial study used to evaluate the process capability models in a feature-oriented detailed design system integrated with a computer-automated process planning system led to the identification and development of models to enhance system performance. Models of shape creation capabilities, models for collision checking and more detailed error models have been discussed as particular areas needing further development.

Acknowledgements

The authors thank Mandelli SpA for taking part in the study that led to this publication, The Royal Society of Edinburgh for funding the six-month placement of Jane Naish in Italy, and the European Commission for funding the development work of FODDS through the SESAME project.

References


Biographies

Jane Naish received a B. Eng. from the Mechanical Engineering Department of the University of Edinburgh in 1991, and has since worked on feature-oriented design, and process capability representation (PCR). Recent research has concentrated on PCR and geometric reasoning for machining with a study performed at a machine tools manufacturer, Mandelli SpA, in Italy. She is currently a lecturer at the University of Abertay Dundee.
Frank Mill graduated in 1983 from Napier College with a B. Sc. in technology with industrial studies. After working at Coventry Polytechnic as both a research technician and research associate he took up a lecturing post in the Department of Mechanical Engineering at The University of Edinburgh in 1988, where he has worked ever since. His research interests include computer-aided process planning, feature-oriented design, concurrent engineering, and genetic algorithms. He is a Senior Lecturer and Director of the Manufacturing Planning Group.

Jonathan Salmon graduated from Imperial College, London, in 1988 with a B. Sc. in physics and then went on to the Artificial Intelligence Department of The University of Edinburgh where he gained an M. Sc. in information technology. After working in industry he joined the Manufacturing Planning Group in 1990 as a research associate to work on feature-oriented design for process planning. He is now a lecturer within the group. His research interests include geometric reasoning, solid modelling, and robotics.
APPENDIX D
FEATURES AS AUTONOMOUS AGENTS: AN ALTERNATIVE PARADIGM
FOR CONCURRENT ENGINEERING.

D. Jacquel, J.C. Salmon and F.G. Mill
Department of Mechanical Engineering
The University of Edinburgh, EH9 3JL, SCOTLAND

ABSTRACT
This paper describes a novel approach to the design of concurrent engineering systems by
reversing the traditional view of such a system as a number of distinct but integrated modules
operating on a data structure that is the product model. In this traditional view the data structure is a
passive entity, and must be acted upon by modules such as design, process planning and NC
generation.

In this new approach we imbue the model with intelligence (or at least a degree of autonomy),
i.e., have an active model surrounded by passive expert modules which are capable of answering
questions appropriate to their area of expertise. Indeed the model is allowed to be composed of any
number of active agents, each responsible for an appropriate portion or feature of the model. Therefore
a multi-agent system is created in which each design feature is an agent that tries to successfully
design, process plan and generate NC code for itself.

KEYWORDS
Feature Oriented Engineering, Agents, Multi-agent System, Active Model, Concurrent
Engineering

1. INTRODUCTION
This paper focuses on a new approach for the design of a concurrent engineering system. This
new approach takes its roots in the Artificial Intelligence field, and consists of applying the emerging
multi-agent paradigm to the design of a feature based design system.

In the traditional approach, a concurrent feature based manufacturing system is an integrated set
of software modules, such as modeller, geometric reasoner, process planner and NC code generator.
They are all acting upon a passive data structure that is the product model, continuously enriching it
so a design can evolve from the conceptual stage to a finished product complete with methods of
production. Using this classical approach, the designer creates a product through a Design/
Evaluation/ Redesign loop, adding new features to the model before using one of several modules to
evaluate the quality of the design against chosen criteria. This evaluation process can pinpoint flaws
in the design, which can then be corrected before enriching the model with new features.

No matter how complex and powerful the modules used on the model are, the classical design
process is not an interactive, concurrent activity but merely a client/server exchange between system
and designer. The system will give answers to specific requests but is incapable of taking initiatives
in the design process.

It is proposed to reverse this traditional approach, and create a system with an active product
model and passive modules. This is achieved by applying the increasingly popular multi-agent
paradigm to the system architecture. Indeed each instance of a design feature is an autonomous agent
in the system. The product model therefore becomes an active community of feature agents that
communicate and cooperate with one another in order to accomplish their goal: engineer themselves.
Thus, the design module looks like a traditional design system from the user’s point of view, but
unbeknownst to the user, the roles have been reversed. The user firmly believes that they are in
control and are designing some component for their own purpose. The product agent on the other
hand is 'using' the designer as an expert tool to fill in details about the design. In this way, user and product work in a symbiotic relationship to achieve a common aim.

An important factor is the interrelationship between the feature agents making up a complete product design. Certain features may intersect with each other or interact with each other in interesting ways that place considerable constraints on the manufacturing solutions for individual features. This problem is solved using two mechanisms. Firstly, all agents must be able to cooperate within the product model that they inhabit. The analogy here is one of a community and though each agent wishes for success as an individual, success of the entire community is paramount. Co-evolution techniques leading to emergent near-optimal solutions is an active research topic in Artificial Intelligence the results of which will be drawn on for applications in the concurrent engineering domain [1].

2. AGENTS

Formalising an agent definition is a difficult task, despite, or indeed because of, the vast amount of activity currently in the field. It is important distinguish the nature of agents at both the agent level and the community level.

2.1 Description of an Agent

Many recent publications related to agents and multi-agent systems (MAS) propose to give different definitions for agents [2][3][4]. At the entity level the most important features of a software agent are its co-operation and sociability aspects. An agent, can be thought of as an entity working inside a community and co-operating with other agents to achieve a goal. Rather than defining the nature of an agent, agency is indirectly described through properties required of agents.

- **Autonomy:** An agent can act with a certain range of autonomy and is capable of spontaneous actions. An agent has the capacity to plan its own actions and follow a self-prescribed schedule. Such autonomy can only be achieved through both synchronous and asynchronous actions.

- **Communication:** An agent must be able to carry bi-directional conversation with other agents in a language rich enough to allow it to express intentions and abilities to the community. Moreover, in order to comply with the autonomy property the communication protocol used by an agent should break the client/server protocol and permit peer-to-peer dialogues. [5]

- **Co-operation:** Using their communication abilities, agents should be able to initiate dialogues with one another in order to achieve their goals. This co-operative behaviour should lead to improved reactivity of the system and additionally, better interaction with users.

2.2 A Multi-Agent System (MAS)

Only considering the notion of agency at the agent level ignores the social dimension of agents. A social software agent is only useful if living inside a community of agents. In such a community, the asynchronous nature of each agent leads to exchanges of messages and decision/actions throughout the system. This asynchronous, peer-to-peer activity between individuals inside the system results in an overall multi-goal, co-operative search for an optimum solution.

Unlike a traditional procedural synchronous system, each feature agent initiates a dialogue only on an asynchronous basis, when it needs to solve a problem or improve its fitness against its goals. This allows the system to only use its processing power on critical parts of the design model, resulting in improved resource allocation.

2.3 Agents and Concurrent Engineering

Concurrent engineering (CE) proposes to make product creation a faster and more efficient process by allowing traditional sequential tasks such as geometry design and process planning to take place simultaneously. By parallelising the different tasks of the engineering process, CE brings early feedback to each task, avoiding costly and time-consuming redesign phases. The multi-agent paradigm and techniques have already been applied in a number of ways to CE [6].
Agents can be used as a mean of integration between the existing engineering tools by using autonomous interface agents between design and process and manufacturing applications [7]. In this approach agents provide their communication and information sharing capabilities to create a dialogue between design and manufacturing, bringing early feedback about the manufacturability of a product being designed.

Using more than interface agents [8] proposes integration of design, manufacturability analysis, incremental process planning, dynamic routing, and scheduling, using both feature agents and module agents (geometric interface, design agent, part agent and machine agents).

This work has arisen out of a previous project entitled *Simultaneous Engineering System for Applications in Mechanical Engineering* (SESAME), (BRITE/EURAM 0565). The goal in SESAME was a Simultaneous Engineering Workstation (SEW) and attempted to unify the tasks of Computer Aided Design, Process Planning and NC Generation in such a way that they could be used by a single engineer on a single seat in a standard environment. This goal SESAME achieved with a significant degree of success [9], but the tasks were still performed in a sequential manner as is evident from the architecture in Figure 1. The system was not truly concurrent, but achieved greater integration than earlier process planning systems.

![Figure 1: SESAME Architecture](image)

A subsequent proposal in order to increase the concurrency of the system was to build intelligent agents capable of incremental design, process planning and NC generation. This coarse grained approach (Figure 2) can be seen in a number of systems [10], and is a valuable approach when trying to integrate existing products into a concurrent agent based systems.

The Edinburgh Features as Agents' approach (Figure 3) uses a fine-grained model with lightweight agents acting for the features in the design and being aided by expert assistant modules such as incremental design, process planning and NC generation systems. This method is not suited to adapting existing commercial systems, but the knowledge gained in the group from SESAME and other projects is present to enable existing modules to be rewritten and adapted.

This fine-grained approach allows solutions to the current design problem to be developed all the time in the dead time between the human's design decisions. The fact that many of these solutions may be unsatisfactory in the finished design is unimportant as the system always holds as complete a picture as it can at any point, and after the last design decision is made, little work need be done to take the current plan and adapt it to the final design (in the majority of cases). In addition the designer is kept constantly aware of the downstream implications of their design.
3. THE EDINBURGH 'FEATURES AS AGENTS' PROPOSAL

The SESAME project was a major improvement over traditional systems, offering a highly integrated set of tools under a single interface allowing a user to cover the complete design process with rapid feedback about manufacturability and permitted the generation of NC code. However the design process, even though integrated, remained sequential rather than concurrent.

The Edinburgh proposal advances the benefits of SESAME by replacing the static product model with an active representation. This can be achieved with the multiagent paradigm applied at low 'fine-grained' level of design features. This radical switch underneath the interface is bringing new benefits to the system. It is also an efficient way of applying and solving design constraints as they can be made part of each agent's goal. It should also ensure quality models at all time and allow complex questions to be answered quickly as agents are constantly working on behalf of the user to find an optimum solution. This constant activity generates much redundant information, but this information is generated in otherwise idle CPU time, increasing the system's apparent performance.

3.1 Small example

To illustrate the proposal, a simple example of how the system works during the design of a small prismatic component is now presented.

Table 1 below shows the evolution of the system during the design process pictured in Figure 4 at three different levels. The user level reflects the designer interaction with the system. The model level represents the product state in terms of geometry. Finally the agent level focuses on the state of and communication between the agents populating the model.

This small example illustrates two important concepts not found in traditional CAD systems. It displays automatic resolution of a thin wall problem within the constraints defined by the user (*). It also illustrates the increased interactivity by allowing an agent to initiate dialogue with the user in the case of an unresolved problem (**).
Table 1 Agent Communication during Design Example

<table>
<thead>
<tr>
<th>User level</th>
<th>Model Level</th>
<th>Agent level (\footnote{\textcircled{\textbullet} shows inter-agents communication)}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creation of blank block with tight size and position constraints</td>
<td>Untouched blank</td>
<td>block: I'm the only agent in the model ... sleep</td>
</tr>
<tr>
<td>Add new pocket with normal position constraints</td>
<td></td>
<td>pocket: I'm not alone in the model.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pocket\footnote{\textcircled{\textbullet}}Broadcast: New pocket at position xyz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>block: thin wall problem, but I can't move/resize</td>
</tr>
<tr>
<td></td>
<td></td>
<td>block\footnote{\textcircled{\textbullet}}pocket: you're creating a thin wall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pocket: I'm moving to new position xyz (*)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pocket\footnote{\textcircled{\textbullet}}Broadcast: pocket at new position xyz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pocket: no more problem ... sleep</td>
</tr>
<tr>
<td>Add new hole with tight constraints on position</td>
<td></td>
<td>block: no more problem ... sleep</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hole: I'm not alone in the model.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hole\footnote{\textcircled{\textbullet}}Broadcast: New hole at position xyz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>block: I can't guarantee access, I can't move/resize</td>
</tr>
<tr>
<td></td>
<td></td>
<td>block\footnote{\textcircled{\textbullet}}hole: access problem block position xyz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pocket: the new hole interacts with me.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pocket\footnote{\textcircled{\textbullet}}hole: there is a pocket at position xyz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hole: try to solve access problem, but I can't move/resize</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hole\footnote{\textcircled{\textbullet}}pocket: Can you move to new position xyz?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pocket\footnote{\textcircled{\textbullet}}hole: I can't comply with request.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hole: I can't solve this problem.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hole\footnote{\textcircled{\textbullet}}user, I have an access problem. (***)</td>
</tr>
<tr>
<td>Receive message from hole. Modify hole position</td>
<td></td>
<td>hole\footnote{\textcircled{\textbullet}}Broadcast: hole new position xyz</td>
</tr>
<tr>
<td></td>
<td></td>
<td>block: no more problem ... sleep</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pocket: no more problem ... sleep</td>
</tr>
</tbody>
</table>

3.2 Current state of work

Using the Feature Based Design knowledge gained in SESAME, a new design tool is being implemented as a multi-agent system. It takes a fairly simple subset of the problems to be solved during the conception process in order to test the validity of our approach.

To an external user, the new implementation appears little different from the previous Feature Based CAD system. It allows the user to add negative (material removal) features to a blank to obtain the desired design. Underneath the user interface, however, radical changes have taken place. Where there used to be a passive product model on which software modules were acting, there is now a community of agents tirelessly working together to reach an optimum solution.

Each time the user adds a new feature to his design, they actually gives birth to a new software agent that joins the existing community inside the system. As soon as this new feature/agent enters the MAS, it can start following its own plans and interact with other agents in order to satisfy its goals. It is this underlying community of agents that represent the living product model.

The current MAS is limited to geometric reasoning such as thin walls detection but it already shows its potential by immediately detecting problems during design. For example, if the user places a new hole too close from another one, creating a thin wall, the system will immediately detect it and take action to solve the problem. This could lead to one of the agents deciding simply to move slightly aside or reduce its diameter (subject to previously supplied constraints), or ask the user for an alternate solution. The system also delivers better performance in terms of response time when adding new features to an existing design because it is incremental by nature. It doesn't need to re-analyse the entire design against the new feature, the new agent interacts locally with existing agents only. Of course a snowball effect is always possible when adding a new feature to the design but this only happens if agents are insufficiently constrained, or if the new feature interacts with many others.

3.3 Future work

The first addition to the current implementation will be to add non-feature agents for conflict resolution and solution proposal. These new agents should share the same structure as the feature...
agent but provide more complex computation power to propose solutions to problems our lightweight feature agents are unable to solve. A second addition will be expert modules to handle support for process-planning and NC code generation to our current agents. This should not be a major difficulty thanks to the modularity of the Edinburgh approach. Adding support for these tasks also requires adding new knowledge/abilities to our existing agents.

It is already possible to express constraints local to a single agent, the next step is to work on constraints propagation inside the agent community forming the product model. It is also intended to investigate the feasibility of applying the multi-agent paradigm to a constraint solving feature based design system that would automatically generate design solutions. Lastly, we will investigate solutions to the problem of combinatorial explosion that we can see looming in the system.

4. CONCLUSION

This new approach, consisting of turning the system upside-down, making each feature an active element of the system has, so far, shown potential. The constant activity of the agent community radically changes the way the system behaves. The active model, always in search for an optimum solution, ensures a quality model at all stage of design. Leading to several major improvement from the traditional design system architecture.

One could argue that constantly analysing the model generates tremendous load on the computing resources but the MAS provides a better use of computing resources compare to the traditional approach. Indeed, it is believed that, despite the large increase in term of number of operations, the active model provide better overall performances by diluting the calculations along the entire design process.

The Design-Evaluation-Redesign loop no longer needs to be performed by the user. This process is achieved individually by each agent in the system during the design phase. Being able to detect potential problems at early stages through this constant self-assessment of the model, the multi-agent approach reduces costly redesign. The active model is also an efficient way of applying constraints to the design since the constraints can be embedded inside each agent. Finally a higher interactivity during the design process is achieved by enabling any agent to initiate a dialogue with the user.

5. REFERENCES


A voxel-based representation for evolutionary shape optimization

PETER BARON,1 ROBERT FISHER,1 ANDREW TUSON,2 FRANK MILL,3
AND ANDREW SHERLOCK3

1Department of Artificial Intelligence, University of Edinburgh, 5 Forrest Hill, Edinburgh EH1 2QL, Scotland, UK.
2Department of Computing, City University, Northampton Square, London, EC1V 0HB, UK.
3Department of Mechanical Engineering, University of Edinburgh, Sanderson Building, The King's Buildings, Mayfield Road, Edinburgh EH9 3JL, Scotland, UK.

(RECEIVED June 6, 1998; REVISED November 27, 1998; ACCEPTED December 7, 1998)

Abstract

A voxel-based shape representation when integrated with an evolutionary algorithm offers a number of potential advantages for shape optimization. Topology need not be predefined, geometric constraints are easily imposed and, with adequate resolution, any shape can be approximated to arbitrary accuracy. However, lack of boundary smoothness, length of chromosome, and inclusion of small holes in the final shape have been stated as problems with this representation. This paper describes two experiments performed in an attempt to address some of these problems. First, a design problem with only a small computational cost of evaluating candidate shapes was used as a testbed for designing genetic operators for this shape representation. Second, these operators were refined for a design problem using a more costly finite element evaluation. It was concluded that the voxel representation can, with careful design of genetic operators, be useful in shape optimization.

Keywords: Shape Optimization; Evolutionary Algorithms; Voxel Representation

1. INTRODUCTION

Shape optimization attempts to find an optimal shape for a component subject to design constraints. Typical problems that are of interest to the research community in this area have been concerned with structural load-bearing components and aerodynamic profiles. Some work has also been reported in areas such as thermal conduction for heat sinks and manufacturing cost minimization. In structural shape optimization, often these studies aim to minimize the amount of material (and hence perhaps cost and weight) needed to support a given load. In aerodynamic optimization, often the aim is to minimize drag subject to constraints on lift and geometry. Almost all of the work to date has described shape representations for single-criterion optimization, although many researchers are interested in multicriteria problems.

Structural shape optimization can be usefully characterized as the integration of geometric modelling, structural analysis, and optimization algorithms (Hsu, 1994). The finite element (FE) method is popularly used to analyze candidate shapes. In early research in shape optimization, the FE mesh itself was used as the geometric model to be manipulated by the optimizer. Optimization techniques then available were based on mathematical methods of function optimization, typically gradient based. The nodal coordinates of the FE mesh were used as design variables. However, it soon became apparent that use of the mesh as the geometric model was impractical due to difficulties in ensuring that the mesh could adequately calculate stresses and in keeping the shape's boundary smooth. Researchers moved to separating the geometric modeler and the FE mesh. Commonly the boundary of the component is modelled using splines, with control point coordinates used as design variables. Splines have the useful property of smoothness and local shape control. Mesh generation techniques then generate an adequate mesh given a description of the candidate shape's boundary.

Gradient-based optimizers can find optima with few design evaluations. This is often extremely important in engineering problems, where the time taken to perform one
design evaluation is often many orders of magnitude greater than the time taken to produce candidate designs. However, such optimizers can often have difficulties in dealing with local optima, discrete design variables, and with noise generated when small changes in the design variables cause changes in mesh topology. Recently, to address these problems, the use of stochastic optimization techniques, such as genetic algorithms (GAs) due to Holland (1975), and simulated annealing (Kirkpatrick et al., 1983), in shape optimization (Chapman et al., 1994; Smith, 1995b) has been a popular area of research. Generally, this research has still retained a parameterized description of the shape's boundary as the geometric model.

The work described in this paper investigated the possibility of replacing this boundary representation of the shape with a cellular representation. The cellular representation chosen in this work used voxels that partition the design space into rectangular regions or boxes that are then assigned a binary full or empty value. This approach was motivated by a number of potential advantages (Smith, 1995b):

- any shape can be represented to an arbitrary accuracy by increasing resolution;
- it is straightforward to convert existing engineering solutions into voxels;
- they map naturally to the representations frequently used by GAs;
- domain knowledge can be readily incorporated;
- geometric constraints can easily be applied; and,
- the topology of candidate shapes is not predefined.

However, in contrast to the successful application of this technique in (Farrell, 1998) for the inversion of geographical and potential-field data, earlier work by Watabe and Okino (1993) states the following objections to the scalability of voxel representations:

- the occurrence of small holes in the final shape;
- the long length of the chromosomes;
- the expectation that crossover operators would be ineffective; and
- the lack of smoothness in the shapes' outlines.

Given the potential advantages of a voxel representation, the Authors considered it worthwhile to address these difficulties. Specifically, the aims of this work were:

- to determine the suitability of voxels as a geometric model for use in shape optimization and any difficulties, such as those outlined above, that may arise;
- to design suitable operators for a GA optimizer to use with such a representation to overcome such difficulties; and
- to investigate and identify issues that will have to be confronted by the practitioner in scaling up this representation to real-world problems.

Therefore this work does not aim to produce a system that returns a usable, improved solution to a real-world problem. Instead, it concerns itself with the more strategic and scientific question of investigating and, where possible, resolving issues that pertain to how a practitioner is to construct such a practical system.

1.1. Experiments

Two experiments were devised to investigate the voxel representation. First, a simplified beam design problem was formulated for which the cost of evaluation would be small. Using this problem as a testbed, a number of operators were designed. Second, an annulus design problem was tackled using a finite element analysis. Thus, the computation cost of evaluation in this case was much greater. The usefulness of the operators designed in the first experiment could then be evaluated with a more difficult design problem and related scalability issues investigated. Baron (1997) gives comprehensive details of all experiments undertaken.

Finally, this investigation will restrict itself to examples where 2D voxels (pixels) are used. This is for reasons of convenience and speed of solution evaluation as FE analyses in three dimensions are more computationally demanding. However, no assumptions are made in this study regarding the dimensionality of the problem and so the results presented here should be generalizable to higher dimensional problems.

2. SIMPLIFIED BEAM DESIGN

A prototypical mechanical engineering problem is that of optimizing a beam to support various loads with a minimal amount of material. Evaluation of the candidate cross sections was made using bending theory for symmetrical beams, considering only normal stresses (Gere & Timoshenko, 1984). This is an oversimplified model, but is sufficient to test whether the potential problems with a voxel representation outlined above do pose a problem in practice. The maximum stress constraint imposed by the physics model used in these experiments is summarized below.

\[
\left| \frac{My}{I} \right| < \sigma_{\text{max}} \text{ for all voxels,}
\]

where \( \sigma_{\text{max}} \) is the maximum stress allowed within any given area (voxel); \( M \) is the bending moment; \( y \), is the distance of the voxel \( i \) from the neutral axis of the shape; and \( I \) is the second moment of area of the candidate cross section. The neutral axis of a shape is defined as a horizontal line that passes through the center of the mass of the shape. As a voxel representation uses areas that are all of uniform size and density, the center of mass can be found by taking the average of the positions of all occupied voxels. The second
moment of area is approximated in the discrete representation by summing the moments of each voxel, that is:

\[ I = \sum_{i=0}^{n} a y_i^2, \]

where \( a \) is the area of a voxel.

In the real world, the solution to this problem would correspond to an I-beam, but that also requires a web to connect the two flanges of the beam together. In a design based on a full calculation with shear stress, the web would be necessary so to counteract this additional stress. However, as shear stress is not represented in this problem, a connectivity requirement in the form of a repair step was added, whereby all pixels must be connected to a seed pixel in the center top edge of the beam. In addition, all vertically central voxels were enabled to provide a straight web before the connectivity repair step. This was found, in formative experiments, to prevent the formation of a crooked web (as the physics model used does not prevent this), and improve slightly the results obtained.

To try to ensure that the alterations and improvements made to the GA here will also prove beneficial to the real-world problem, it was decided not to concentrate on fine-tuning any of the various parameters available, but rather to focus on the design and operation of various new operators. Therefore, parametric variations were restricted to an absolute minimum and were used only to determine the approximate values required to gain reasonable advantages from the new operators. Therefore, in the following experiments, the following parameter settings remain constant unless mentioned otherwise:

- Beam dimensions = 0.05 \( \times \) 0.10 m
- Bending moment = 13,000 Nm
- Voxel grid = 32 \( \times \) 64 voxels
- Max. stress allowed = 2 \( \times \) 10\(^8\) Nm\(^{-2}\)

### 2.1. Experiments using the naïve GA

The first set of experiments with a 2D representation treated the chromosome as a long 1D binary string that wrapped around at the vertical edges onto new lines to form the 2D cross section. Standard two-point crossover \( (p_c = 0.35) \) and bitwise mutation \( (p_m = 0.001) \) were used in conjunction with a generational GA with a population of size 20. GENITOR-style rank-based selection (Whitley, 1989) was used throughout. From the above, the fitness function, \( F \), to be minimized was of the following form:

\[ F = V + S/(1000 \times \sigma_{\text{max}}) + k \times \max\{(S - \sigma_{\text{max}}),0\}, \]

where \( V \) was the count of active voxels (proportional to weight), \( S \) the maximum stress of any voxel, \( \sigma_{\text{max}} \) the value of the maximum stress constraint, and \( k \) the constraint penalty multiplier (set to \( 5 \times 10^{-3} \) according to the results of formative experiments).

With this particular optimization problem, the difficulty lay not in getting a valid solution, but in getting a near optimal-mass solution. The first experiments were relatively unsuccessful in this regard: the results after 2000 generations were full of small holes and had extremely uneven inner edges. This can be seen in the typical end-of-run results shown in Figure 1 (the numbers represent the fitness values of each individual).

The stresses were concentrated at the vertical extremes of the beam, so the material in the middle contributes less toward the beam’s ability to withstand the load, and therefore as we are trying to minimize the mass of the beam, the material is more usefully used at the extremes of the beam.

The GA, even in this simple standard form, rapidly removed material from the middle of the cross section, and in the later stages of the experiments was observed to be moving material from low-stress areas into high-stress areas where holes were left near the extremities.

However, this first naïve GA approach took an extremely large number of evaluations to make significant progress,
and this is not acceptable as later experiments would have a greatly increased evaluation time due to the integration of the FE package. The rate of improvement was also seen to decrease as the run continued, levelling off to almost none at all by the end of the run. This means that the GA was not finding any further improvements to the chromosome and, as the results are visibly poor, it indicates a general weakness in the operators being applied.

Attention was therefore concentrated towards improving the GA operators to achieve greater benefits during the early search period, and to produce better quality final results.

2.2. The smoothing mutation operator

The smoothing operator experiments were an attempt to address directly some of the weaknesses of the voxel representation by devising a new specialized operator, which should aid the search by reducing the number of small holes and ragged edges produced by the GA. The new operator was intended to be capable of easy expansion from two dimensions to n-dimensions that it would continue to be useful in the case of higher dimensional problems using the voxel representation.

This operator selects a rectangle with random position and size ranging from 2 pixels to one quarter of the dimensions of the grid. The most common value for the pixels in the area selected was then found and written to all of the pixels in that area (Fig. 2).

The GA parameters used were the same as before and the new operator was applied in addition to the previous mutation and crossover operators—application of this operator to 60% of the chromosomes in the population was found, in formative experiments, to give the best results. The GA configuration was otherwise unchanged, though the number of generations was limited to 1500 in this case.

Comparing Figure 3, which displays some typical end-of-run population members with earlier results (shown in Fig. 1), shows just how effective this domain specific approach to operator design has been, especially at eliminating isolated holes and reducing ragged edges.

2.3. UNBLOX: An N-dimensional crossover operator

The two-point crossover operator, which had been used up to this point, treated the chromosome as a 1D string of bits and therefore suffered from a problem with linkage—voxels that are adjacent in a 2D grid are not necessarily adjacent in the 1D string. This separation increases the possibility that useful building blocks (areas of the grid which contribute to a higher overall fitness evaluation) will be disrupted during the crossover procedure.

Cartwright and Harris (1993) describe the use of the UNBLOX crossover operator, which was specifically designed to overcome these limitations with conventional two-point crossover. This operator swaps a rectangular area of the grid instead of the substring swapped by two-point crossover. If the area overlaps an edge of the grid then it is made to “wrap-around” to the opposite side—this convention was adopted from the original paper, though its effect on edge smoothing is somewhat unclear. The size and location of the area

![Fig. 2. The smoothing operator.](image)

![Fig. 3. Typical end-of-run results with the smoothing operator.](image)
to be swapped are selected at random, and in this implementation the area was restricted to a minimum size of two voxels per dimension so that the operator would always have some effect when applied.

The crossover operators were used with the standard probability of 0.3 per chromosome and no changes were made to the standard algorithm or to any of the other parameter settings described earlier. The graph in Figure 4 shows the results of 10 trials using three alternative crossover operators, including the UNBLOX operator. The other two crossover operators were the standard two-point crossover and uniform crossovers (Goldberg, 1989).

The results confirm that the UNBLOX operator does indeed perform better than either the two-point crossover or the uniform crossover techniques on this problem. The rate of descent of the UNBLOX line is quicker, indicating that the population converged to good solutions faster with this approach than with the other operators, and the eventual end result after 1500 generations had a slightly better fitness value than those produced by the other techniques.

2.4. Two-dimensional mutation operators

A new mutation operator was designed which scrambles the contents of a randomly selected rectangular area of the voxel grid, it is referred to here as the "two-dimensional" operator. This operator can be easily modified to work in N-dimensions, and affects a relatively small area of the chromosome rather intensively in the selected rectangular selected area in the same way as for the smoothing mutation operator. A second, somewhat altered version of this mutation operator was also designed and tested in these experiments called the "two-by-two" area mutation operator. This operator uses a fixed mutation square of two-by-two voxels and was designed to be applied only if at least one voxel in the mutation area is already active. The theory behind this operator is that most of the modifications need to be made to the surface or interior of the evolving shape and that little benefit will result from flipping isolated voxels in the middle of the void areas. The choice of a fixed two-by-two area was motivated by the observation that most of the irregularities on the surfaces would fit into such an area and that with only 16 permutations possible (4 binary bits), the probability of mutating a poor-quality area into a more fit variation would be reasonably high.

The new operators were again applied in addition to the original bitwise mutation operator, with a probability of 0.25 per chromosome of being applied. After each application there was a decreased probability of the same operator being applied again, with the probability of a further application being decreased to one half of its previous value each time. The experiments were performed 10 times for each of the 3 alternative mutation combinations, over a period of 1500 generations.

The graph in Figure 5 shows the effect of the two new mutation operators alongside the results obtained when neither of them was applied. The generation number is plotted along the horizontal axis and the average fitness of the best individual from the population at each generation is plotted vertically.
The addition of the 2D operator generally results in better performance than the bitwise operator alone, though the two lines do meet between generations 300 to 400. The steeper descent of the 2D operator line indicates that early performance was especially improved, and the final result after 1500 generations is significantly better than previously. The two-by-two operator offers a similar rate of improvement during the early stages of the trial, a slightly better performance between generations 100 to 600 and finally converges with the 2D operator's line at about generation 1000. This seems to indicate that although offering early benefits to the optimization, it is not better than the 2D operator in the long run.

In conclusion, two new mutation operators were designed with the particular intention of directly addressing the perceived problems with the prior optimizations. Both of the new operators were found to be more effective than the previous uninformed bitwise mutation, producing benefits to the rate of early improvement and the final quality of solution generated.

In the absence of any other clearly distinguishing features, the two-by-two operator will be used during the further experiments, as it offers a speed advantage over the 2D mutation operator outlined above.

2.5. Conclusions about the beam design problem
The results have shown that although a naïve GA does indeed suffer from the problems suggested by Watabe and Okino (1993), a small selection of operators informed only by domain knowledge about the representation will effectively solve each of these difficulties.

To see whether the above improvements can be usefully combined to produce the desired behavior, and improve further upon Figures 1 and (especially) 3, Figure 6 depicts a number of typical end-of-run results for the complete system with all operators active. Comparison with the earlier results shows that the complete system produces superior results with no holes or large protuberances. In addition, the dramatically improved performance of the final system in terms of the solution quality-time tradeoff surface it exhibits is shown clearly by Figure 7.

In summary, the final system uses a normal bitwise mutation operator in addition to the two new mutation operators, smoothing, and two-by-two. The smoothing operator rapidly cuts away unwanted areas of material during the early stages of the optimization and can help to smooth ragged edges and fill small holes later on. The two-by-two mutation operator is highly effective at smoothing off ragged edges and at filling in small holes in the material if they occur in undesirable places. Finally, the two-point crossover operator has been replaced by the n-dimensional UNBLOX operator, to fully exploit the 2D structure of the problem.

3. ANNULUS DESIGN PROBLEM USING FE ANALYSIS
The experiments undertaken with the simplified beam design problem outlined in Section 2 led to the design of ef-
effective GA operators for manipulation of 2D shapes. This section details further experiments undertaken to apply these operators to a more difficult design problem. The problem chosen was to design a jet-engine annulus. The finite element method was chosen as the design evaluation/analysis technique. Initially, for ease of implementation, the voxel shape description was directly used as the finite element mesh.

3.1. The annulus design problem

The full original specification of this problem was taken from Smith (1995a). The problem is to design a jet-engine annulus that is subjected to loading due to rotation and due to the attachment of the turbine blades to its outer circumference. The part is axisymmetric around the axis of rotation, and consequently it reduces to the 2D shape optimization problem shown as Figure 8.

The optimization involved reducing the mass of the annulus while observing a series of four separate stress constraints at discrete locations in the annulus. The constraints relate to the hoop stresses at the inner and outer circumferences and the radial stresses along the center line of the annulus. The stress constraints to be observed were, in descending order of importance:
3.2. The fitness function

The GA fitness function was defined as an objective (the weight of the annulus in kg, and a factor to minimize the total stress, in MPa) plus a sum of penalty terms if one of the 4 stress constraints was broken. The function maximized

\[ F = \frac{\Sigma_i \sigma_{max(i)}}{\Sigma_i 1000 \times S_i} - \text{annulus_weight} - \Sigma_k k \times i \times \max\{S_i - \sigma_{max(i)}, 0\} \]

Constraint penalties were applied if any of the four constraints limits \( \sigma_{max(i)} \) were exceeded by the stress, \( S_i \), measured (in MPa). The constraints were ordered in importance by using \( 4 \times k \) for the most important, \( 3 \times k \) for the second most important, \( 2 \times k \) for the next and \( 1 \times k \) for the least important constraint, the (decreasing) order of importance was as for the constraints limits listed above.

3.3. Results from the basic system

Again, a generational GA with a population of size 20 and GENITOR-style rank-based selection was used. The UNBLOX, smoothing mutation, and 2-by-2 mutation operators were applied sequentially with probabilities 0.3, 0.8, and 0.8 respectively (on the basis of formative experiments). A 62-by-27 voxel grid was used to represent the annulus and the constraint penalty, \( k \), was set to 0.00005. The settings used for the annulus were:

- Dimensions of design space = 0.25 \( \times \) 0.05 m
- Radius of hole = 0.10 m
- Blade force = \( 10 \times 10^5 \) N rad\(^{-1}\)
- Young's modulus = \( 2.238 \times 10^{11} \) N m\(^{-2}\)
- Material density = \( 8.221 \times 10^3 \) kg m\(^{-3}\)
- Revolution speed = 1571.0 rad s\(^{-1}\)

The basic system was first applied without further modifications to the annulus optimization. However, the problem, as specified, was very tightly constrained, which meant that the attempts to solve this problem using random population initialization violated all of the stress constraints by large amounts. Also, the rate of improvement in the population, when extrapolated beyond the time period allocated to the experiments, indicated that a valid solution would not be found for some considerable number of generations.

To circumvent this problem, the population was instead initialized with a selection of variations on the annulus design supplied with the original specification, which were modified further by an aggressive random mutation operator that added and removed small areas of material over the surface of the annulus design. This kind of intelligent initialization was thought reasonable as a user will often want to start the GA with existing designs to see what improvements can be made. Even when a totally new shape is being designed, the user would normally have some expectation about the final form, which could easily be used to initialize the population. The intelligent initialization approach meant that the initial population was not unreasonably far outside of the stress constraints, yet supplied the optimization with sufficient variation that the population did not rapidly converge onto a single solution. Some of the results from this basic system can be seen in Figure 9, which shows six members of the population after 75 generations.

The results shown in Figure 9 were poor. The lack of symmetry around the horizontal axis and the uneven edges were just the most visible failings in this set of results. A second problem was the occurrence of large stresses at the corners of elements on the edge of the shape. These failing need to be addressed if any claims as to this representation's scalability can be made.

3.4. Improvements made to the system

Attention was now turned to resolving the issues and shortcomings highlighted by the above investigation in turn.

3.4.1. Use of symmetry

It was known that a solution to the annulus design problem should be symmetric about a radial axis. It was there-
fore decided to utilize this domain knowledge and thus reduce the search space of the problem. The GA was modified to reconstruct the final shape in its entirety only when producing the element definition files to be accessed by the FE package. This simple modification reduces the search space from a typical size of $2^{2542}$ for a 62-by-41 voxel grid, to $2^{1302}$, which represents a 62-by-21 voxel half-grid. The central line of voxels along the axis of symmetry is not mirrored as it is now enforced by the GA to be always turned on—this also provides a guaranteed central line of elements for the stress measurements to be taken from.

### 3.4.2. Mesh improvement

It was found in the initial experiments for the annulus design problem that directly using the voxel description of the geometry as the FE mesh caused problems with high stresses caused by corners in the mesh. It was therefore decided to separate the geometry model and mesh. There were several possible approaches that could have been taken. An approach that was considered was to use interpolation splines to form a smoothed edge. The voxels would then act as a "skeleton" and the spline as a "skin." A mesh generator could then produce a mesh whose density could then be independent of the voxel model. However, for this prototype system it was decided simply to add triangular elements at the corners. While this was a far less elegant solution, it was much simpler to implement. 

These new triangular elements were created by specifying connections between groups of three nodes in the element connection file. These triangular elements were added to the shape at all suitable "steps," which were identified by convolving the voxels in the shape against a series of four matching template masks. If each square in the mask matched the value of the voxels surrounding an empty voxel then the appropriate triangular element was created in the "step." The convolution masks and the triangles that they caused to be inserted are shown in Figure 10.

### 3.4.3. Design of operator to remove holes

The two-by-two mutation operator (which can either fix holes or cause them to appear) was modified to only mutate areas where, as well as at least one voxel being turned on, at least one of the four voxels is also turned off. The result of this modification is that the two-by-two mutation operator can now only mutate at the boundaries of the shapes being formed, and consequently it should also help reduce the number of small protuberances.

### 3.5. Results of improved system

The improved GA for annulus optimization used the same settings as the basic system for all parameters except that the chromosomal grid was set to 21 voxels high, which is mirrored due to the symmetry used to produce a voxel grid height of 41 voxels. The analysis was permitted to continue for 114 generations and this took approximately 24 hr in total. Some of the final population created by the improved GA are shown in Figure 11. This displays 3 of the 20 individuals and shows a clear improvement in quality over the results generated previously. The small protuberances have been totally eliminated and only a few members of the population contain small holes. The rate at which a valid solution was found is considerably faster than the basic implementation, and once found, the GA continued to improve upon this solution even to the very last pass of this trial.

The annulus shapes produced can be seen to be unusual. It is proposed that the "overhangs" present at the cob and the thinness of the neck are due to the inadequate specification used for the annulus and the method used to penalize

![Fig. 10. Convolution masks for triangle insertion process.](image)

![Fig. 11. Final annulus cross sections from improved GA.](image)
constraint violation. Stress constraints were defined for four discrete points in the specification that was intended to be used with a parameterized shape description. This specification would be adequate for such a representation. However, with the voxel representation the optimizer was able to remove material with greater flexibility. At an optimal solution one of the stress constraints is just inactive. Removing more material would then increase the stress to above the maximum value. However, the GA could improve the fitness value if, by adding material elsewhere, the position of high stress was moved from the point at which the constraint was assessed, as long as the amount of material added was less than that removed. Given that this explanation is correct, the problems do not lie with the voxel representation and could be solved by improving the specification and method of penalizing constraint violation.

After using the FE package to examine the solutions produced by this optimization, it was possible to confirm that the use of the triangular elements to smooth the boundary worked as expected in reducing the amount of stress in the regions immediately surrounding a step. Figure 12 shows the stress values calculated by the FE package for the voxels surrounding steps in two typical runs and clearly shows how the triangles permit the excess stress to be distributed in a more even pattern. Darker shades indicate higher stress levels in both of these pictures.

3.6. Conclusions for the annulus design problem

It was found that the use of unmodified operators from the beam design problem was unsuccessful. However, when the operators were modified, taking into account knowledge held about the annulus design problem, the results were more successful (Fig. 13).

Difficulties were encountered in the direct use of the voxel shape representation as the FE mesh. These were, to some extent, alleviated by the use of smoothing triangular elements. However, the full decoupling of the primary voxel-based shape description and FE mesh would be desirable in future studies.

Unfortunately, due to the flexibility of the voxel representation in removing and adding material coupled with the GA's ability to exploit the whole search space, it was found that the specification of the problem needed to be more tightly defined, as unwanted overhangs were present in the final solution. In response it should be noted that, in the authors' experience, there are often a number of possible problem formulations for a parametric approach, each with differing suitability to the problem at hand and ability to represent only feasible solutions. Therefore, the above should not be taken to be a severe criticism of the voxel representation—for any approach, a significant amount of experimentation will be required to identify a suitably constrained problem formulation.

The unwanted overhangs aside, a comparison of the mass of the annulus produced by the voxel representation (41 kg), compares well against the original annulus design (68.6 kg), and that produced by the parametric GA described in (Smith, 1995a) which achieved an annulus of mass 40.9 kg. All of these annulus designs satisfied the stress constraints, though given that these designs were evaluated using different FE packages, a fine-grained comparison needs to treated with some caution.

Finally, and rather unfortunately, the voxel GA did not perform as well in regard to time to solution. The parametric GA found its solution in 400 evaluations compared to the 1000 evaluations required by the voxel-based GA—this was felt to be a result of the GA having to search a much larger and less constrained search space when using a voxel representation.

4. CONCLUSION

Voxels were found to be a viable representation for shape optimization with an evolutionary algorithm in 2D problems. They have a number of potential advantages over other
A voxel-based representation for evolutionary shape optimization

representations such as parameterized boundary descriptions. Topology is not predefined, domain knowledge is easy to incorporate, geometric constraints can be easily applied, and it is straightforward to convert existing solutions into such a description to "seed" an initial population of shapes.

Experiments were undertaken on two design problems to investigate the effectiveness and scalability of this representation: a simplified beam design and a jet-engine annulus design using finite element analysis. During these experiments, a number of difficulties inherent with this representation were addressed, primarily by use of specifically designed genetic algorithm operators that utilized domain knowledge held about the problems tackled. An N-dimensional crossover operator was used that provided linkage between adjacent rows of voxels and thus avoided the slow convergence found with a conventional crossover operator. An operator was designed to remove unwanted holes produced in candidate shapes and to smooth boundary edges.

On the annulus design problem, the direct use of the voxels as the finite element mesh was found to be inadequate, and a convolution-mask-based solution to this issue was devised. That said, further work in this regard will involve the further decoupling of the voxel representation and mesh.

Furthermore, the flexibility of the voxel representation, along with the GA's exploitation of a much expanded search space uncovered deficiencies in the specification used for the annulus design problem, leading to unwanted "overhangs" in the solutions obtained. Although the results obtained were roughly equivalent in terms of the mass of annulus produced, they compared poorly with regard to the number of evaluations required to find such a solution.

Finally, it should be noted that GA optimizers can easily be modified to be used as interactive optimization systems (Tuson et al., 1997). In this case, the computer would rely on an engineer's practical experience and knowledge of the problem domain to direct key choices in the optimization process. Given the diversity of possible shape optimization problems, such flexibility will be required to deal with the constraint handling issue noted above. The lack of initial assumptions in the voxel representation could be seen to be an advantage here as the engineer has, in effect, a tabula rasa to start work from, and constraints on the solutions obtained can be expressed directly. Given the amount of experimentation required to find a good problem formulation for both parametric and voxel approaches, such an interactive approach would be highly desirable in any case. Further research into principled methods for allowing the user to interact with such a system is therefore recommended.

ACKNOWLEDGMENTS

The authors thank the Engineering and Physical Sciences Research Council (EPSRC) for partly funding this work via two research studentships (95303677 and 95306458).

REFERENCES


Peter Baron is a professional video games programmer who received his M.Sc. in Information Technology: Knowledge-Based Systems at Edinburgh University through part-time study, completing his thesis in 1997. His thesis "Evolutionary Shape Optimization using a Voxel Based Representation" forms the basis for this paper. He has spent a year working on a "Virtual Reality" advertising system for In-
puting at City University, London, U.K., where his research interests include methodologies for the principled design of neighborhood search and evolutionary optimizers, and their application to domains such as operational research and engineering.

Robert B. Fisher received a B.S. with honors (Mathematics) from California Institute of Technology (1974) and a M.S. (Computer Science) from Stanford University (1978). He received his Ph.D. from University of Edinburgh (awarded 1987), investigating computer vision in the Department of Artificial Intelligence. Dr. Fisher is a Reader in the Division of Informatics at the University of Edinburgh. His research covers topics in high-level computer vision, and he directs a research project investigating three-dimensional model-based vision and automatic model acquisition. He teaches general and industrial vision courses for undergraduate, M.Sc.- and Ph.D.-level students.

Frank Mill is a senior lecturer in manufacturing engineering at the University of Edinburgh. He has published over 30 papers on optimization in design, process planning, and manufacturing systems, for example scheduling, in the last 12 years. His primary interests are in the use of evolutionary algorithms to such problems.

Andrew Sherlock is presently a Ph.D. student in the Manufacturing Planning Group in the School of Mechanical Engineering at Edinburgh University. His interests lie in the application of biologically inspired processes to Mechanical design. Specifically, he is using morphogenetic approaches to shape generation, coupled with a genetic algorithm, for structural shape optimization.
APPENDIX F
Biological analogies in manufacturing

F. Mill*, A. Sherlock

School of Mechanical Engineering, The University of Edinburgh, King’s Buildings, Mayfield Road, Edinburgh, EH9 3JL, UK

Abstract

Biological organisms and manufacturing facilities are both examples of complex systems that exist in changing environments. It may be that some of the lessons that have been learned from nature are applicable to engineering. The purpose of this paper is to examine analogies from nature and to discuss their relevance to engineering systems. In particular systems that exhibit emergent behaviours or are subject to evolutionary forces are studied. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Evolutionary design; Biologically inspired control systems; Genetic algorithms; Emergent systems; Artificial life

1. Introduction

The ‘design’ of biological systems and the ‘evolution’ of engineering artefacts have interested the public at large for some time but our understanding of both these areas is currently undergoing a period of rapid development. In biological sciences genetic engineering is having a fundamental impact because deliberate, planned changes can be made to an entity’s genotype. In engineering, a growing ability to analyse and model the process of evolution has made changes to the way the design and development of artefacts and systems are understood. Furthermore, engineers are also becoming interested in looking directly at how biological systems function in an effort to design manufacturing systems that exhibit adaptive emergent behaviour.

Automated strategies based on evolution, e.g. genetic algorithms (GA), genetic programming (GP) and other evolutionary algorithms have been the focus of attention in the automation of some shape design and planning functions. Architectures for learning complex control strategies have also been an area of interest for many engineers through the application of neural nets. At the same time animal behaviourists, psychologists, game theorists, economists, mathematicians and others have sought to produce a genetic explanation of the emergence of strategies for competition and co-operation between entities. The artificial life community has grown, fuelled by increased computing power and a better understanding of the agent-based approaches to problem modelling often learned from the emergent behaviours of bee or ant colonies. Now engineers have become interested in using these paradigms in the hope that mechanisms might be learned from nature that can in turn be used to make manufacturing systems adaptive and self organising.

Manufacturing facilities and biological entities are both examples of complex systems that operate in
rapidly changing environments. The purpose of this paper is to examine current areas of interest in which analogies from nature are being claimed to have direct relevance to engineering systems. Fig. 1 shows the major stages and some of the biological analogies that are considered in relation to them.

2. Evolutionary design

There has been a long held popular view that technical processes such as design may be partially controlled by evolutionary forces. It is well known that nature has produced complex shapes that are well adapted to their environment. The mechanism for this evolution of shape is not well understood however. Near perfect geometric shapes may be produced but there are no drawings or mathematical expressions that are encoded in any explicit way in an organism.

In contrast, the designs of most artefacts are supposed to be explicitly encoded through such mechanisms as engineering drawings or less commonly, through prescribed process plans. When the evolution of products is talked about the actual mechanism of selection, inheritance and reproduction is not necessarily clear and the analogy with biological evolution is incomplete.

Biological evolution works through replicators and their mutation in the presence of selection. In our normal understanding of the process with regard to biological agents, the replicators are the genes that are encoded in the DNA. Near perfect copies of the individual’s set of chromosomes are in every cell in the body. The DNA holds the necessary information required to make proteins and because of this it has been likened to a recipe [7], albeit a very complex one. The exact details of what proteins to produce and when are conditional upon environmental factors which are evident in a cell’s chemical surroundings. Thus a body’s development is dependent on both its genetic code and its environment. The success of the genes held in the chromosomes are influenced by the success of the organism they have built in being able to ensure that copies of the genes are copied into future organisms. Commonly, organisms that are successful might breed more and produce more and ‘better’ offspring. In this way ‘good’ genes tend to propagate while ‘poor’ ones tend not to. If evolution does take place in engineering design it is unclear if similar mechanisms exist. Engineering drawings are not directly analogous to DNA replicators, the marketplace does not function directly as a selection environment and engineering products do not reproduce.
2.1. Product evolution

Replicators of engineering products might be considered to be attributes of the product such as shape elements, colour, surface finish or perhaps manufacturing process [22]. Importantly an engineering drawing does not fully specify the exact dimensions and appearance of an object. A period of manufacture is required to develop a part's plans into the part itself. This process is analogous to embryonic development.

At the simplest level a product may appear to be selected, i.e. bought or not bought. A more complex model might allow us to consider whether a product is bought 'more quickly' or bought at a higher price. Through the mechanisms of the market, demand might lead to the supply of further products and a process of selective reproduction may be realised. The selection decision of whether to reproduce part designs or re-use some of their attributes, however is a function of management. Only they can decide the 'fitness' of a part in terms of its ability to help management realise their goals for the firm. Decision parameters may be short-term profit, market share or other strategic goals.

Moreover, variants, e.g. of size and colour may also provide a form of diversity. Changes in the environment, evident in changing customer demand due, e.g. to changing fashion or weather, may favour previously less popular alternatives. The diversity therefore provides flexibility that might accommodate an ability to adapt.

Change in the biological world is made possible as a result of change either in the genotype through mutation or through characteristics that are acquired through an organism's development. This second type of change in characteristics cannot be passed on in reproduction, although in the 19th Century it was thought by Lamarck [7] and others that it could. In this area, the analogy between biology and engineering is weak. Firstly, whilst mutations do have an analogy in manufacturing (due to unforeseen effects resulting from design or process changes and which can therefore be regarded as being random), their timing is not random. Furthermore, most design or process alterations result in deliberate changes in the functionality of a product. There is a biological analogy evident in the practice of genetic engineering but it is fair to assume that this accounts for very little of biological 'design'. The practice of engineering design does make possible an analogy of Lamarckism. Acquired characteristics on a product may be evident due to a customer making changes, either deliberately or through use and these could make the product more useful or 'better' in some way. Subsequently engineers may incorporate such changes into the product when new through a design change (examples include pre-faded jeans).

2.2. Simulated evolution in design and planning

The evolution of a design may be accelerated on a computer where simulations of the environment provide the necessary feedback to allow selection to take place. Genetic algorithms can then be used to simulate analogies of biological processes associated with reproduction. These techniques are in abundance in the literature, e.g. see [3-5,8,10,16]. Mill et al. [14] give an overview of the use of shape representations for genetic algorithms in topology and shape design. Fig. 2 shows the results of one of the authors' test runs using GAs to design an aerofoil.

The use of GAs in the optimisation of process plans has been reported some 10 years ago by, e.g. Vancza [17], Husbands et al. [13], Husbands and Mill [12].

2.3. Biologically inspired heuristics

Another analogy that has been used in design uses a heuristic that appears to be borrowed from biological life. In the design of structural members some researchers have used the heuristic, 'put more material at points of high stress' [2,6]. This can be implemented in an automated design system using the finite element method to assess engineering stresses. Commonly, the process runs through many iterations applying a fictitious 'swelling strain' which is proportional to the difference between a node's von Mises stress and a base stress. The shape is then modified using the swelling strain. This continues until a convergence criterion is met. This is similar to the way that developing bones and trees appear to attempt to adjust their shapes in order to adapt to their environment, except that the process is continuous though limited in time. Vaario et al. [21] describes the use of a sophisticated development model that can achieve such design processes.
Fig. 2. A GA used to optimise an aerofoil.
3. The firm and the manufacturing system

It is not only in product design, however, that biological analogies can be used to explain phenomena. The nature of a manufacturing firm or any firm it has been argued, is also subject to evolutionary forces. Perhaps the mechanism for this is even less like that which biological entities find themselves subjected to. The basic argument goes thus. Firms are selected in the market place, good ones making more money and gaining more power and so expanding and perhaps creating offshoots. Unsuccessful firms are selected against and go bust. The meaning of reproduction here is clearly complex. Whereas in biological systems reproduction is carried out by making mutant copies of one or in sexual reproduction, two individuals, in the case of a new firm there may be many inputs to the genotype. A new firm may be an offshoot of several parents, inheriting their genes or perhaps more accurately, 'memes'. A meme is an idea or value that might be held [7]. These memes may be associated with management structures, management techniques or attitudes to workers and they may be installed in a company from the parent companies or they may be brought in from other firms through their employees. This is a particular feature where companies start up offshoots in other countries and often find working practices becoming evident which are imported from the local work force. Another aspect of this evolution is that it is perhaps Lamarckian. A firm on a daily basis acquires new characteristics and these may in turn be passed on directly to offshoot firms when they are set up.

It is the ability of the firms to acquire characteristics and respond to their environment that has recently become of interest to manufacturing engineers. Whilst the process of evolution may be operating in an important way it is fundamentally different to that which a biologist would recognise. If the workings of a firm are a result of evolutionary pressures, it is likely that this forms only a limited analogy with the biological evolution of entities. Rather than consider the organisation as characterised mainly by a top down structure where characteristics can be inherited, it is also possible to view a firm as a collection of individuals working together to achieve the firm's aims (through achieving their own). In classical economics it is assumed that the owner and therefore the firm seeks to maximise profit and that this approach explains the behaviour of the firm. However, firms do not always behave in the ways predicted by models based on this assumption and this has given rise to behavioural theories of the firm. These seek to explain the firm as a result of the coalition of its members, i.e. its owners, managers and workers.

The view of the organisation of any economic system as being made up of individual agents is not new but recently it has become realistic to think about modelling based on this view. An agent-based paradigm may be used to model a set of (usually identical) individuals who behave according to simple rules and this results in a whole system that can be observed. Most recently engineers have become interested in the development of manufacturing systems that exhibit emergent behaviours. This approach lends itself to the agent-based paradigm. The idea is based on the hive or ant colony model of a manufacturing system. Individual agents or ants can be programmed to exhibit simple behaviours that do not in themselves appear complex or even purposeful. A system consisting of the sum of these behaviours, however, could exhibit useful global behaviour. The analogy with biology here is clear. The approach is based on the assumption that beehives and animal colonies are similar, in some ways at least, to factories. Although the attributes of emergence and adaptability are found in biological systems, the means proposed so far for achieving these in manufacturing systems are quite different from the way they are achieved in nature [9, 18, 20].

Most of the work carried out in this area has been based around pre-programmed agents. These may have adaptive capabilities programmed into them and they may also be given some capability to learn. Groups such as those at The University of Sussex [11] have tried an alternative approach which breeds programmes using evolutionary algorithms. Their work is concerned with the behaviour of small mobile robots. The evaluation of performance of the robots is carried out on real world tasks, the best performers on tasks being selected for breeding. This approach uses genetic algorithms to evolve the control programs of the robots. Their approach is akin to the 'design' of biological beings. Biological systems, whether they are human bodies or beehives evolve by a process that receives no
feedback. The replicators that hold the design of an entity are in fact usually set aside long before the selection process takes place. In the case of an animal the reproductive cells are chosen early in embryonic development. The other cells in an animal therefore may be identical except for mutations but they cannot take part in reproduction. Their best means of transferring their genetic material is, therefore to work to ensure that their reproductive kin are successful. Similarly, in many beehives, the queen is chosen early with the worker bees being incapable of breeding. Because their genetic material is similar to that of the queen’s, their best chance of ensuring their own genetic material transfer to new generations is to work to enable the queen to breed successfully. This has often, wrongly been observed as the individuals working ‘for the good of their species’. Given the chance, a worker bee might attempt to mate and produce her own offspring. When this happens it is likely that her own kin will kill her offspring because genetically they have less in common with the offspring than with those of the queen [15].

These models of biological evolution are obviously fundamentally different than those so far proposed for manufacturing systems.

Another aspect of evolution that might be considered in the context of a firm is the area of strategic behaviour. Writers such as Axelrod [1], de Wall [23] and later Ridley [15] have shown how complex competitive and co-operative strategies develop in biological systems and how these complex strategies can also be transferred to the world at large, including our own and organisations’ behaviours. This work uses game theory as its basis, in particular the Prisoner’s Dilemma game and the strategies appear to explain, e.g. pricing and other behaviours, particularly in oligopolistic situations.

3.1. Biological manufacturing systems

The main attraction of the so-called ‘biological manufacturing system’ (BMS) is that it is potentially self-organising and adaptive. Interest in this concept is particularly strong in Japan where a group led by Fuji working in partnership with Fujitsu, Honda, Komatsu and Sony as well as academic groups at Kobe and Kyoto Universities have set up two projects to investigate the concept, ‘biological manufacturing systems’ as part of the next generation manufacturing systems project of the intelligent manufacturing systems programme and ‘modelling emergent synthesis’. The aims of these projects are to investigate the usefulness of imitating, within a manufacturing system, the self-organisation and evolutionary optimisation of a biological system.

In the BMS project the manufacturing system is seen as an organism which must respond to external stimuli and create products. The manufacturing system has ‘genetic information’ which describes the system. This project has two principal themes: ‘self-organisation’ and ‘evolutionary simulation of production network (ESProN)’.

In the first theme self-organisation is investigated as a method of controlling factory activities. Manufacturing cells are modelled as autonomous entities which can self-organise into an assembly line that can re-configure itself as requirements change or machinery malfunctions [18]. This work makes use of a virtual reality model of a factory [20]. An example application of this work has been the lineless production of a car chassis [19].

In the second theme (ESProN) product lifecycles and production networks have been modelled as evolutionary processes. Evolution is also used in the design of new products.

The second project ‘modelling emergent synthesis’, can also be split into two themes. In the first theme human participation is allowed with a manufacturing system which has the self-organising properties developed in the BMS project. The second theme looks at the emergence of new products by examining their interaction with the environment [22].

3.2. Santa Fe Institute

The Santa Fe Institute in the USA, is a multidisciplinary research institute dedicated to work in the area of complex systems. These are systems that can be characterised as involving numerous interacting parts functioning as a whole. They display emergent behaviour which is only seen at the level of the system. Such emergent behaviour would not be seen by observing only the parts. Many processes and artefacts in manufacturing, e.g. a manufacturing facility or an assembly of parts can be viewed as a complex
system. Equally, biological organisms and ecologies can be seen as complex systems. Research at the Santa Fe Institute ranges from economics to adaptive computing to molecular biology, seeking to develop an understanding of the fundamental properties and behaviours exhibited by such complex systems. Whilst most of this work is not directly related to manufacturing the insights gained are applicable both in creating adaptive systems for manufacturing and in using adaptive systems (often drawing on biological metaphors, e.g. genetic algorithms and neural nets) for design or control.

There is some work emanating from the Santa Fe Institute which is directly concerned with manufacturing. This involves looking at some specific applications of modelling and simulating complex adaptive behaviour in problems based on assembly lines and order fulfilment processes. They are also studying meme propagation during a business re-engineering project. Several other research groups (including the authors of this paper) are also active in this general area.

4. Conclusions

The entire life-cycle of a product from market research through design, planning, manufacture, supply chain and decommission involves many complex systems. These complex systems involve many interacting individuals, customers and processes. It is very difficult to model or predict the behaviour of such complex systems. Many biological entities can also be characterised as complex systems that have desirable properties such as self-organisation and adaptivity. There has therefore been interest in using biological analogies as a route towards understanding and using such complex systems. In this paper some of these analogies were discussed: modelling product change as an evolutionary process, viewing firms and manufacturing in an evolutionary context, insect societies as analogies for manufacturing systems. These analogies were found to provide some insights but were limited in providing complete explanations. The application of biologically inspired techniques, e.g. genetic algorithms and simulated biological growth, in design and planning were also discussed and were found to be useful.

References

[16] M. Schoenauer, Shape representations and evolution schemes,


Frank Mill graduated in 1983 from Napier College with a BSc (Honours: 1st class) in Technology with Industrial Studies. In 1988, after working at Coventry Polytechnic as a research associate he took up a lecturing post in the School of Mechanical Engineering at Edinburgh University where he has worked since. His research interests include Computer Aided Process Planning, Feature Oriented Design, Concurrent Engineering and Genetic Algorithms. He is a senior lecturer and is also currently working part time for a Computer Aided Design consultancy.

Andrew Sherlock graduated in 1995 from The University of Edinburgh with a BE (Honours: 1st class) in Electrical and Mechanical Engineering. Since 1995, he has been working at the School of Mechanical Engineering at Edinburgh University towards a PhD in shape optimisation, using genetic algorithms with a cellular shape representation. He is currently employed as a research engineer for Mathématiques Appliquées SA, developing industrial applications of adaptive technologies.
Diverless Weld Inspection and Repair Using ECM/ACFM Techniques

D. CLIFTON¹, F. MILL¹, J. ESNAOLA¹, R. KARE² and W. D. DOVER³

¹ School of Mechanical Engineering, University of Edinburgh, King's Buildings, Edinburgh EH9 3JL. ² TSC Ltd, 6 Mill Square, Wolverton Mill, Milton Keynes MK12 5RB, ³ Department of Mechanical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

Abstract

This paper describes the development and application of an integrated tooling/NDT system which provides the potential to undertake a weld inspection, carry out removal of the defective zone, and achieve validation of the repair, all within a single ROV deployment operation. Removal of the defect zone is achieved using a novel application of the electrochemical machining process (ECM). This system is integrated with the alternating current field measurement (ACFM) system, for crack detection, location and sizing. The paper describes the development of electrochemical machining process models to cover situations encountered in automated weld sampling operations. From these models, and data collected from experimental trials, it has been determined that the sampling/removal operation can be carried out at tool feed rates of up to 3 mm/min. The model has been extended for determining the change in gap dimensions for the case of zero tool feed. This relationship can be used as a control parameter when determining the degree of overcut required when inclusions are encountered and the tool is required to remain stationary for a certain duration. Tank trials of the system have been carried out, successfully demonstrating effective operation of the integrated strategy in a submersed environment.

1. Introduction

In many offshore oil and gas production platforms extension of use beyond the original design life is being sought. To achieve this, NDT inspections, and the application of defect repair techniques, will have to be undertaken. At shallow sea depths, the use of diver deployment can often be considered for such operations, although in many cases this is considered hazardous, cumbersome and costly. In some of these situations the use of ROV based inspection, repair and maintenance (IRM) systems would be considered advantageous. At depths greater than 200 m, ROV based IRM would be a common if not essential requirement.

IRM is commonly directed to structural welds. The first part of this operation involves the accurate detection of defects through the use of NDT technologies. In some cases a methodology is then deployed to remove the crack and repair the defective zone. NDT re-examination may then take place to assess the effectiveness of the repair operation. This paper describes the development and application of an integrated tooling and NDT system which provides the potential to undertake a weld inspection, carry out removal of the defective zone, and achieve validation of the repair, all within a single ROV deployment operation.

One way to remove a weld defect is by the application of conventional grinding systems. However, it has been reported [1] that the use of grinding can impart additional stress at the machined site and this can result in a residual surface which is prone to further crack growth. The system described in this paper achieves crack removal by a novel application of the electrochemical machining (ECM) process. This technique does not impart additional stress into the machined zone and is known to leave behind a surface of high structural integrity [2]. Also, because the process is essentially forceless it can readily be applied without the use of bulky mechanisms. Thus the system readily lends itself to ROV deployed operations. Furthermore, unlike grinding, the ECM system will enable a sample of the defective zone to be retained for subsequent metallurgical analysis. Such analysis can often provide an estimate of the residual life of a structure or be used to provide information about whether the defect was introduced at the time of manufacture or has arisen during service.

The basis of the ECM defect sampling system has previously been described [3]. This paper describes the development of process models to cover situations encountered in automated weld sampling operations. The model is used to provide parameterisation of the sampling process which is then validated through experimental trials. The paper describes the systems re-engineering for submersed operations and integration of the ECM tool with the alternative...
current field measurement (ACFM) system, for crack detection, location and sizing. The tank trials of the integrated assembly are described.

2. ECM Sampling Process Model Developments and Validation

ECM is carried out by applying an electric potential across an electrolyte flowing in the gap between an electrode tool (biased as the cathode) and the welded structure (biased as the anode). This arrangement forms an electrochemical cell through which electric charge flowing as a result of the applied potential causes electrochemical dissolution of the weld zone in the immediate region of the cathode tool. The tool electrode is then traversed in an appropriate way so as to remove and/or sample the defect. For general background and overview on the ECM process see references [3] and [4].

The ECM sampling tool bit consists of a section of tubing shaped so that the outer cutting face forms an approximate inverse of the required sample site profile. This has commonly been a 'U'-shaped section of the form shown in Figure 1. The tube is slotted around the forward facing cutting surface to enable the supply of electrolyte into the active cutting zone in the direction of feed. During the sampling operation the tool is fed into the workpiece along a 'boat-shaped' profile. When set to machine a deep cut, the operation can be configured to remove a sample of material containing the defect. In the case of shallow depths of machining the process is used to dissolve the surface of the workpiece to remove the defect.

The important design constraints have been the degree by which the tool tip is deflected during the sampling operation and the accuracy of the cut. These constraints have been examined through the development of both a tool deflection and a cutting dynamics model, as described in the following sections.

2.1 Tool deflection model

An important dependency with supply pressure will be the tool tip deflection; defined as the distance that the bottom tip of the tool is deflected in the opposite direction of tool feed, caused as a reaction to the momentum of the electrolyte flow rate in the direction of tool feed. This is considered as an important variable since if for any reason the supply was interrupted during machining then the tool tip would relax back to its non-pressurised position. If the deflection distance was greater than the tool to workpiece gap (in the direction of feed) then it would be possible for physical contact to occur between the tool tip and the workpiece. This is considered undesirable since tool damage may result if the situation is not recognised and the tool power remains on. To incorporate a fail-safe feature in this respect, a design criterion has been set to ensure that the tool deflection is less than the equilibrium gap. This criterion would enable re-entry of the tool path in cases where, for some reason, the original profile was not complete.

The tool tip deflection model, described in detail in reference [5], calculates the forces acting to deflect the tool due to the outflow of electrolyte, by calculating the change in fluid momentum at the supply slot. From this force, the strain energy as the function of the modulus of elasticity of the tool material is determined. Castigliano's theorem [6] is then applied to determine the maximum deflection occurring at the tool tip calculated as the sum of the individual deflections occurring at the curved section of the slot, the straight slotted section and the straight unslotted section of the tool.

The tool deflection model has been used to calculate the degree of tool deflection in terms of the physical tooling parameters of outside diameter, o.d. (or $2r_1$, where $r_1$ is the tool radius), the internal diameter, i.d., and electrolyte supply pressure. An important interdependence has been the relationship between the electrolyte supply pressure and level of electrolyte flow rate. The flow of electrolyte is required to remove solid and gaseous products of the dissolution process (i.e. mainly oxides of the metal being machined). There will exist a certain level of electrolyte flow rate, for a given level of forward feed rate, at which dissolution becomes unstable. Initial trials have shown a linear dependency between the limit of stable dissolution and the required flow rate of approximately 2.01/min for every 1 mm/min velocity of the tool feed.

2.2 ECM sampling process cutting models

The primary cutting parameters are the frontal gap, $y$, maximum overcut, $g$, and the maximum cut width, $w$, where it can be seen from Figure 2 that $w = 2r_1 + (2g)$. Note that the overcut will be in all directions around the tool, and along the
length of the tool traverse, except at the tool face in the direction of feed. In this case the gap between the tool and workpiece is the frontal gap, \( y \), which after a short period of machining, once equilibrium conditions have been established, becomes equal to the equilibrium gap, \( y_{eq} \). The equilibrium gap will be the closest approach between the tool and the workpiece in the direction of the feed, and is therefore the parameter that must be quantified in relation to the tool deflection model. The process overcut (which corresponds to the accuracy of the cut) can be of critical importance when the wall thickness at the weld site is marginal within operation requirements.

A theoretical analysis of overcut and frontal gap can be made by considering the dynamic equations that relate the rate of material removal to the tool velocity and other set process parameters. This relationship can also be used to estimate the degree of overcut that will occur when the tool velocity is zero. Such a situation would occur when the tool encounters a non-conducting inclusion within the weld. In this case the tool is held stationary in order to increase the degree of overcut so as to machine around and free the inclusion from the surrounding material. The following theoretical analysis considers and develops relationships to address these situations.

Although seawater could be used as the machining electrolyte, the concentration is usually too low (at around 3%) to enable significant dissolution currents. The achievable removal rates would therefore be limited using seawater for the machining electrolyte. To enable relatively high removal rates, machining was carried out using a 15% NaCl solution made up in a separate tank and pumped to the tool tip. Because enriched electrolyte is directed only in the direction of tool feed, the gap model has been developed with the assumption that only the forward facing gap contains electrolyte at the enriched conductivity. The trailing face of the tool will only be in contact with a solution at the surrounding ambient concentrations and therefore components of erosion due to the trailing half circumference of the tool will be small, and have been neglected.

Figure 3 shows the two erosion situations that occur during the sampling operation; Figure 3(a) represents the situation for a finite constant tool feed and Figure 3(b) is a representation of the situation when the ECM tool bit is held static. The erosion model is first developed to determine the frontal equilibrium gap, \( y_{eq} \), as shown in Figure 3(a), by considering the basic equations of ECM dynamics developed for the case where the tool surface is perpendicular to the direction of feed and where the lines of electric flux are parallel to the direction of feed. This situation occurs along the \( y \)-axis (representing the relative position of the workpiece surface referenced to the tool surface) as illustrated in Figure 3(a). In this area, the gap between the tool surface and the workpiece surface can be calculated using a general dynamic process description developed as follows.

Considering Figure 3(a), the rate of dissolution, as determined by Faraday’s laws of electrolysis, is proportional to the current density, \( J \), flowing in the electrolyte solution between the tool and workpiece surfaces. The current density, given
by \( J = \kappa(V - V_0)/y \), can also be expressed in terms of the erosion rate, \( dy/dt \), of the workpiece. Subtraction the change in position of the tool surfaces (i.e. the feed rate, \( f \)), moving along the \( y \)-axis, gives:
\[
\frac{dy}{dt} = \frac{J\varepsilon}{\rho F} - f = \frac{\kappa(V - V_0)\varepsilon}{\rho F y} - f
\]
or
\[
\frac{dy}{dt} = \frac{k}{y} - f
\]
where \( k \) is a machinability constant given by:
\[
k = \frac{\kappa(V - V_0)\varepsilon}{\rho F}
\]
\( y \) is the gap measured along the \( y \)-axis as shown in Figure 3(a), \( V - \Delta V \) is the voltage available to drive the current through the electrolyte (\( V \) is the applied machining voltage and \( \Delta V \) is the total voltage loss at the tool and workpiece surfaces), \( \kappa \) is the conductivity of the electrolyte, \( \rho \) is the density of the workpiece material, \( \varepsilon \) is an electrochemical parameter calculated as the ratio of the average molecular weight of the workpiece material to the average valency of the ions precipitating into solution. It is assumed that the \( \kappa, \varepsilon \) and \( V_0 \) are constant and independent of other process parameters; \( F \) is the Faraday \((96,487 \text{ Cmol}^{-1}) \) and is a constant by definition.

In the limit \( t \rightarrow \infty, dy/dt = 0 \), equilibrium conditions are reached and \( y \rightarrow y_{eq} \), the steady state gap value from which any steady state gap function can be formulated as:
\[
y_{eq}(V, f) = \frac{\kappa(V - V_0)\varepsilon}{\rho F f}
\]
Equation 2 can be used to determine the equilibrium gap conditions as a function of set gap voltage and feed rate.

Examining the static tool condition illustrated in Figure 3(b) it can be seen that in this case the lines of electric flux, at equilibrium after a short period of machining, will be normal to the tool surface at all points on the tool. The gap can thus be considered to be equal to the difference between the radius of the cut \( r_c \) (the width of cut being \( 2r_c \)) and the tool radius \( r_t \), as illustrated in Figure 3(b). The volume of electrolyte between the tool/workpiece surfaces will increase with time, as the gap opens out, and hence the resistance of the electrolyte, \( R \), in the gap, can be obtained by integration, thus:
\[
R = \frac{1}{2\pi} \int_{r_t}^{r_c} \frac{dr}{r} = \frac{1}{2\pi k} \left( \ln \frac{r_c}{r_t} \right)
\]
The current density at the workpiece surface, \( J_{rc} \), is given by:
\[
J_{rc} = \frac{V}{2\pi r^2} = \frac{kV}{r_c \ln (r_c/r_t)}
\]
And the rate of metal removal at the workpiece surface is given by:
\[
\frac{dr_t}{dt} = \frac{eJ_{rc}}{F\rho} = \frac{k}{r_c \ln (r_c/r_t)}
\]
Where \( k \) is the machinability constant as defined above.

Equation 3 can be solved, by integration, using the boundary conditions that \( r_c = r_t \) at \( t = 0 \) and \( r_c = r \) at \( t = t \). Experimentally, the former boundary condition would correspond to touching electrodes, but in practice the experimental condition \( r_c = r_t + \delta r \) is used, with \( \delta r \) made sufficiently small that \( r_c \approx r_t \). This can be achieved by allowing the electrodes to touch, then separating them a small distance \( \delta r \) to prevent shorting. This gives the non-explicit relationship between the radius of the cut, \( r_c \), and time, \( t \), as follows:

\[
G-5
\]
Solving the equality of equation 4 for $r_c$ in terms of $r$ for a fixed value of $r_t$, will enable the interval of time for a required gap radius to be determined, for the case of zero tool feed.

In Section 5, the relationships developed above are applied to parameterise the defect removal/sampling process.

3. The ACFM Array Probe

Prior to defect sampling/removal, sizing and locating the defect is achieved using the ACFM system. The technique uses a uniform input current, arranged to flow in a direction normal to the expected crack direction, and requires measurement of two components of magnetic field, one parallel to the crack ($B_x$) and one normal to the surface under inspection ($B_z$). The presence of a surface breaking crack perturbs the uniform current flow, which in turn produces perturbations in both components of the magnetic field measured by the sensors. The $B_x$ field responds to the surface current density, which decreases locally by an amount determined by the crack depth. The $B_z$ field responds to circulation in the current flow, which is in a clockwise direction around one crack end and anticlockwise around the other end. In this way, the $B_z$ signal exhibits a peak and trough, the separation of which determines the length of the crack.

In order to inspect the machined zone produced by the ECM tool, a purpose-built probe was produced. This has an interchangeable nose sized to fit the particular radius groove being cut. The nose produced contained eight sensor coils (four $B_x$ and four $B_z$) arranged around a semi-circular tip with an 8 mm radius. The array probe body included ‘O’ ring seals around the nose and on the removable side plate to prevent water access down to depths of around 10 m (this was the maximum depth required for the tank trials of this system).

4. ECM and ACFM Probe Carriage Assembly

A detailed view of the ECM/ACFM carriage/scanner assembly is shown in Figure 4. Both the ACFM probe and the ECM sampling tool are positioned within a single scanner assembly. The scanner assembly is driven by a single screw and nut arrangement, and guided via two guide rods positioned symmetrically about the drive screw. The drive screw, seen in Figure 4, is driven via a DC servo motor situated in the end enclosure. The end closure is sealed to operate at pressure in a submersed environment. The sampling tool and the ACFM probe are attached on opposite sides of a single sliding carriage to enable movement along a vertical axis. The sliding carriage is positioned via the position of a roller riding on a cam. The cam has two positions which are enabled by the hydraulic cam actuators. When positioned in a vertical orientation, during the ECM cutting phase, the cam acts to hold the ECM tool in position so that actuation of the feed motor moves the ECM tool along the required weld profile. When the cam is in the horizontal position the ECM tool retracts to its home position, and movement of the main screw then enables the ACFM probe to move along the surface of either the weld, when searching for a defect, or the sampled site, when the system is being used to validate the repair. The roller is of sufficient width and the cam has rounded edges to ensure smooth relocation of the roller and cam between the ACFM and ECM modes.

Figure 4 Schematic overview of ECM tooling and ACFM probe mounted on carriage/scanning assembly.
5. System Testing and ECM Tool Parameterisation

System testing and ECM tool parameterisation has been carried out for a range of process variables and two different tool configurations, made from copper tubing, as follows: tool A, 1.47 mm o.d./1.1 mm i.d. and tool B, 1.24 mm o.d./0.85 mm i.d. These sizes have been chosen as representing what is considered to be the lower and upper limit of outside tool diameter likely to be used in most practical situations. The tubing has been bent into a 'U' shape with a diameter across the machining face of 15 mm. The electrolyte supply slot runs across the face to cover a maximum machining depth of 10 mm from the surface of the weld.

Gap geometries and tool deflections have been measured by interrupting the process at the required conditions and then casting (using Permadyne dental casting paste) the region between the tool tip and the machined surface. The geometry is then measured by first sectioning the casting and then using a travelling microscope to read off the required dimensions. The angle that the cutting face makes with the horizontal reference position that the tool retracts to when the flow is turned off, is used to deduce the tool deflection during cutting when the flow is turned on.

Electrolyte solution has been made up as 15% wt/wt sodium chloride in water. The 15% concentration has been selected since it is easy to achieve such concentrations by agitation of the mixture for only a few seconds. In fact, some preliminary trials have suggested that this concentration could be achieved using an in-line salination system. If lower concentrations are used in subsea trials then the overcut and achievable feed rate will decrease, all other conditions remaining constant.

Trials have been undertaken using a workpiece in the form of a flat plate and also using flat T-butt welded sections. Figure 5 shows a typical example of the removed sample and sample site in the case of the T-butt welded section. These trials have been used to examine the following parameter trade-offs:

- Tool internal diameter against external diameter in relation to tool tip deflection for a range of pressure/flow characteristics. Any trade-offs considered for this relationship will be constrained by the requirement that the tool tip deflection should not exceed the forward cutting gap. The upper limit of tool deflection can then be used to confine other trade-offs.
- Inlet pressure against electrolyte flow rate. It is a requirement of ECM that the electrolyte is passed at high flow velocities into the machining gap. Such flow velocities should usually be in the region 2-4 l/min to achieve feed rates of between 1.0 and 3 mm/min.
- Machining current against feed rate. As the tool feed rate is increased the frontal cutting gap will decrease leading to a decrease in the gap resistance and hence an increase in the machining current. Limiting the machining current is of particular importance in a subsea environment where providing significant amounts of power could be cumbersome and expensive.
- Overcut against feed rate. The degree of accuracy of the sampling operation can be parameterised by the value of overcut. The accuracy of the operation will need to be specified if wall thickness tolerances at the sample are to be determined and held within specification.

6. Tool Parameterisation Results and Conclusions

Tool parameterisation trials have been carried out over feed rates of 1-3 mm/min, 3 mm/min being considered to be the maximum feed rate that can reasonably be sustained given the required trade-offs on other parameters. From these trials it was noted that stable dissolution can be achieved at up to 3 mm/min when using a flow velocity of 4 l/min, for tool A (this flow velocity limit is slightly lower for B, because of the reduced area at the machining face). If the flow velocity is decreased then the feed rate needs to be reduced by a corresponding amount. From the experimental plot shown in Figure 6, inlet pressures of about 20 bar are required, in the case of tool A, to produce the flow rate of 4 l/min (slightly lower for tool B). Although increasing the pressure beyond this would provide greater flow rates, this would be at the added expense of requiring higher integrity enclosure seals and increased pump capacity. Therefore, if 20 bar is considered as a reasonable upper limit of tool supply pressure, then as an initial parameter constraint, tool tip deflections can be considered for supply pressure of up to 20 bar.

Figure 5 Removed sample and sampled site on T-section weld.
Theoretical deflections are shown in Figure 7 plotted against variables of tool inside and outside diameters. The experimental points have been plotted with error bars of ±0.05 mm which is an estimated measurement tolerance. A close correspondence is noted between the experimental data and the theoretical surface in Figure 7, giving confidence in application of the model to other tooling configurations and process parameters.

To provide parameter ranges for tool deflection constraints, equation 2 has been used to examine the sensitivity of the frontal gap to the primary process parameters of $V$ and $f$. Plotting this function, using the parameters of $\varepsilon = 28$ and $\rho = 7.8 \text{ gcm}^{-3}$ for steel (as used in test trials, but these values should also be representative of the types of steels most commonly used subsea) and using a value of $\kappa = 0.2 \text{ Scm}^{-1}$ (corresponding to the electrolyte concentration of 15% sodium...
chloride as used in the trials), gives the gap plot illustrated in Figure 8. Measured values of the frontal gap from experimental trials, plotted for the case of 10, 15 and 20 volts, show good correspondence to the theoretical line, thus providing confidence as to the validity of the plot for other sets of parameters. From this plot it can be seen that at the upper limit of feed rate of 3 mm/min when machining at 20 volts, the frontal gap at equilibrium is 0.2 mm. The tool deflection for this condition is in the region of 0.4 mm (Figure 9). Thus the tool deflection exceeds the frontal gap for this condition. A decrease in the supply pressure would decrease the tool deflection, but only at the expense of reducing the flow volume and hence the upper limit of attainable tool feed. Thus if feed rates of 3 mm/min are required, then the machining voltage would have to be raised (to in the region of 35 volts) according to equation 2, to increase the machining gap to 0.4 mm.

Now considering the parameter of overcut, $g$, in the case of a finite tool feed; it is noted that as the feed rate, $f$, is increased then the frontal gap will decrease (equation 2) and the overcut, which is also a function of the feed rate, will decrease. A
direct analytical solution for determination of the overcut is not readily achieved (numerical minimisation methods would probably have to be used). However, using an exponential minimisation fit to the experimental measured values of the overcut, \( g \), as shown in Figure 9, the following parametric relationship has been obtained when machining with a sodium chloride electrolyte at 15%; \( g \) (mm) = 1.24 \( e^{-0.39f/(mm/min)} \) so that the total cut width is then given by; \( w = 2r_t + 2g \).

The overcut for the case of zero tool feed can be examined by applying equation 4 for the same conditions used for the plot of Figure 8, and for a range of values of \( r_c \). Solving equation 4 (using Mathad's Minerr function) gives the theoretical function of the increase in cut width with time, as shown in Figure 10. The fitting of experimental data to Figure 12 shows close correspondence.

Finally, experimental data can be examined to investigate the relationship between the feed rate and tool power requirements. This trend is illustrated in Figure 11 for tool A and tool B. The machining current, \( I \), will be slightly higher or lower for the larger diameter (tool A) or smaller diameter (tool B) respectively. The base current shown in Figure 11 is the electrolysis current that will flow before the feed rate is switched on. The value of 40 A occurred in this case because

---

**Figure 10** Theoretical and experimental data showing change in the cut radius, \( r_e \), for zero feed using tool A and tool B.

**Figure 11** ECM experimental current data for tool A and tool B shown with linear fit.
sampling head was fully submerged within a tank of electrolyte at a concentration of 15%. Some current would then flow between the workpiece and body of the sampling head. In the case of a subsea environment where the surrounding electrolyte will be at a concentration of only a few percent, the base current will probably be in the region of a few amperes. The important information in Figure 11 is the gradient of the relationship between feed rate and machining current, which will be the same regardless of the base current value. From this relationship, it can be seen that the machining current, for a 15 mm width tool machining at a depth of 10 mm, will rise in approximate accordance to: \( I(A) = 22f/(mm/min) + \) base current.

7. System Validation and Tank Trials
The system has been tested at a depth of 10 m using a tank testing facility at Oceaneering Ltd in Aberdeen. The test, shown in Figure 12, was carried out on a tubular node, submerged to a depth of about 10 m, in a tank of fresh water. The tubular welded T joint was made from BS4360 50D steel welded to typical offshore specifications. The T joint contained a fatigue crack at the hot spot stress site that had been produced using out of plane bend loading in a test rig at UCL. Figure 13 shows a close-up of the ACFM probe and the ECM sampling tool positioned over the weld toe. A 15% sodium chloride electrolyte was made up in a separate tank and supplied to the tool tip at a rate of 2 l/min. The sampling operation was carried out at a feed rate of 0.8 mm/min using a machining voltage of 12 V.

The ACFM array probe and ECM drive actuator were connected through a 25 m underwater umbilical to the monitoring and control system, situated at the side of the tank. The weld toe at the cord side was scanned to determine the location and size of the defect. With the ACFM probe deployed and the ECM tool retracted the probe was driven along the brace/cord intersection. At the end of the scan the ACFM software displayed the crack signal as shown in Figure 14(a). The troughs in the Bx and Bz plots are typical of a defect present at the toe. The crack was sized as 20.9 mm long and 2.7 mm depth.

Once the presence of the defect was confirmed the ACFM array probe was retracted. The hydraulics were then activated to deploy the cam so that the ECM tool moved into position. The ECM tool was set to cut out the defect which it successfully completed in 55 min. The ECM tool was then retracted and the cam actuated to bring the ACFM probe back inline for scanning along the repaired surface. ACFM re-inspection was then carried out to confirm if the ECM had successfully removed the crack. Figure 14(b) shows an ACFM scan of the repaired surface. The absence of peaks and troughs in the ACFM traces confirm the complete removal of the crack.

8. Conclusions
The development of ECM parameter models, together with data from machining trials, has enabled the sampling operation to be accurately parameterised. From these models, and the data given in Figures 6 to 11, the following overview of tool parameterisation can be made. The defect sampling/removal operation can be carried out at tool feed rates of up to 3 mm/min. However, at these upper feed rates the quantity of electrolyte needed to obtain stable dissolution is in the region of 4 l/min supplied at a pressure of 20 bar. At this condition the tool deflection will exceed the frontal gap limit unless the machining voltage is raised to a value in the region of 35 V.

As an example of overcut parameterisation, consider using a 1.24 mm outside diameter tool,
at a feed rate of 1 mm/min and a machining voltage of 15 V, using a 15% sodium chloride electrolyte, then the overcut will be 0.58 mm giving a total cut width of the order of 3.10 mm. Decreasing the overall cut width can be achieved by either decreasing the machining voltage or increasing the feed rate. However, increasing the feed rate must be accompanied by an increase in the flow velocity, by increasing the electrolyte pressure.

With regard to tooling parameters; larger internal diameters will provide greater flow output for the same supply pressure, but will consequently exhibit larger deflections. This problem can be offset by increasing the outside diameter of the tool, and hence the wall thickness, but this will be at the expense of increasing the cut width and the overall ECM power requirement.

The model developed for determining the change in gap dimensions, for the case of zero tool feed, has been successfully validated. This relationship can be used as a control parameter when determining the degree of cut required when inclusions are encountered and the tool is required to remain stationary for a certain duration.

The ECM defect removal/sampling system has been shown to be effective under remote control in a submersed environment. Integration of the ECM tooling with the ACFM system has enabled complete IRM operation to be achieved within a single deployment set-up.

Acknowledgements

The authors would like to thank the following: Andrew Holman at TSC Ltd, Milton Keynes, for carrying out the design and fitting of the ACFM probe array to the carriage assembly; Bob Gusthart and Derek Jardine for carrying out the design and manufacture of the equipment and for setting up and providing assistance at the validation trials; UCL for supplying some of the welded sections for validation trials; Oceaneering Ltd, Aberdeen, for providing facilities and backup personnel in support of the tank trials; and OSO and IMP for financial support.

References

A direct analytical solution to the tool design problem in electrochemical machining under steady state conditions

G M Alder, D Clifton* and F Mill
School of Mechanical Engineering, University of Edinburgh, Scotland, UK

Abstract: A method for the direct computation of two-dimensional electrochemical machining tool designs is described. The required workpiece geometry is represented by a Fourier series. Conformal transformation is then used to express the tool shape in series form, each term being a direct analytical function of the corresponding workpiece harmonic. Tool designs are thus achieved without numerical iteration. The model has been experimentally validated for a required workpiece geometry consisting of two harmonics, for which a tool was designed and manufactured. An In 718–15 per cent NaCl workpiece–electrolyte system was used to produce a machined surface, whose Fourier transform was obtained. The measured and predicted harmonic amplitudes agree closely. This harmonic design method is also shown to give insight into the relationship and limitations between tool design and achievable workpiece detail.

Keywords: electrochemical machining, tool design, conformal transformation

NOTATION

\( a_r \) coefficient of Fourier series

\( E \) applied electrode potential

\( F \) Faraday's constant

\( h \) nominal equilibrium gap

\( j_n \) normal component of current density

\( j_x, j_y \) current density components in the coordinate directions

\( J_{eq} \) nominal equilibrium current density

\( k \) electrochemical equivalent of the work material

\( L \) width of tool and workpiece surface

\( r \) an integer denoting harmonic in series

\( u \) electric field potential

\( V \) potential driving the dissolution current

\( U_0 \) total overpotential at the electrode surfaces

\( V_f \) tool feed velocity

\( x, y \) physical coordinates

\( Z \) complex function

\( \eta \) current efficiency

\( \theta \) local angle that the machined surface makes with the \( x \) axis

\( \kappa \) electrolyte conductivity

\( \rho \) density of the work material

\( \phi \) dimensionless electric potential

\( \psi \) dimensionless electric flux

1 INTRODUCTION

Electrochemical machining (ECM), in which metal is removed by electrochemical dissolution of a workpiece material, is a process that has the potential to produce complex shapes at high metal removal rates. A significant constraint in attaining improved manufacturing effectiveness in ECM applications is that of achieving the correct tool design for a specified workpiece profile. In this short communication a method for a new direct analytical approach to tool design is outlined.

Much work has been reported on the problem of ECM process simulation [1, 2], where the tool geometry is specified and the resulting workpiece geometry generated. The basis of these approaches is a solution of the field equations using finite difference, finite element or boundary element methods. The design problem, which is addressed here, is the inverse of simulation, attempting to determine the tool profile needed for a given workpiece surface. Previous workers addressing the design problem [3–5] have employed computational iteration to satisfy boundary conditions at both tool and
workpiece surfaces. The new approach is analytical and computationally very efficient when compared with the numerical optimization methods.

The method presented is applicable for two-dimensional tool design, in the case of equilibrium machining along a single axis. This is a common configuration in ECM applications. The method starts with a Fourier transform of the required two-dimensional workpiece shape. Conformal transformation is then used to express the tool shape analytically as a direct function of the harmonics of the Fourier transform. Although conformal transformation has previously been applied to ECM by, for example, Nilson and Tsuei [3], in this case the transformation was determined numerically. The present analysis is particularly applicable when the nominal equilibrium gap is known in advance (for specified machining conditions) or can be measured during machining. In this short communication the concept is described and then applied to a test shape for experimental validation.

2 THEORY

The two-dimensional configuration considered is shown in Fig. 1a. The tool surface AB moves downwards, parallel to the y axis with a steady feed velocity $V_f$. The coordinates $x$ and $y$ are non-dimensionalized by the width of both surfaces, $L$. The machined surface CD lies below the tool. The analysis relates to the electric field in the inter-electrode gap. The boundary conditions are constant potentials at the tool surface and the machined surface. The boundaries AC and BD are insulated.

The spacing between the two surfaces is expressed in terms of the nominal equilibrium gap $h$ defined as

$$h = \frac{\eta \kappa \kappa E}{\rho V_f} \tag{1}$$

where $\eta$ is the current efficiency, $\kappa$ is the electrolyte conductivity, $\rho$ and $k$ are the density and electrochemical equivalent of the work material, defined as $k = nA/F$, where $n$ is the valency, $A$ is the atomic number and $F$ is Faraday's constant. $E$ is the potential difference between the surfaces which can be decomposed as the difference between, $U$ is the potential driving the dissolution current and $U_0$ is the total overpotential at the electrode surfaces. The definition of $h$ is the nominal gap and would exist in the case where both the tool and the workpiece are planes normal to the feed direction.

The analysis below uses the dimensionless potential $\phi = (u/E)(h/L)$, where $u$ is the electric potential at any point in the field, and the dimensionless flux $\psi$ is defined as

$$j_x h = \frac{\partial \psi}{\partial y}, \quad j_y h = \frac{\partial \psi}{\partial x} \tag{2}$$

where $j_x$ and $j_y$ are the current density components in the coordinate directions. The current density components can also be derived from the potential:

$$j_x h = \frac{\partial \phi}{\partial y}, \quad j_y h = \frac{\partial \phi}{\partial x} \tag{3}$$

Thus, the following relationships can be obtained from equations (2) and (3):

$$\frac{\partial \phi}{\partial y} = \frac{\partial \psi}{\partial x}, \quad \frac{\partial \phi}{\partial x} = -\frac{\partial \psi}{\partial y} \tag{4}$$

The analysis follows the approach of Nilson and Tsuei [3], in which it is noted that $\psi$ and $\phi$ are orthogonal (because lines of flux are normal to lines of equi-potential), and that they constitute an alternative coordinate system (Fig. 1b). Equations (4) are the Cauchy–Riemann equations of complex variable theory and mean that the relationship between the two coordinate systems can be written

$$Z = Z(W) \tag{5}$$

where $Z$ is an analytic function relating the complex numbers $Z = y + ix$ and $W = \phi + i\psi$. Of the many possible functions $Z$, it is proposed here that letting $Z = W + e^W$ provides a particular basis for the ECM configuration of Fig. 1. This choice can be generalized as follows:

$$Z = W + \sum_{r=1}^{N} a_r (e^{r \psi W} + e^{-r \psi W}) \tag{6}$$
where the integer \( r \) is the wave number of a harmonic series and \( a_r \) is the corresponding constant coefficient. Application of equation (5) then leads to a direct analytical relationship between a desired machined surface and the tool required to make it.

At the machined surface, the rate of removal of material is related to the current density normal to the surface by Faraday's law:

\[
V = \frac{\eta j}{\rho}
\]

where \( V \) is the velocity at which the surface erodes (normal to the surface) and \( j_n \) is the normal component of current density at the surface. In steady state ECM, the local shape is maintained with corresponding points moving in the \( y \) direction at the tool feed velocity. Thus

\[
V = \frac{\eta j_n}{\rho} = V_f
\]

where \( \theta \) is the local angle that the machined surface makes with the \( x \) axis. Using equation (1) leads to

\[
\frac{j_n}{\kappa E \cos \theta} = 1
\]

The vertical component of current density is \( j_v = j_n \cos \theta \) so that, with substitution from equation (3), equation (9) becomes

\[
\frac{j_v}{\kappa E \cos^2 \theta} = \frac{1}{\cos^2 \theta} \frac{\partial \phi}{\partial x} = 1
\]

Equation (10) applies along the machined surface, which is a path of constant \( \phi = 0 \). By making a change in coordinates, it can be shown that:

\[
\left( \frac{\partial \phi}{\partial x} \right)_\phi = \left( \frac{\partial \phi}{\partial y} \right)_\phi \cos^2 \theta
\]

where the subscripts denote the path of partial differentiation. Thus, the boundary condition at the machined surface [equation (10)] becomes

\[
\left( \frac{\partial \phi}{\partial x} \right)_\phi = 1
\]

Equation (12) represents a uniform horizontal current distribution along the machined surface, which is consistent with the underlying model of steady state erosion.

When equation (6) is expanded, and real and imaginary parts equated, the following equations are obtained:

\[
x = \psi + \sum_{r=1}^N a_r \sinh(r \pi \psi) \sin(r \pi \psi)
\]

\[
y = \phi + \sum_{r=0}^N a_r \cosh(r \pi \psi) \cos(r \pi \psi)
\]

On the equipotential line \( \phi = 0 \), equations (13) become

\[
x = \psi
\]

\[
y = \sum_{r=0}^N a_r \cos(r \pi \psi)
\]

Equation (14a) satisfies equation (12), confirming that the line \( \phi = 0 \) represents the machined surface. The procedure thus starts with the decomposition of the desired machined surface profile into harmonic components, giving the coefficients \( a_r \) in equation (14b). Any other equipotential line \( \phi = \text{constant} \) in equations (13) then represents a corresponding tool surface lying at a particular value of the equilibrium gap.

3 EXPERIMENTAL VALIDATION

In order to validate the model, the double cosine case with \( a_1 = 0.5 \) and \( a_2 = -0.125 \) in equation (14b) has been considered. These coefficients become \( a_1 = 5 \times 10^{-3} \) m and \( a_2 = -1.25 \times 10^{-3} \) m when expressed in physical dimensions for a tool of \( 40 \times 10^{-3} \) m overall width, as shown in Fig. 2. Equations (13) were then applied to give the required tool profile for a nominal machining gap [as defined by equation (1)] of \( 0.6 \times 10^{-3} \) m. This tool profile was then manufactured.

The workpiece material was the nickel-based alloy In 718. Electrolyte was made up as a solution of 15% w/w NaNO\(_3\) in water. During machining the workpiece was fully enclosed within a flow cell that acts to direct and confine the electrolyte along the line of the workpiece surface (left to right in Fig. 2). Flow conditions were set to pass a constant 20 l/min through the machining gap. This rate has been determined from previous work [6], to be sufficient that the geometric distortions caused by variations in effective electrolyte conductivity, thought to be due to the accumulation of machining products [7], are less than \( 0.01 \times 10^{-1} \) m and can be neglected.

An initial set of calibration trials was undertaken to determine the nominal gap–voltage function according to

\[
h(U) = \frac{(U - U_0) \kappa}{J_{eq}}
\]

where \( J \) is the steady state current density.

From previous work [6] on the In 718–15 per cent NaCl system the constants \( \kappa = 0.22 \times 10^2 \) S/cm and \( U_0 = 3.0 \) V were obtained. At the fixed feed rate of \( 1.0 \) mm/min used in the experiments, \( J_{eq} = 70.4 \times 10^4 \) A/m\(^2\). The nominal gap was then set, by adjusting the voltage, according to equation (15). The value of \( h \) used was \( 0.60 \pm 0.01 \times 10^{-3} \) m at a gap voltage of \( 20.0 \pm 0.1 \) V. In order to characterize the sensitivity of
the analytical solution to deviations in the actual gap, additional machining operations were carried out at
gaps slightly below and above the nominal value of
0.60 ± 0.01 × 10⁻³ m. These were 0.40 ± 0.01 × 10⁻³ m,
using a voltage of 14.0 ± 0.1 V, and 0.80 ± 0.01 × 10⁻³ m using a gap voltage of 26.0 ± 0.1 V.

Machined profiles, plotted with the theoretical profile,
for the computed nominal gap at 0.6 × 10⁻³ m and the
additional profiles at gaps of 0.4 × 10⁻³ m and
0.8 × 10⁻³ m are shown in Fig. 2. A close spatial
conformance to the model workpiece shape can be
seen at the set gap of 0.6 × 10⁻³ m while, as would be pre-
dicted, at the gap of 0.8 mm the machined profile is wider
than the model shape and narrower for the 0.4 mm gap.
However, because the convergence to conformity
between the theory and machined profile is difficult to
quantify in the spatial domain, a series of Fourier trans-
forms has been applied to produce the coefficients \( a_1 \) and
\( a_2 \) for the experimental profiles. Convergence is then
quantified as the minimization of the difference

\[
\begin{align*}
\alpha_1(\text{diff}) &= \alpha_1(\text{theory}) - \alpha_1(\text{exper}) \\
\alpha_2(\text{diff}) &= \alpha_2(\text{theory}) - \alpha_2(\text{exper})
\end{align*}
\]

This gives \( \alpha_1(\text{diff}) \) and \( \alpha_2(\text{diff}) \) for the
0.4 × 10⁻³ m gap profile of 0.14 and 0.31 respectively
and for the 0.8 × 10⁻³ m gap profile of −0.13 and 0.27
respectively. This compares with the case of the
0.6 × 10⁻³ m gap, on which the actual tool design was
based, of 0.01 and 0.00 respectively. From these data,
close convergence to correspondence at the predicted
gap of 0.6 × 10⁻³ m can be clearly seen.

4 DISCUSSION

The experiments described in Section 3 above success-
fully demonstrate the application of the theory outlined
in Section 2. This success arises even though the theory
uses an ideal model covering only the electric field with
assumption of uniform gap conductivity. This is in line
with the set experimental conditions that non-ideal
effects become small at high flowrates.
An important advantage of the design method described here is that larger machining gaps can be contemplated, allowing higher electrolyte flowrates, with consequent benefits with regard to the removal of gaseous and solid products. The harmonic design method also gives insight into the profile relationship between an ECM tool and the workpiece. For example, Fig. 3 shows a set of tool surfaces, generated by equations (13), for equilibrium gap values $h/L = 0.025, 0.05$ and 0.075. It can be seen that, as the gap is increased, the tool shape needed to generate the required profile becomes more distorted, eventually reaching a limiting state, beyond which physically realizable tools are not possible. This limit is in line with ECM experience, where it is well known that fine detail cannot be copied across significant machining gaps. Inspection of equations (13) shows that limiting cases arise when the sinh and cosh terms significantly distort the original harmonics of equation (14b). That happens at the larger gaps and/or higher wave numbers. The example in Fig. 4 shows a workpiece consisting of straight line and radius segments. The harmonic decomposition leads to a large number of high wave number components whose amplitudes, although initially small, are amplified and distorted by equations (13) when the tool shape is calculated (Fig. 4a). Filtering the workpiece surface leads to a feasible tool shape (Fig. 4b), but at the expense of some loss of workpiece detail. However, the example illustrates the ability of equations (13), not only to generate tool profiles, but also to provide information about how feasible it might be to machine a given geometry in the first place.

REFERENCES


