
Thompson, Mary Kathryn

Publication date: 2011

Document Version
Publisher’s PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Interdisciplinary Design

Proceedings of the 21st CIRP Design Conference

Mary Kathryn Thompson
Editor
21st CIRP Design Conference
Interdisciplinary Design

27-29 March 2011, KAIST, Korea

Organized by
Korea Advanced Institute of Science and Technology
Conference Chairmen
N P Suh KAIST, Korea
M K Thompson KAIST, Korea

International Program Committee

CIRP
A Bernard, France
D Brissaud, France
C Brown, United States
J Corbett, United Kingdom
J Duflo, Belgium
H ElMaraghy, Canada
W ElMaraghy, Canada
A Fisher, Israel
L Galantucci, Italy
P Gu, Canada
K Hon, United Kingdom
F Houten, Netherlands
J Jawahir, United States
B Kaftanoglu, Turkey
S Kim, Korea
S Kim, United States
F Kimura, Japan
T Kjellberg, Sweden
F Krause, Germany
S Kumara, United States
B Lauwers, Belgium
S Lu, United States
E Lutters, Netherlands
P Maropoulos, United Kingdom
N Masayuki, Japan
L Mathieu, France
L Monostori, Hungary
D Mourtzis, Greece
A Nee, Singapore
R Roy, United Kingdom
G Seliger, Germany
M Shpitalni, Israel
G Sohlenius, Sweden
N Suh, United States
M Thompson, United States
S Tichkiewitch, France

T Tolio, Italy
T Tomiyama, Netherlands
M Tseng, Hong Kong
J Vancza, Hungary
B Wilhelm, United States
D Yang, Korea

Non CIRP
M Cantamessa, Italy
G Cascini, Italy
D Cavallucci, France
J Clarkson, United Kingdom
A Duffy, United Kingdom
Y Jang, Korea
D Lee, Korea
T Lee, Korea
L Leifer, United Kingdom
J Morrison, Korea
A Riel, France
R Stark, Germany
B Tjahjono, United Kingdom
P Xirouchakis, Switzerland
P Zwolinski, France

Local Organizational Committee
C Brown, United States
Y Jang, Korea
S Kim, Korea
H Kwak, Korea
T Lee, Korea
J Morrison, Korea
M Pena, Korea (Secretariat)
H Rosas, Korea
M Thompson, Korea
D Yang, Korea
Forward

The first CIRP Design Conference was held at MIT in 1990. A great deal has been achieved in design research for production and manufacturing since that first meeting. This can be seen in the dramatic increase in the number of research papers written on the subject in the past two decades and in the continued success of meetings such as this. This field has also seen, and will continue to see, many changes.

In today's world, system design and issues associated with large systems such as complexity and integration are critical. As such, the importance of cooperation, collaboration, integration and exchange of information between different elements of these systems (particularly between hardware and software), different parts of the design team, and different phases of the design process will only increase. Academic research must rise to address these new challenges.

These changes are also causing engineering design research, which has always spanned the silos of technical knowledge, to reach beyond engineering and become truly interdisciplinary. For this reason, the theme of the 21st CIRP Design Conference is Interdisciplinary Design.

The 21st CIRP Design Conference features 35 technical papers and 2 keynote presentations by leading scholars from academia and industry from 17 countries around the world. The meeting will discuss research topics from production and manufacturing, mechanical engineering, biomedical engineering, industrial design, and architectural engineering including: cost, value, complexity, sustainability, integration, computation, information, robustness, creativity, collaboration, rapid prototyping, virtual design, and green transportation. The papers included in this book include contributions from the CIRP community and beyond.

We would like to take this opportunity to thank the CIRP STC Dn and the leadership of the CIRP for their support of this conference. We would also like to thank all of the authors for sharing their exceptional research with us, the members of the international program committee for their support in reviewing the papers, and the local organizing committee for their assistance with conference preparations. We would especially like to thank Ms. Jieun Choi, Ms. Monica Pena, and Mr. Harvey Rosas for their tireless efforts. Finally, we would like to thank the Korea Advanced Institute of Science and Technology, the KAIST Department of Industrial and Systems Engineering, and the Daejeon Convention and Visitor's Bureau for their financial and logistical support for the conference.

President Nam P. Suh

Mary Kathryn Thompson
Chairmen, 21st CIRP Design Conference
# Table of Contents

## Integration and Sustainability

An Integrated Approach to Sustainable Product Development at the System Design Stage  
C.H. Chu, J.C.P. Su, Y. T. Chen .......................... 1

Identification of the Optimal Design Based on Evaluation of Product Configurations  
Considering Specific Product Adaptabilities  
J. Man, L. Zhang, Y. Chen, D. Xue, P. Gu ......... 9

Generating Optimal Disassembly Process Plans from AND/OR Relationships using a Hierarchical Genetic Algorithm  
R. Edmunds, M. Kobayashi, M. Higashi ............. 16

Integrated Design for Solving Imaginary Complexity in Design  
K. Pimapunsri, S. Tichkiewitch .......................... 24

Obstacles and New Opportunities for Integrated Design  
L. B. Jensen, T. R. Nielsen .............................. 32

## Integration

Design of an Integrated Process Chain to Manufacture Titanium Components with Micro Features  
A. V. Vargas, U. Berger, H. N. Hansen .............. 37

Conceptual Product-Service System CAD - An Application  
G. V. Annamalai Vasantha, R. Roy .................... 46

Developing an Integrated Design Strategy for Chip Layout Optimization  
W. W. Wits, J. M. Jaureguı-Becker, F. E. van Vliet, G. J. te Riele ................................. 55

## Innovation and Creativity

Cognitive Designers Activity Study, Formalization, Modelling, and Computation in the Inspirational Phase  
C. Bouchard, J. E. Kim, J. F. Omhoever, A. Aoussat ............................................. 63

A Process-Based Model of New Venture Creation: Toward Modelling a Practical Application of Extant Theory using SADT Diagrams  
T. Hernandez Rico, P. Dubois, M. Le Coq ....... 69

Function Representations in Morphological Charts: An Experimental Study on Variety and Novelty of Means Generated  
J. L. Richardson III, J. D. Summers, G. M. Mocko ....................................................... 76

A Synthesis Decision Framework for Early-Stage Innovative Design  
S. C.-Y. Lu, A. Liu .................................... 85

## Value, Cost and Risk

An Approach to Improving the Cooperation between the Departments of Product Design and Technical Production Planning  
L. Weyand, H. Bley ....................................... 93

Risk Reduction via Prototyping in Customized Product Development  
K. Sun, S. L. Chen ........................................ 101

Evaluating the Effects of Poorly Performed Product Development Phases on Customer Satisfaction  
Y. Borgianni, G. Casaini, F. Rotini ................. 108

State of the Art - Lean Development  
U. Dombrowski, T. Zahn, S. Schütze ............... 116

## Robust Design

Robustness of Inventive Design Solutions - Transferring the Robust Design Focus from Production Process to Development Process  
F. Wuttke, M. Bohn, N. A. Suyam-Welakwe, A. Bohn .................................................... 123

Approaches to the Design of Micro Mechanical Systems  
H. N. Hansen ............................................. 130
Random Design: A Design of Experiment Method to Integrate Geometrical Deviations throughout the Product Lifecycle into Performance Simulation
F. Vignat, D. S. Nguyen, D. Brissaud..............135

A Discrete Geometry Framework for Geometrical Product Specifications
M. Zhang, N. Anwer, L. Mathieu, H. B. Zhao..........................142

Complexity
Individual Contour Adapted Functional Implant Structures in Titanium
C. Schoene, R. Stelzer, P. Sembdner, J. Markwardt, B. Reitemeier, G. Engel.............149

A CAD Based Tool for the Support of Modular Design
J. Pandremenos, G. Chryssolouris..................154

Modeling the Structure and Complexity of Engineering Routine Design Problems
J. M. Jauregui-Becker, W. W. Wits, F. J. A. M. Van Houten.............................159

Lead Time Reduction and Efficiency Enhancement Show Strong Interference with Customer Constraints in Banking Service Process Design
M. Nakao, M. Okano and D. Dvivedi.............167

Axiomatic Design for Understanding Manufacturing Engineering as a Science
C. A. Brown........................................173

Axiomatic Design of Carbon Composite Bipolar Plates for PEMFC Vehicles
H. N. Yu, J. W. Lim, M. K. Kim, J. D. Suh, and D. G. Lee.................................176

Virtual Design
Which Virtual Reality Environment for Usage Evaluation of Innovative Surgical Instrument in Minimally Invasive Surgery?
G. Thomann, Q. V. Dang, J. Tonetti, F. Villeneuve..................................180

Exploring the Potential of 3D Visualization Techniques for Usage in Collaborative Design
W. W. Wits, F. Noël, C. Masclet.........................187

A Virtual Assembly Approach for Product Assemblability Analysis and Workplace Design
Z. Y. Ding, K. K. B. Hon, F. Shao....................194

Computation and Information
Using Existing Standards as a Foundation for Information Related to Factory Layout Design
D. Chen, M. Hedlind, A. von Euler-Chelpin, T. Kjellberg...............................199

Product Lifecycle Management Model for Design Information Management in Mechanical Field
J. Le Duigou, A. Bernard.................................207

Method Applying System Models as Catalog Entities for Emergent Motion Design
T. Brett......................................214

Special Session on OLEV
Design of Wireless Electric Power Transfer Technology: Shaped Magnetic Field in Resonance (SMFIR)
N. P. Suh..............................................222

Application of Shaped Magnetic Field in Resonance (SMFIR) Technology to Future Urban Transportation
I. S. Suh..............................................226

Magnetic Field Design for Low EMF and High Efficiency Wireless Power Transfer System in On-Line Electric Vehicles
An Integrated Approach to Sustainable Product Development at the System Design Stage

C. H. Chu*, J. C. P. Su, Y. T. Chen
Department of Industrial Engineering and Engineering Management, National Tsing Hua University
101 Kuang Fu Rd, Sect. 2, Hsinchu 300, Taiwan
*chchu@ie.nthu.edu.tw

Abstract
This paper presents an integrated approach to reducing the environmental impact of product development at the system design stage, by a simultaneous consideration of product design, manufacturing, and the supply chain. The approach incorporates a number of factors that ecologically influence the product lifecycle activities into product architecture design. CAD-based functions are developed to systematically generate manufacturing bills of material (BOM) by varying these factors. Lifecycle assessment is applied to evaluate the environmental impact of the generated BOM's. Optimization schemes guide the variation process to search for optimal results. This work realizes computer-aided sustainable product development by offering a system design tool of ecological decision making.

Keywords:
Lifecycle assessment, sustainable design, product architecture, bill of materials, product development

1 INTRODUCTION
Environmental issues like global warming and energy consumption have become an imperative for the contemporary world. However, current product development activities in manufacturing companies are still mainly driven by cost/profit analysis. A concern is that to integrate the eco-design principle into product development could impose additional design constraints and costs. Most ecological design methods were developed as assessment tools that can only estimate the environmental impact of an existing product or finished design [1]. They are difficult to be integrated into modern product development.

Grote et al. [2] proposed an eco-design method that helps a design engineer make design decisions without a trade-off on economic issues, with a focus on concept design and detail design phases. Feldmann et al. [3] proposed Green Design Advisor (GDA) that computes an overall score for environmental impact using multi-attribute value theory, considering metrics related to the number of materials, materials used in the product, and the disassembly and recyclability of the product. Mascle and Zhao [4] evaluated the environmental impacts for parts, assembly and operations during material extraction, material processing, manufacturing, usage and product disposal using feature modelling techniques. Most previous sustainable design methods were developed for facilitating decision making in the detail design stage. Product architecture has a profound impact on the entire product lifecycle [5]. Most of the past studies investigated the influence of product architecture from the perspectives of design for assembly/disassembly. Product architecture has been identified as the crucial factor that links product design and supply chain activities for environmental decision makings [6-7]. Supply chain considerations should be incorporated early in the design process to ensure the greatest possible reduction in environmental impacts. However, fewer methods have been developed to reduce environmental impacts with approaches at the system design stage.

In this research, we have developed an integrated framework that reduces the environmental impact of product development which considers design, manufacturing, and the supply chain at the same time. This framework incorporates a number of factors into the system design stage that ecologically influence the product development activities. CAD-based functions are applied to generate manufacturing BOM's (bills of material) by varying these factors. The environmental impact of each generated BOM is estimated based on LCA databases for three product development stages: raw material production, part assembly/manufacturing, and the supply chain. Optimization schemes are integrated with the variation process to search for the optimal result with a minimized amount of CO₂ emission. A real bicycle design is tested to validate the proposed framework. The test results show that it provides a useful tool for environmental decision making in the system design stage.

Figure 1: Framework of the proposed methodology
2 METHODOLOGY

The motivation for this research is to provide a systematic approach that assists engineers to make environmentally-friendly decisions in the system design stage. There are several pieces of functional elements in our method, as shown in Figure 1. A mechanism is first proposed to generate various options of the product architecture that satisfy product functional specifications. The next step is to estimate a quantitative measure for each design with a given environmental metric. A computation procedure is then applied to guide the variation mechanism based on the measure to produce optimal results. Several data sources are required to support the computation process. A component database offers the information of all part models that can be used in the product architecture variation. An ecological database enables product lifecycle assessment and the supply chain design depends on the supplier information.

2.1 Variation of Product Architecture

Product architecture links physical components or modules with functional specifications. To automatically generate all feasible options of product architecture is a challenging task. We assume that the components to be used are given. It is assumed that the relationships between physical chunks and product functions have been known. A finished product model contains the information of function-feature mapping specified by the designer. Such relationships need to be specified for all of the components that will be used for BOM variation. Many factors can be varied to generate different options for product architecture. The variation result is described as a manufacturing BOM with an assembly plan that specifies the assembly sequence and the assembly method for each step. Material type is perhaps the simplest factor to be varied. Some components have several variants produced from the past design process and each variant is denoted by a revision. These variants can be used to produce different designs of product architecture while all satisfying the product functions.

A second category of factors is related to product assembly. There are different assembly methods in practice. An assembly operation is accomplished with certain geometry or group of geometric elements in each method. It is manifested with rules like co-axis, co-plane, extended angle, and offset distance between geometric elements in a CAD model. In this work, we only consider the assembly operations defined by the co-axis and co-plane conditions [7]. A valid assembly operation needs to satisfy the geometric and physical constraints during the assembly. Components to be assembled must satisfy two conditions: (1) have a compatible assembly interface and (2) they do not interfere with each other when put together in space. The assembly interface is defined as the set of assembly features used in an assembly operation, which involves one or multiple assembly features in a component.

A component may contain multiple assembly interfaces when it can be assembled in different ways. The following procedure is applied to determine if two given components satisfy the assembly conditions:

I. **Match assembly interface:** this step identifies all of the assembly interface pairs of the same number and types of assembly features. For each assembly interface, components are automatically assembled in SolidWorks™ to check if the corresponding assembly features are aligned correctly.

II. **Check interference:** for each interface pair passing the first step, we conduct an interface check on the assembled result.

III. **Update assembly information:** the assembly interface that has been used in an assembly operation cannot be utilized any more. These features will be set as inactive.

**Assembly sequence**

A feasible assembly sequence satisfies both geometric and physical constraints in product assembly, referred to as the precedence relationship [9]. The approach of Disassembly Precedence Matrix (DPM) is used to characterize the spatial relationships among the components to be assembled. In addition, we only consider the component disassembly from the six nominal directions (±X, ±Y, and ±Z) [7]. Each component is moved along a given direction until reaching out of the bounding box of the assembly. The components interfering with the one moving along the path are identified and denoted in the DPM.

Variation of assembly sequence is achieved by two steps. The first step generates different assembly hierarchies. An example consisting of five components is shown in Figure 2. For simplification purposes, we assume that an assembly process only involves two components. There are three different topologies under this assumption. The second step is to arrange the components in each assembly hierarchy. There are n! possible sequences of arrangement for n components.

**Component Merge**

To merge components is an effective approach to varying product architecture. Here we propose a simplified and semi-automatic method that assists designers to combine components in an assembly and thus leverage this flexibility. The components to be united should satisfy several conditions. First, the material of the components has to be identical. Besides, they need to be adjacent with each other in the assembly. For simplification purposes, we only consider the components with compatible assembly interfaces. Some component features may vanish after the merge and the merged components still need to provide the original design functions.

![Figure 2: Assembly hierarchies of five components](image-url)
This condition can be validated based on the function-feature mapping relationship. Lastly, the merged result has a higher geometric complexity, thus decreasing its manufacturability.

The variation procedure of product architecture is shown in Figure 3. The designer is required to provide product design functions, an instance of valid product architecture, and the mapping between features and product functions. The mapping has been manifested during the construction process of component models in SolidWorks™ prior to the procedure. A design of product architecture can result in different BOM’s by varying related factors in manufacturing and the supply chain. The later section of optimization framework will explain how to incorporate these factors into the BOM generation process.

LIFE CYCLE ASSESSMENT

In this work, a cradle-to-gate approach is employed to evaluate the environmental impact. It is an assessment of a partial product life cycle. We only consider the impacts of a product in three phases of its lifecycle: preparation of raw materials, manufacturing, and distribution. In addition, the following assumptions are made for each phase. The component suppliers can acquire raw materials from a nearby area and the impact induced by the transportation of raw materials is thus ignored. We only consider the assembly and manufacturing processes of a component. The component material is limited in metals and plastics. They correspond to three manufacturing processes: metal processing, plastic deformation, and injection molding. The possible assembly processes include interference fit, bolt/screw hole, welding, and adhesive bonding. The logistics among the component suppliers and the final assembly plant is our major concern. The transportation of the finished product from the assembly plant to the end customer is not taken into account.

In this work, we adopt DoItProm™ [9] for evaluating the environmental impact of a product during its development process. DoItProm™ was originally developed in 2000 under the support of Ministry of Economic Affairs (MOEA) and the Industrial Technology Research Institute (ITRI) in Taiwan to address the LCA needs of the country. The basic data inventory consists of inputs such as materials, fuel, electrical energy, water, and outputs such as gaseous emissions, water emissions, and solid waste. The major discrepancy between ecological assessment tools and CAD systems is that neither materials nor components result in environmental damage, but only lifecycle activities, like production and transportation. To overcome this problem, we propose the concept of “manufacturing complexity” that links design features and environmental assessment.

OPTIMIZATION FRAMEWORK

An optimization framework has been developed to integrate the variation mechanisms described previously and to generate optimal BOM’s. Different optimization algorithms are used in the framework to guide those variation mechanisms, as each has different characteristics, problem complexity, and the variation process. For simplification purposes, the following assumptions hold in the optimization. First, the manufacturing and assembly capabilities are known for each supplier. Their individual production capacity is infinite. Second, the outsourcing order of the component manufacturing and assembly cannot be split among multiple suppliers. The batch size in transportation is not considered.

4 OPTIMIZATION FRAMEWORK

A GA-based optimization scheme is adopted for determining the optimal product architecture. Product architecture is represented as a chromosome as shown in figure 4. The fitness value of a chromosome is calculated to determine the probability of its survival during the evolution process. The fitness value consists of four parts corresponding to each major activity in the product development. The probability of the chromosome chosen by the next generation is the inverse of the fitness function. Crossover and mutation operators are used to produce a second generation population in the evolution process. Performing of these operations is determined by the probability values Pc and Pm. A pair of parent chromosomes is selected among the population by the roulette wheel selection mechanism.

Figure 4: Encoding of chromosome in the GA algorithm

4.1 Genetic Algorithm

A GA-based optimization scheme is adopted for determining the optimal product architecture. Product architecture is represented as a chromosome as shown in figure 4. The fitness value of a chromosome is calculated to determine the probability of its survival during the evolution process. The fitness value consists of four parts corresponding to each major activity in the product development. The probability of the chromosome chosen by the next generation is the inverse of the fitness function. Crossover and mutation operators are used to produce a second generation population in the evolution process. Performing of these operations is determined by the probability values Pc and Pm. A pair of parent chromosomes is selected among the population by the roulette wheel selection mechanism.
4.2 Tabu Search
A Tabu search is applied to determine an optimal assembly sequence in a given assembly hierarchy, which falls into the problem category. A local search procedure is used to iteratively move from the current solution to a new one until reaching a stop criterion. The neighborhood of each solution is modified as the search progresses. The Tabu list is the most critical parameter controlling the search process. It includes the solutions that have been visited in the previous iterations. Tabu Tenure determines the number of previous solutions to be stored. We need to convert the problem (in this case the assembly sequence) into a formulation that would be efficiently processed by a Tabu search. There are n components to be arranged in n slots of an assembly hierarchy, as shown in Figure 8. An assembly sequence π where π(i) ∈ {p_1, p_2, … p_n} for i = 1, 2, … n. π is a sequence of component indices. A feasible solution has to satisfy the assembly constraints specified in the DPM. A neighboring solution is generated by a move from the current solution in the search space. In our approach, the move is defined as switching two component indices randomly selected.

4.3 Dynamic Programming
The problem here is to arrange the manufacturing tasks of a given assembly sequence among a group of suppliers subject to individual manufacturing constraints. We assumed that the final product is assembled from n components. n_1 of them are raw parts and n_2 = n-n_1 of them are subassembly. A_i is the set that contains the raw parts and subassembly of component i and M_i is the set that contains suppliers who are capable of producing component i. c_i is the environmental impact generated during the manufacturing process when component i is produced by supplier j ∈ M_i, d_{ij} denotes the environmental impact generated when component i is transported from supplier j to supplier j′. At each stage of dynamic programming, we evaluate suppliers for one component starting from the first to the nth component. The first component represents the final product. Let f_i(S_i, X_i) = f_i(S_i, (X_{i+1})) be the total environmental impacts for the remaining stages given at stage i component i is produced by supplier S_i. Component k, k ∈ A_i, is produced by the supplier X_k. Given S_i, X_i denotes the value of X_i that minimizes f_i(S_i, X_i) and f^*_i(S_i) is the corresponding minimum value of f_i(S_i, X_i).

\[
f_i(S_i, X_i) = \text{immediate environmental impact (stage i) + minimum future environmental impacts (stage i+1 onward)}
\]

The immediate environmental impact includes the environmental impact generated during the manufacturing process when component i is produced by supplier S_i or C_i and the environmental impact induced by moving raw parts and subassemblies of component i from the locations of their suppliers to the location of S_i. Thus,

\[
f_i(S_i, X_i) - C_iS_i + \sum_{k \in A_i} d_k(X_k, S_i) + \sum_{k \in A_i} f^*_k(S_k)
\]

If component i is a raw part, A_i is empty. Then

\[
f_i(S_i) = C_iS_i
\]

The objective is to find the minimum environmental impact of the final product, i.e.,

\[
\min_{S_i} f_i(S_i)
\]

Dynamic programming finds the solution by successively searching f^*_i(S_i) for all i as:

\[
f_i(S_i) = C_iS_i + \min_{X_i \in A_i} \{ \sum_{k \in A_i} d_k(X_k, S_i) + \sum_{k \in A_i} f^*_k(S_k) \} \quad (5)
\]

Recursion is solved by backward induction.

5 IMPLEMENTATION AND TEST RESULT
This work integrates product design and supply chain design for reducing the environmental impact of product development. CAD-related functions are required to ensure that the design solution produced in the search process satisfies the product functional requirements and that the assembly plan is feasible. The proposed methodologies have been implemented using C++ in MS Visual Studio™ 2005. SolidWorks™ 2008 offers a complete set of CAD library in C++ and individual functions can be accessed via DLL calls by external programs. It also contains a simple product data management module, which works as the component database. The ecological data related to the impact assessment is output from the systems into a text file.

Proprietary C++ programs have been developed for the optimization schemes based on GA, Tabu search, and dynamic programming.

Figure 5: A test example of bicycle design

An example of bicycle design (see Figure 5) is used to verify the proposed framework. The available materials and variants of individual components are shown in Table 1. Suppose that the bicycle company is developing a new product with a group of suppliers. Some suppliers can only produce certain components due to limitations of their manufacturing capability, as shown in Table 2. Some of them can provide both manufacturing and assembly services. These suppliers are located in different cities in Taiwan. The transportation distances among them are shown in Table 3. The assembly capabilities are listed in Table 4. The LCA data related to CO_2 emission is shown in Table 5. The parameter settings in the GA and Tabu search are shown in Table 6. The search terminates in both algorithms when the maximal number of iterations is reached. The Tabu length is chosen as 7 in the implementation.
<table>
<thead>
<tr>
<th>Part #</th>
<th>Component name</th>
<th>Material</th>
<th># of Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Grip</td>
<td>PP, PE</td>
<td>2</td>
</tr>
<tr>
<td>P2</td>
<td>Pedal</td>
<td>PE, PVC</td>
<td>2</td>
</tr>
<tr>
<td>P3</td>
<td>Connection tube</td>
<td>6061 Al, low carbon steel</td>
<td>1</td>
</tr>
<tr>
<td>P4</td>
<td>Back frame</td>
<td>6061 Al, gray cast iron</td>
<td>2</td>
</tr>
<tr>
<td>P5</td>
<td>Back frame connector</td>
<td>6061 Al, low carbon steel</td>
<td>1</td>
</tr>
<tr>
<td>P6</td>
<td>Pedal rod</td>
<td>6061 Al, low carbon steel</td>
<td>2</td>
</tr>
<tr>
<td>P7</td>
<td>Handle bar</td>
<td>6061 Al, low carbon steel</td>
<td>2</td>
</tr>
<tr>
<td>P8</td>
<td>Main frame</td>
<td>6061 Al, low carbon steel</td>
<td>3</td>
</tr>
<tr>
<td>P9</td>
<td>Rim</td>
<td>6061 Al, low carbon steel</td>
<td>2</td>
</tr>
<tr>
<td>P10</td>
<td>Seat tube</td>
<td>6061 Al, low carbon steel</td>
<td>1</td>
</tr>
<tr>
<td>P11</td>
<td>Rear sprocket</td>
<td>6061 Al, gray cast iron</td>
<td>1</td>
</tr>
<tr>
<td>P12</td>
<td>Free wheel</td>
<td>6061 Al, gray cast iron</td>
<td>1</td>
</tr>
<tr>
<td>P13</td>
<td>Saddle connector</td>
<td>6061 Al, gray cast iron</td>
<td>1</td>
</tr>
<tr>
<td>P14</td>
<td>Front connector</td>
<td>6061 Al, gray cast iron</td>
<td>2</td>
</tr>
<tr>
<td>P15</td>
<td>Suspension</td>
<td>6061 Al, gray cast iron</td>
<td>1</td>
</tr>
<tr>
<td>P16</td>
<td>Suspension connector</td>
<td>6061 Al, gray cast iron</td>
<td>1</td>
</tr>
<tr>
<td>P17</td>
<td>Front wheel shaft</td>
<td>6061 Al, gray cast iron</td>
<td>1</td>
</tr>
<tr>
<td>P18</td>
<td>Front fork</td>
<td>6061 Al, low carbon steel</td>
<td>2</td>
</tr>
<tr>
<td>P19</td>
<td>Saddle</td>
<td>PUR flexible foam</td>
<td>1</td>
</tr>
<tr>
<td>P20</td>
<td>Tire</td>
<td>synthetic rubber</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Component materials and variants in the bicycle design

<table>
<thead>
<tr>
<th>Component name</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grip</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedal</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connection tube</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back frame</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back frame connector</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedal rod</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handle bar</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main frame</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rim</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>Seat tube</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear sprocket</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free wheel</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saddle connector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front connector</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspension</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspension connector</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front wheel shaft</td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front fork</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>Saddle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>Tire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>V</td>
</tr>
</tbody>
</table>

Table 2: Manufacturing capability of suppliers

<table>
<thead>
<tr>
<th>distance (km)</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1(M3)</td>
<td>440</td>
<td>140</td>
<td>0</td>
<td>480</td>
<td>300</td>
<td>490</td>
<td>900</td>
<td>700</td>
</tr>
<tr>
<td>A2(M4)</td>
<td>40</td>
<td>320</td>
<td>480</td>
<td>0</td>
<td>150</td>
<td>20</td>
<td>380</td>
<td>160</td>
</tr>
<tr>
<td>A3(M6)</td>
<td>50</td>
<td>330</td>
<td>490</td>
<td>20</td>
<td>90</td>
<td>0</td>
<td>390</td>
<td>130</td>
</tr>
<tr>
<td>A4(M7)</td>
<td>600</td>
<td>750</td>
<td>900</td>
<td>380</td>
<td>600</td>
<td>390</td>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>A5</td>
<td>50</td>
<td>300</td>
<td>500</td>
<td>10</td>
<td>150</td>
<td>10</td>
<td>400</td>
<td>150</td>
</tr>
</tbody>
</table>

Table 3: Geographic locations of individual suppliers and their distances
<table>
<thead>
<tr>
<th>Grip</th>
<th>Pedal</th>
<th>Connection tube</th>
<th>Back frame</th>
<th>Pedal rod</th>
<th>Handle bar</th>
<th>Main frame</th>
<th>Rim</th>
<th>Seat tube</th>
<th>Rear sprocket</th>
<th>Free wheel</th>
<th>Saddle Connector</th>
<th>Front connector</th>
<th>Suspension connector</th>
<th>Front wheel shaft</th>
<th>Front fork</th>
<th>Saddle</th>
<th>Tire</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedal</td>
<td></td>
<td>A3,A4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tube</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back frame</td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td>A4,A5</td>
<td>A4,A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back frame</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A4,A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>connector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A3,A4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedal rod</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handle bar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main frame</td>
<td></td>
<td></td>
<td>A5</td>
<td>A4,A5</td>
<td>A5</td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td>A4,A5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rim</td>
<td></td>
<td></td>
<td>A5</td>
<td>A4,A5</td>
<td>A5</td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td>A4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seat tube</td>
<td></td>
<td></td>
<td>A4</td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rear sprocket</td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free wheel</td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saddle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front</td>
<td></td>
<td></td>
<td>A3,A5</td>
<td></td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>connector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A3,A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspension</td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>connector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front wheel</td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shaft</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front fork</td>
<td></td>
<td></td>
<td>A5</td>
<td>A4,A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saddle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4 Assembly capability of suppliers
### Table 5: LCA data required by the assessment of environmental impacts

<table>
<thead>
<tr>
<th>Material</th>
<th>Unit</th>
<th>CO2 (kg)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>kg</td>
<td>1.9</td>
<td>0.89 DoItProTM</td>
</tr>
<tr>
<td>PE</td>
<td>kg</td>
<td>1.8</td>
<td>0.917~0.952    DoItProTM</td>
</tr>
<tr>
<td>PVC</td>
<td>kg</td>
<td>2.7</td>
<td>1.29~1.30 DoItProTM</td>
</tr>
<tr>
<td>gray cast iron</td>
<td>kg</td>
<td>1.5</td>
<td>7.30 DoItProTM</td>
</tr>
<tr>
<td>6061 Al</td>
<td>kg</td>
<td>4.5678</td>
<td>2.70 SimaProTM</td>
</tr>
<tr>
<td>low carbon steel</td>
<td>kg</td>
<td>0.653</td>
<td>7.80 SimaProTM</td>
</tr>
<tr>
<td>synthetic rubber</td>
<td>kg</td>
<td>4.3</td>
<td>1.00 DoItProTM</td>
</tr>
<tr>
<td>PUR flexible foam</td>
<td>kg</td>
<td>4.2</td>
<td>0.05 DoItProTM</td>
</tr>
</tbody>
</table>

### Table 5: LCA data required by the assessment of environmental impacts

<table>
<thead>
<tr>
<th>Material</th>
<th>Unit</th>
<th>CO2 (kg)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVC – Injection molding</td>
<td>kg</td>
<td>0.28</td>
<td>DoItProTM</td>
</tr>
<tr>
<td>PP, PE – Injection molding</td>
<td>kg</td>
<td>2.657</td>
<td>DoItProTM</td>
</tr>
<tr>
<td>Machining</td>
<td>kg</td>
<td>1.1158</td>
<td>DoItProTM</td>
</tr>
<tr>
<td>Sheet metal working</td>
<td>kg</td>
<td>1.3938</td>
<td>DoItProTM</td>
</tr>
</tbody>
</table>

### Table 6: Parameter setting in the optimization schemes

<table>
<thead>
<tr>
<th>Scheme</th>
<th>crossover rate</th>
<th>mutation rate</th>
<th>population size</th>
<th>number of iterations</th>
<th>Tabu length</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>0.5</td>
<td>0.03</td>
<td>40</td>
<td>75</td>
<td>7</td>
</tr>
<tr>
<td>TS</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Figure 6: Search process of optimal manufacturing BOM with reduced CO2 emissions

### Table 7: Breakdown of CO2 emission for the optimal manufacturing BOM

<table>
<thead>
<tr>
<th></th>
<th>raw material preparation</th>
<th>manufacturing</th>
<th>assembly</th>
<th>supply chain</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2 emission (kg)</td>
<td>26.28</td>
<td>2.32</td>
<td>0.018</td>
<td>5.55</td>
<td>34.17</td>
</tr>
<tr>
<td>CO2 emission (percentage)</td>
<td>76.9%</td>
<td>6.8%</td>
<td>0.05%</td>
<td>16.2%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 8: Optimal product design and supplier selection

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
</tr>
</thead>
<tbody>
<tr>
<td>variant</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>material</td>
<td>PE</td>
<td>PVC</td>
<td>steel</td>
<td>Cl</td>
<td>steel</td>
<td>steel</td>
<td>steel</td>
<td>steel</td>
<td>Al</td>
</tr>
<tr>
<td>supplier</td>
<td>M1</td>
<td>M2</td>
<td>M6</td>
<td>M8</td>
<td>M5</td>
<td>M6</td>
<td>M8</td>
<td>M8</td>
<td>M6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
<th>P16</th>
<th>P17</th>
<th>P18</th>
<th>P19</th>
<th>P20</th>
</tr>
</thead>
<tbody>
<tr>
<td>variant</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>material</td>
<td>Al</td>
<td>Cl</td>
<td>Cl</td>
<td>Cl</td>
<td>Cl</td>
<td>Cl</td>
<td>Al</td>
<td>foam</td>
<td>rubber</td>
</tr>
<tr>
<td>supplier</td>
<td>M4</td>
<td>M5</td>
<td>M3</td>
<td>M5</td>
<td>M5</td>
<td>M5</td>
<td>M5</td>
<td>M8</td>
<td>M2</td>
</tr>
</tbody>
</table>

Table 8: Optimal product design and supplier selection
The search process of optimal manufacturing BOM in terms of reduced CO\textsubscript{2} emission is shown in Figure 6. The improvement becomes less significant after 60 iterations in the GA algorithm. The best solution after 75 iterations generates 34.17-kg CO\textsubscript{2} emission. As shown in Table 7, the breakdown of the result indicates that raw material preparation dominates the environmental impact of product development. This conclusion is similar to most LCA practices and previous analysis. The logistics in the supply chain also consumes a larger portion of CO\textsubscript{2} and outweighs the manufacturing or assembly activities. The corresponding results of product design and manufacturing supplier selection are summarized in Table 8. The optimal assembly sequence is shown in Figure 7.

![Figure 7 Optimal assembly sequence](image)

6 CONCLUSION

Sustainable product development becomes an imperative for modern manufacturers and consumers all over the world. Product design is considered the critical stage involving decisions that concern environmental impact most. Most design for environment methodologies only facilitate decision making in the detail design stage. The supply chain activities need to be incorporated early in the design process to provide the greatest improvement in sustainability. This paper presented an integrated framework for product designers to make environmentally friendly decisions in consideration of the product design, manufacturing, and the supply chain simultaneously. It incorporates a number of factors into the system design stage that ecologically influence the product development activities. These factors, including component selection, assembly sequencing, assembly method, component merge, and supplier selection, allow automatic variation of manufacturing BOMs. The variation result is guaranteed to provide all product functions and to be interference-free during assembly. LCA was conducted to estimate the amount of CO\textsubscript{2} emissions for raw material production, part assembly/manufacturing, and the supply chain. Three optimization schemes were applied to search for better BOMs with minimized CO\textsubscript{2} emissions. An example of bicycle design was tested to demonstrate the capability of the proposed framework. The test results show that the system design stage offers a feasible means to significantly improve the environmental impact of product development. This research can be improved by incorporating probability-based optimization to account for the uncertainty in estimation of environmental impact.

7 REFERENCES


Identification of the Optimal Design Based on Evaluation of Product Configurations Considering Specific Product Adaptabilities

J. Man¹, L. Zhang¹, Y. Chen¹, D. Xue², P. Gu²,³

¹School of Mechanical Engineering, Tianjin University, Tianjin, China
²Department of Mechanical and Manufacturing Engineering, University of Calgary, Calgary, Canada
³Department of Mechatronics Engineering, Shantou University, Shantou, China

manjiajz982@qq.com, zhanglh@tju.edu.cn, chenyongyangtju.edu.cn, dixue@ucalgary.ca, pgu@ucalgary.ca

Abstract
Adaptable design is a new design approach to create products that can be easily adapted to satisfy the changed functional requirements. Adaptable design approach can reduce the environment impact as well as to improve product competitiveness by replacing multiple products with a single adaptable product. In this research, a method to evaluate adaptabilities of mechanical products based on configurations of these products is introduced. In this method, different functional requirements are described by adaptation tasks. Each adaptation task is achieved by a product whose configuration is modeled by a number of modules based on modular design approach. Adaptation from one product to another one is conducted by changing modules (i.e., adding new modules and removing existing modules) of this product. Adaptabilities of a product to change to other products considering the adaptation tasks are obtained based on the required costs to change from configuration of this product to the configurations of the adapted products. An industrial application is provided to show the effectiveness of the introduced method.

Keywords:
Adaptable Design, Product Platform, Gear Cutting Machine Tool

1 INTRODUCTION
Adaptable design is a new design paradigm with both economical and environmental benefits [1]. The fundamental principle of adaptable design is the ability of a design or a product to be adapted to a new one based on the changed requirements by reusing some of the components in the existing design or product in the new one. Adaptable design approaches can reduce the costs of customers by replacing multiple products with a single adaptable product as well as to reduce the effort of manufacturers by reusing the knowledge in the existing design. Adaptable design approaches can also be used to improve the environment by reducing the number of totally manufactured products.

Adaptability is defined as the ability for a design or product to be adapted to other ones. Two types of adaptabilities are considered in adaptable design: design adaptability and product adaptability [1]. Design adaptability is the capability of an existing design to be adapted to create a new or modified design based on the changed requirements. Product adaptability is the capability of a physical product to be adapted to satisfy the changed requirements. Since design adaptability can be achieved by developing knowledge-based systems, this research focuses on adaptable design considering product adaptability.

Since the introduction of the adaptable design concept, many design methods and processes have also been developed to identify the adaptable products [2]. Some traditional design methodologies can also be used to design the adaptable products. For example, the reconfigurable design method can be used to create a reconfigurable machine to achieve functions of several machines by reconfiguration of the components of this machine [3]. The modular, product platform, and product family/portfolio design methods can be used to improve the structures of products, so these products can be changed easily to achieve different functions [4]. Various kinds of multi-functional machine tools have been developed and used to satisfy the specific demands of customers [5]. However, a systematic methodology to design configurations of the machines considering adaptability has not been introduced. Due to the increasing demands on multi-functional machine tools, more effort must be devoted to develop a scientific methodology for the design of these multi-functional machine tools with high efficiency and quality.

Because many design candidates can be achieved in adaptable design, evaluation of these candidates considering their product adaptabilities needs to be conducted to identify the best design from all these candidates. In adaptable design, product adaptabilities are classified into specific product adaptabilities and general product adaptabilities, depending on whether planned information for specific adaptations is available [1]. When certain adaptabilities and their probabilities can be predicted, the product can be designed to accommodate these specific product adaptabilities. For accommodating some unpredictable requirements and changes, the product can be designed to have some general product adaptabilities by its product architecture and interfaces.

For achieving specific product adaptabilities, Gu et al. developed a method to measure specific product adaptability by comparing the relative efforts of product adaptation and new product creation [1]. Li et al. extended this specific product adaptability evaluation method by considering three types of product adaptation tasks: extendibility of functions, upgradeability of modules and customizability of components [6]. Fletcher et al. developed a method to quantify general product adaptability by comparing the actual product structure with its ideal product structure that can be easily changed [7]. In addition, modularity, commonality [8], customizability...
[9], etc. used in modular, product platform, and product family/portfolio design methods can also be employed to evaluate product adaptabilities.

The research presented in this paper focuses on the development of an evaluation method to measure specific product adaptabilities considering the contributions of the modules to different configurations of the required adaptable products. A design process is introduced to create a family of products where each product in this family can be easily adapted to other products in the same family. The developed specific product adaptability evaluation method has also been employed in the design of a heavy-duty gear cutting machine.

2 A METHOD TO EVALUATE SPECIFIC PRODUCT ADAPTABILITIES

2.1 Specific Product Adaptability

Gu et al. [1] developed a method to measure specific product adaptability by comparing the relative efforts of product adaptation with new product creation. Suppose \( Tp_i \) is the \( i \)-th adaptation task, the effort for this task, according to the information axiom in axiomatic design theory [10], can be modeled by its information content described by \( \text{Inf}(Tp_i) \). Cost is usually used for modeling the effort. When \( S_1 \) is the current state of the existing product, \( A S_2 \) is the state after adaptation; the effort for this adaptation is then described by \( \text{Inf}(S_1 \rightarrow AS_2) \). In the same way, the effort to develop a new product from scratch is described by \( \text{Inf}(\text{ZERO} \rightarrow AS_2) \), where \( \text{ZERO} \) is the state to design a new product from scratch, and \( IS_2 \) is the state of product to satisfy only the new requirements. The relative saving of effort is modeled as adaptable factor, \( AF(Tp_i) \).

\[
AF(Tp_i) = \begin{cases} 
1 - \frac{\text{Inf}(S_1 \rightarrow AS_2)}{\text{Inf}(\text{ZERO} \rightarrow IS_2)} & \text{if } \text{Inf}(\text{ZERO} \rightarrow IS_2) > \text{Inf}(S_1 \rightarrow AS_2) \\
0 & \text{if } \text{Inf}(\text{ZERO} \rightarrow IS_2) \leq \text{Inf}(S_1 \rightarrow AS_2)
\end{cases}
\]

(1)

When less effort is required to adapt a product than to develop a new one, the \( AF(Tp_i) \) is described by a measure between 0 and 1. When it takes more effort to adapt a product than to develop a new product (i.e., \( \text{Inf}(S_1 \rightarrow AS_2) \geq \text{Inf}(\text{ZERO} \rightarrow IS_2) \)), product adaptation should not be considered (\( AF(Tp_i) = 0 \)). When no additional effort is required for product adaptation (i.e., \( \text{Inf}(S_1 \rightarrow AS_2) = 0 \)), the product is a perfect adaptable product (\( AF(Tp_i) = 1 \)).

When \( n \) product adaptation tasks, \( Tp_i (i = 1, 2, \ldots, n) \), and their probabilities, \( Pr(Tp_i) \), are considered, the specific product adaptability is then modeled by:

\[
A(Tp) = \sum_{i=1}^{n} [Pr(Tp_i)AF(Tp_i)]
\]

(2)

2.2 Modeling of Configurations of Adaptable Products

Modular design is a popular method for obtaining adaptable products. In modular design, similar components are grouped into modules according to their functions, technologies or physical structures [4]. Since the modules are relatively independent, these modules can be disassembled non-destructively from the product as units. For an adaptable product created using modular design, the modules can be attached, detached, modified, relocated and replaced easily for achieving the changed design functional requirements.

Suppose for an existing product, \( P \), \( n \) adaptation tasks are described by \( Tp = \{Tp_1, Tp_2, \ldots, Tp_n\} \), and these adaptable tasks are achieved by products \( AP_j (j = 1, 2, \ldots, n) \) through adding new modules to the existing product, removing modules from the existing product, and replacing the modules of the existing product by new modules. The probabilities of these adaptation tasks are described by \( Pr = \{Pr(Tp_1), Pr(Tp_2), \ldots, Pr(Tp_n)\} = \{Pr_{T1}, Pr_{T2}, \ldots, Pr_{Tn}\} \).

Since each product created in adaptable design is composed of modules, product adaptability can then be achieved by sharing modules among the products. Suppose each of the \( n \) products \( AP_j (j = 1, 2, \ldots, n) \) is modeled by a configuration with a number of modules, and a total of \( m \) modules, \( M = \{M_1, M_2, \ldots, M_m\} \), are used to describe the \( n \) products \( AP_j (j = 1, 2, \ldots, n) \), configurations of these \( n \) products \( AP_j (j = 1, 2, \ldots, n) \) can then be modeled by the \( m \) modules \( M_j (j = 1, 2, \ldots, m) \) by the following matrix.

\[
A = \begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
\]

(3)

where each element \( a_{ij} = \{1, 2, \ldots, m; j = 1, 2, \ldots, n\} \) is described by a Boolean value of either 1 or 0.

• \( a_{ij} = 1 \): product \( AP_j \) is composed of the module \( M_i \).

• \( a_{ij} = 0 \): product \( AP_j \) is not composed of the module \( M_i \).

Costs of these \( m \) modules are modeled by:

\[
C = \begin{bmatrix}
  c_1 \\
  c_2 \\
  \vdots \\
  c_m
\end{bmatrix}
\]

(5)

2.3 Calculation of Specific Product Adaptabilities

Suppose \( IP_j (j = 1, 2, \ldots, n) \) represents the ideal product to achieve the functions specified in \( Tp_0 \). Based on equation (1), adaptable factor considering the change from the product \( P \) to the product \( AP_j (j = 1, 2, \ldots, n) \) can be modeled by:

\[
AF(Tp_j) = \begin{cases} 
1 - \frac{\text{Inf}(P \rightarrow AP_j)}{\text{Inf}(\text{ZERO} \rightarrow IP_j)} & \text{if } \text{Inf}(\text{ZERO} \rightarrow IP_j) > \text{Inf}(P \rightarrow AP_j) \\
0 & \text{if } \text{Inf}(\text{ZERO} \rightarrow IP_j) \leq \text{Inf}(P \rightarrow AP_j)
\end{cases}
\]

(6)

In equation (6), \( \text{Inf}(\text{ZERO} \rightarrow IP_j) (j = 1, 2, \ldots, n) \) describes the effort to create the ideal product \( IP_j \) from scratch to satisfy the functions given in the adaptation task \( Tp_0 \), while \( \text{Inf}(P \rightarrow AP_j) (j = 1, 2, \ldots, n) \) describes the effort to change from the existing product \( P \) to the product \( AP_j \) considering the adaptation task \( Tp_0 \).

In this work, cost is employed to model the effort given in equation (6). For example, the effort to produce product \( AP_j \) from scratch is obtained by adding all the costs of the individual modules of the product \( AP_j \) using:

\[
\text{Inf}(AP_j) = \sum_{i=1}^{m} c_i a_{ij}
\]

(7)

Since many modules are used in this research to build the products by sharing these modules in different configurations, the cost to create the ideal product \( IP_j \) is approximated by the cost to create the product \( AP_j \) from scratch to achieve the same functional requirements. Therefore the \( \text{Inf}(\text{ZERO} \rightarrow IP_j) \) in equation (6) is calculated by:
Inf\{(ZERO \rightarrow AP)\} \approx \inf\{(ZERO \rightarrow AP)\} \sum_{i=1}^{m} c_{i}a_{ij} \quad (8)

Inf\{(P \rightarrow AP)\} in equation (6) is obtained by adding the costs of the modules which are added to \( P \) to form the \( AP_j \) using:

\[\inf\{P \rightarrow AP\} = \inf\{AP\} - \inf\{P \cap AP\}\]

\[= \sum_{i=1}^{m} c_{i}a_{ij} - \sum_{i=1}^{m} q_{i}b_{i}a_{ij} \quad (9)\]

\[\sum_{i=1}^{m} (c_{i}(1-b_{i})k_{ij})\]

By replacing the \( \inf\{(ZERO \rightarrow P)\} \) and \( \inf\{P \rightarrow AP\} \) in equation (6) using the results calculated by equations (8) and (9), the adaptable factor considering the change from the product \( P \) to the product \( AP_j \) \((j=1,2,...,n)\) can be calculated by:

\[A F(T_{Pj}) = 1 - \frac{\inf\{(P \rightarrow AP)\}}{\inf\{(ZERO \rightarrow P)\}} - \frac{\sum_{i=1}^{m} c_{i}(1-b_{i})k_{ij}}{\sum_{i=1}^{m} c_{i}a_{ij}} \quad (10)\]

When \( n \) product adaptation tasks, \( T_{Pj} (j=1,2,...,n) \), and their probabilities, \( P_{Tj} (j=1,2,...,n) \), are considered, based on equation (2), the specific product adaptability is calculated by:

\[A(P) = \sum_{j=1}^{n} P_{Tj} A F(T_{Pj}) \]

\[= \sum_{j=1}^{n} P_{Tj} \left[ 1 - \frac{\sum_{i=1}^{m} c_{i}(1-b_{i})k_{ij}}{\sum_{i=1}^{m} c_{i}a_{ij}} \right] \quad (11)\]

When many design candidates can be achieved, the optimal design of the product, \( P^* \), with the maximum specific product adaptability can be identified based on evaluations of all design candidates in terms of their specific product adaptabilities.

3 AN ADAPTABLE DESIGN PROCESS BASED ON SPECIFIC PRODUCT ADAPTABILITIES

The major advantage of adaptable design is the use of an adaptable product to replace a number of products which are demanded based on different customer requirements. By employing a modular design approach, a product can be changed to another one by adding new modules to the existing product, removing modules from the existing product, and replacing the modules of the existing product by new modules.

When an adaptable product is designed to achieve multiple functions from different customer requirements, evaluation of each configuration of the product considering the efforts to adapt to other configurations is required. When multiple candidate adaptable products can be created to satisfy the same adaptation tasks, the one with the best average specific product adaptability considering the adaptation from any configuration of this adaptable product to its other configurations is selected as the optimal design.

This adaptable design process is composed of 4 steps as shown in Figure 1.

1. Defining multiple functions as adaptation tasks

\[T_{P} = \{T_{P1}, T_{P2}, ..., T_{Pn}\}\]

2. Modeling the configurations of the adaptable product

\[A_{mn} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{mn} & a_{m2} & \cdots & a_{mn} \end{bmatrix}\]

3. Calculating adaptable factors

\[A_{nm} = \begin{bmatrix} 1 & A_{F12} & \cdots & A_{F1n} \\ A_{F21} & 1 & \cdots & A_{F2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{Fn1} & A_{Fn2} & \cdots & 1 \end{bmatrix}\]

4. Obtain the product adaptabilities

\[A = \{A_1, A_2, ..., A_n\}\]

Figure 1: An adaptable design process.

(1) Defining multiple functions as adaptation tasks.

In this step, the \( n \) different functions of the customer requirements are defined by \( n \) adaptation tasks \( T_{Pj} \) \((j=1,2,\ldots,n)\).

(2) Modeling the configurations of the adaptable product.

A total of \( m \) modules, \( M_i \) \((i=1,2,\ldots,m)\), are used to model the \( n \) configurations of the adaptable product, \( AP_j \) \((j=1,2,\ldots,n)\), for achieving the \( n \) adaptation tasks. Each configuration is composed of modules selected from the \( m \) modules. An \( m \times n \) matrix, \( A_{mn} \), is constructed to show what modules have been used for each of the \( n \) configurations using equation (3).

(3) Calculating adaptable factors.

For any two configurations, \( AP_i \) and \( AP_j \) \((i=1,2,\ldots,n; j=1,2,\ldots,n)\), calculate the adaptable factor, \( A_{F_{ij}} \), based on equation (10) to establish an \( n \) by \( n \) matrix \( A_{F_{nm}} \).

(4) Obtain the product adaptabilities.

Based on equation (11), the \( n \) product adaptabilities, \( A_j \) \((j=1,2,\ldots,n)\), for the \( n \) configurations of this adaptable product are then obtained.

4 AN INDUSTRIAL APPLICATION

Heavy-duty gear cutting machines are mainly used for manufacturing large gears, including spur gears and helical gears with internal and external gear teeth, through the manufacturing processes of milling (roughing), hobbing (semi-finishing) and teeth grinding (finishing). A
traditional gear cutting machine is primarily designed for a specific type of gear machining process. Such a machine cannot be adapted to satisfy the changed functional requirements of a new type of gear process. Therefore the development of an adaptable product using an adaptable design method is necessary. Based on market analysis, product family planning and module planning, the adaptability evaluation method for a variety of products derived from the modular design platform is used in this research to identify the best adaptable product by prioritizing the different design candidates.

Each gear cutting machine in the product family is composed of modules. Since the same modules are shared by different machines in this product family, a machine in this family can be adapted to others in the same product family by adding new modules to the existing machine, removing modules from the existing machine, and replacing the modules of the existing machine by new modules. By evaluating product adaptability of each product derived from the same product family, the product family with the maximum adaptability is selected as the optimal design.

### 4.1 Identification of the Adaptable Tasks

The gear cutting machines considered in this application are designed according to different requirements which are shown in Table 1. These gear cutting machines are classified into two categories considering whether the workpiece or the cutting tool is moved in the manufacturing process. When different processes are used, different tools with different tool holders have to be considered. Therefore the employment of an adaptable design method in the design of the tool turrets in this application is not considered. To support different sizes of workpieces, the table of the machine needs to be adapted to different sizes. Even though these work tables are different in size, the structures of these work tables are similar. Therefore adaptable design can be considered for the design of work tables.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum gear diameter (mm)</td>
<td>1,200 - 3,000mm</td>
</tr>
<tr>
<td>Maximum gear modulus (mm)</td>
<td>16 - 35 mm</td>
</tr>
<tr>
<td>Maximum gear width (mm)</td>
<td>400 - 600 mm</td>
</tr>
<tr>
<td>Maximum table load (Tons)</td>
<td>15 - 40 tons</td>
</tr>
<tr>
<td>Gear type</td>
<td>Internal, external</td>
</tr>
<tr>
<td>Process</td>
<td>Milling, hobbing</td>
</tr>
</tbody>
</table>

Table 1: Variations of Requirements for Gear Cutting Machines

In this application, the considered adaptation activities include two kinds of processes (i.e., milling and hobbing), two types of gears (i.e., gears with internal teeth and gears with external teeth), three maximum sizes of workpieces (i.e., 1,200 mm, 2,000 mm and 3,000 mm), and so on as shown in Table 2. As hobbing can only be used for cutting external teeth, a total of 9 sets of product specifications are created. These specifications are defined as adaptation tasks as shown in Table 2.

### 4.2 Modeling of Product Configurations in Different Product Families

Modular design approach is employed in this work to design the structure of this adaptable machine. First each adaptable task is achieved by a configuration of the adaptable product. When a different adaptable task is required, the existing configuration is then adapted to a different configuration to achieve the new adaptable task.

<table>
<thead>
<tr>
<th>Adaptation task</th>
<th>TP₁</th>
<th>TP₂</th>
<th>TP₃</th>
<th>TP₄</th>
<th>TP₅</th>
<th>TP₆</th>
<th>TP₇</th>
<th>TP₈</th>
<th>TP₉</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gear type</td>
<td>internal</td>
<td>internal</td>
<td>internal</td>
<td>external</td>
<td>external</td>
<td>external</td>
<td>external</td>
<td>external</td>
<td>external</td>
</tr>
<tr>
<td>Process</td>
<td>milling</td>
<td>milling</td>
<td>milling</td>
<td>milling</td>
<td>milling</td>
<td>milling</td>
<td>hobbing</td>
<td>hobbing</td>
<td>hobbing</td>
</tr>
<tr>
<td>Maximum gear diameter (mm)</td>
<td>1,200</td>
<td>2,000</td>
<td>3,000</td>
<td>1,200</td>
<td>2,000</td>
<td>3,000</td>
<td>1,200</td>
<td>2,000</td>
<td>3,000</td>
</tr>
<tr>
<td>Maximum gear module (mm)</td>
<td>16</td>
<td>25</td>
<td>35</td>
<td>16</td>
<td>25</td>
<td>35</td>
<td>16</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>Maximum gear width (mm)</td>
<td>400</td>
<td>500</td>
<td>600</td>
<td>400</td>
<td>500</td>
<td>600</td>
<td>400</td>
<td>500</td>
<td>600</td>
</tr>
<tr>
<td>Maximum table Load (Tons)</td>
<td>15</td>
<td>30</td>
<td>40</td>
<td>15</td>
<td>30</td>
<td>40</td>
<td>15</td>
<td>30</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 2: Adaptation Tasks for Gear Cutting Machines
### Product Category Features

<table>
<thead>
<tr>
<th>Maximum gear diameter (mm)</th>
<th>1,200</th>
<th>2,000</th>
<th>3,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trunk-type milling head for internal teeth</td>
<td>Internal milling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hob frame for external teeth</td>
<td>External milling</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Configurations in Product Family I

<table>
<thead>
<tr>
<th>Tasks/adaptable factors</th>
<th>TP₁</th>
<th>TP₂</th>
<th>TP₃</th>
<th>TP₄</th>
<th>TP₅</th>
<th>TP₆</th>
<th>TP₇</th>
<th>TP₈</th>
<th>TP₉</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP₁</td>
<td>1.00</td>
<td>0.30</td>
<td>0.30</td>
<td>0.90</td>
<td>0.30</td>
<td>0.90</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>TP₂</td>
<td>0.25</td>
<td>1.00</td>
<td>0.25</td>
<td>0.25</td>
<td>0.89</td>
<td>0.25</td>
<td>0.25</td>
<td>0.89</td>
<td>0.25</td>
</tr>
<tr>
<td>TP₃</td>
<td>0.90</td>
<td>0.30</td>
<td>0.30</td>
<td>1.00</td>
<td>0.30</td>
<td>0.94</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>TP₄</td>
<td>0.90</td>
<td>0.30</td>
<td>0.90</td>
<td>0.25</td>
<td>1.00</td>
<td>0.25</td>
<td>0.94</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>TP₅</td>
<td>0.90</td>
<td>0.30</td>
<td>0.30</td>
<td>0.94</td>
<td>0.30</td>
<td>0.30</td>
<td>1.00</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>TP₆</td>
<td>0.25</td>
<td>0.89</td>
<td>0.25</td>
<td>0.25</td>
<td>0.94</td>
<td>0.25</td>
<td>0.25</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>TP₇</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.20</td>
<td>0.95</td>
</tr>
<tr>
<td>TP₈</td>
<td>0.20</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.95</td>
</tr>
<tr>
<td>TP₉</td>
<td>0.20</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.90</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Table 4: Adaptable Factors for Product Family I

### Table 5: Configuration in Product Family II
By analyzing the functions of the 9 adaptable tasks, 3 gear cutting machine families are identified as shown in Tables 3, 5 and 7. In Family I, shown in Table 3, the width of the bed is fixed, and the size of the table can be changed. In Family II, shown in Table 5, since the bed is integrated with the table, sizes of both the table and the bed can be changed. In Family III, as shown in Table 7, the width of the bed is fixed, and the column and the bed can be adapted to different travel lengths.

### 4.3 Calculation of the Adaptable Factors

By using the configuration information provided in Tables 3, 5 and 7, the costs of the modules, and equation (10), adaptable factors from the $i$-th product ($i=1,2,...,9$) to the $j$-th product ($j=1,2,...,9$) considering all 9 products in each product family can be achieved as shown in Table 4, 6, and 8.

### 4.4 Calculation of the Product Adaptabilities

From Tables 4, 6 and 8, the specific product adaptability from the $i$-th product to all other products in the same product family can be achieved using equation (11). The

<table>
<thead>
<tr>
<th>Tasks/adaptable factors</th>
<th>TP_1</th>
<th>TP_2</th>
<th>TP_3</th>
<th>TP_4</th>
<th>TP_5</th>
<th>TP_6</th>
<th>TP_7</th>
<th>TP_8</th>
<th>TP_9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TP_1$</td>
<td>1.00</td>
<td>0.30</td>
<td>0.30</td>
<td>0.88</td>
<td>0.30</td>
<td>0.30</td>
<td>0.88</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>$TP_2$</td>
<td>0.20</td>
<td>1.00</td>
<td>0.20</td>
<td>0.20</td>
<td>0.89</td>
<td>0.20</td>
<td>0.20</td>
<td>0.89</td>
<td>0.20</td>
</tr>
<tr>
<td>$TP_3$</td>
<td>0.12</td>
<td>0.12</td>
<td>1.00</td>
<td>0.12</td>
<td>0.12</td>
<td>1.00</td>
<td>0.12</td>
<td>0.12</td>
<td>1.00</td>
</tr>
<tr>
<td>$TP_4$</td>
<td>0.88</td>
<td>0.31</td>
<td>0.31</td>
<td>1.00</td>
<td>0.31</td>
<td>0.31</td>
<td>0.94</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>$TP_5$</td>
<td>0.21</td>
<td>0.90</td>
<td>0.21</td>
<td>0.21</td>
<td>1.00</td>
<td>0.21</td>
<td>0.21</td>
<td>0.94</td>
<td>0.21</td>
</tr>
<tr>
<td>$TP_6$</td>
<td>0.12</td>
<td>0.12</td>
<td>0.92</td>
<td>0.12</td>
<td>0.12</td>
<td>1.00</td>
<td>0.12</td>
<td>0.12</td>
<td>0.95</td>
</tr>
<tr>
<td>$TP_7$</td>
<td>0.88</td>
<td>0.30</td>
<td>0.30</td>
<td>0.94</td>
<td>0.30</td>
<td>0.30</td>
<td>1.00</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>$TP_8$</td>
<td>0.21</td>
<td>0.89</td>
<td>0.21</td>
<td>0.21</td>
<td>0.94</td>
<td>0.21</td>
<td>0.21</td>
<td>1.00</td>
<td>0.21</td>
</tr>
<tr>
<td>$TP_9$</td>
<td>0.12</td>
<td>0.12</td>
<td>0.92</td>
<td>0.12</td>
<td>0.12</td>
<td>0.95</td>
<td>0.12</td>
<td>0.12</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 6: Adaptable Factors for Product Family II

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Features</th>
<th>Maximum gear diameter (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1,200</td>
</tr>
<tr>
<td>Trunk-type milling head for internal gear</td>
<td>Internal milling</td>
<td></td>
</tr>
<tr>
<td>Hob frame for external gear</td>
<td>External milling</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Configurations in Product Family III

<table>
<thead>
<tr>
<th>Tasks/adaptable factors</th>
<th>TP_1</th>
<th>TP_2</th>
<th>TP_3</th>
<th>TP_4</th>
<th>TP_5</th>
<th>TP_6</th>
<th>TP_7</th>
<th>TP_8</th>
<th>TP_9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$TP_1$</td>
<td>1.00</td>
<td>0.56</td>
<td>0.56</td>
<td>0.90</td>
<td>0.56</td>
<td>0.56</td>
<td>0.90</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>$TP_2$</td>
<td>0.42</td>
<td>1.00</td>
<td>0.42</td>
<td>0.90</td>
<td>0.42</td>
<td>0.90</td>
<td>0.42</td>
<td>0.90</td>
<td>0.42</td>
</tr>
<tr>
<td>$TP_3$</td>
<td>0.34</td>
<td>0.34</td>
<td>1.00</td>
<td>0.34</td>
<td>0.34</td>
<td>0.90</td>
<td>0.34</td>
<td>0.90</td>
<td>0.34</td>
</tr>
<tr>
<td>$TP_4$</td>
<td>0.90</td>
<td>0.56</td>
<td>0.56</td>
<td>1.00</td>
<td>0.56</td>
<td>0.56</td>
<td>0.94</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>$TP_5$</td>
<td>0.43</td>
<td>0.90</td>
<td>0.43</td>
<td>0.43</td>
<td>1.00</td>
<td>0.43</td>
<td>0.43</td>
<td>0.95</td>
<td>0.43</td>
</tr>
<tr>
<td>$TP_6$</td>
<td>0.35</td>
<td>0.35</td>
<td>0.91</td>
<td>0.35</td>
<td>0.35</td>
<td>1.00</td>
<td>0.35</td>
<td>0.35</td>
<td>0.95</td>
</tr>
<tr>
<td>$TP_7$</td>
<td>0.90</td>
<td>0.56</td>
<td>0.56</td>
<td>0.94</td>
<td>0.56</td>
<td>0.56</td>
<td>1.00</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>$TP_8$</td>
<td>0.42</td>
<td>0.90</td>
<td>0.42</td>
<td>0.42</td>
<td>0.94</td>
<td>0.42</td>
<td>0.42</td>
<td>1.00</td>
<td>0.42</td>
</tr>
<tr>
<td>$TP_9$</td>
<td>0.35</td>
<td>0.35</td>
<td>0.91</td>
<td>0.35</td>
<td>0.35</td>
<td>0.95</td>
<td>0.35</td>
<td>0.35</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 8: Adaptable Factors for Product Family III
specific product adaptabilities for all of the products in the three product families are shown in Table 9. In this case study, the probabilities of all 9 of the adaptable tasks are selected as 1s, since all these adaptable tasks have to be satisfied. From this table, Product Family III is selected as the one with the maximum measure of the average product adaptability. The product platform is then selected to build the 9 configurations of the adaptable product in this product family.

<table>
<thead>
<tr>
<th>Tasks/Family</th>
<th>Product family I</th>
<th>Product family II</th>
<th>Product family III</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP₁</td>
<td>4.61</td>
<td>4.57</td>
<td>6.15</td>
</tr>
<tr>
<td>TP₂</td>
<td>4.29</td>
<td>4.00</td>
<td>5.35</td>
</tr>
<tr>
<td>TP₃</td>
<td>3.98</td>
<td>3.54</td>
<td>4.86</td>
</tr>
<tr>
<td>TP₄</td>
<td>4.66</td>
<td>4.65</td>
<td>6.21</td>
</tr>
<tr>
<td>TP₅</td>
<td>4.35</td>
<td>4.07</td>
<td>5.40</td>
</tr>
<tr>
<td>TP₆</td>
<td>4.06</td>
<td>3.59</td>
<td>4.95</td>
</tr>
<tr>
<td>TP₇</td>
<td>4.65</td>
<td>4.65</td>
<td>6.20</td>
</tr>
<tr>
<td>TP₈</td>
<td>4.34</td>
<td>4.07</td>
<td>5.39</td>
</tr>
<tr>
<td>TP₉</td>
<td>4.06</td>
<td>3.59</td>
<td>4.94</td>
</tr>
<tr>
<td>Average</td>
<td>4.33</td>
<td>4.08</td>
<td>5.49</td>
</tr>
</tbody>
</table>

Table 9: Evaluation of Product Adaptability Considering All Three Product Families

5 SUMMARY

A new method to measure specific product adaptabilities is introduced in this work. In this method, different requirements are described by adaptation tasks. By using modular design, each adaptation task can be achieved by a configuration of the adaptable product which is modeled by a number of modules. Adaptation from one configuration to another one is conducted by changing the modules of this product. Adaptabilities of a product to other products considering the adaptation tasks are achieved based on the required costs to change from configuration of this product to the configurations of the adapted products. An industrial application has been provided to show the effectiveness of the introduced method.

The characteristics of this adaptable design approach can be summarized as follows.

1. Modular design is an effective approach to improve product adaptabilities in adaptable design by developing relatively independent modules according to the functions of these modules in the adaptable product.

2. The introduction of a new specific product adaptability measure considering product configuration is effective to evaluate how easily a product can be adapted to other products based on the costs for achieving these adaptation tasks.

3. The industrial application demonstrates that the newly introduced approach can be employed in industrial equipment design to replace multiple products with a single adaptable product to reduce the environment impact as well as to improve product competitiveness.

6 ACKNOWLEDGMENTS

This research has been supported by the Science and Technology Program of Guangdong Province (No: 2006D90301005, No:2007B080401008) through grants to the Shantou Light Industry Equipment Research Institute, Shantou University, China. This project has also been supported by the Ministry of Science and Technology of China through the National High Technology Development Program (863 Program) (No: 2006AA04Z107) and Special Project for Technology Innovation in Tianjin City, Tianjin, China (No: 07FDZDGX01600).

7 REFERENCES


Generating Optimal Disassembly Process Plans from AND/OR Relationships using a Hierarchical Genetic Algorithm

R. Edmunds, M. Kobayashi, M. Higashi
Engineering Design Laboratory, Research Center for Sustainable Mechanical Systems, Toyota Technological Institute, 2-12-1 Hisakata, Tempaku-ku, Nagoya 468-8511, Japan
Email: edmunds@toyota-ti.ac.jp

Abstract
A Hierarchical Genetic Algorithm (HGA) is presented which, not only successfully finds the optimal disassembly sequence - under a set of given criteria - using information derived from AND/OR relationships, but also reduces the problem size. Whilst many authors agree that AND/OR graphs are the most complete representation of Disassembly Process Plans (DPPs), few have generated optimal sequences from AND/OR information due to the rapid increase of solution paths. For complex systems having a natural hierarchical structure often the optimal solution can be missed, HGA overcomes this and has been proven to be faster and more accurate than traditional Genetic Algorithms.

Keywords:
Disassembly, AND/OR Graph, Numerical Optimization, Hierarchical Structure

1 INTRODUCTION
The biggest damage to the environment is done when a product completes its useful life. However, many products are already in existence for which their End-Of-Life (EOL) use was never considered. Disassembly of a system into its component parts or subassemblies is vital to most EOL strategies. This paper uses a Hierarchical Genetic Algorithm (HGA) to find optimal disassembly paths from the space of all feasible Disassembly Process Plans (DPPs) described using AND/OR information.

Development of a disassembly theory is fundamental for practically all EOL policies devised for the disposal of products – in particular, the careful selection and removal of components for recycling and reuse.

With environmental ideologies ingrained in the public consciousness, and with large new deposits of natural resources becoming increasingly rare, there is extra pressure on industry to take back possession of products at the end of their useful life for reclamation of parts and materials, plus safe disposal of waste products. For the companies, this involves extra time, effort and money, and so the process must be done in the most economically and environmentally viable way [1].

Therefore, there are three elements that will be evaluated when assessing an EOL decision [2]: the costs and benefits of each recovery option; the present disposal cost and the possibility that disassembly is required to recover a valuable part.

If disassembly is needed, the arrangement of the components will, most likely, constrain the sequences such that removal of parts in perfect order (e.g. that the most expensive are removed first and the heaviest last) is unlikely [2-3].

As a consequence, it is necessary to represent the relationships between the components. The most common methods are via AND/OR graphs, which show the possible operations applicable to the current assembly, and precedence graphs, which show the geometrical relationships between the components. Both lead to a hierarchical structure.

For problems with an inherent hierarchical arrangement, traditional Genetic Algorithms (GAs) can miss the true optimal solution [4]. Even if the correct optimum is found, this can take increased computational time when compared to systems with vector arrangements. The Hierarchical Genetic Algorithm [4-5] was developed to overcome these issues and has been proven to work better than standard GAs for hierarchical systems [4-8].

Using the permissible operations upon the assembly to find the antecedents for each of its subassemblies, the configuration required for optimization with HGA is constructed. In doing this, for complete disassembly, the problem size is appreciably reduced. Extra constraints upon the removal order are easily added under this formulation.

The possibly of partial disassembly can also be introduced. This increases the dimensions of the hierarchy, but as it is mainly the width that is augmented, HGA still allows rapid and efficient resolution to obtain the optimal disassembly path.

The paper continues by giving a brief overview of disassembly and explains the difference between the two graphical representations. This is followed by an explanation of the HGA routine, which is subsequently used to model two examples from the literature and optimal disassembly paths are generated. Thus it is shown how disassembly problems represented via AND/OR graphs can be solved using HGA and highlights the benefits of doing this. Finally conclusions are drawn.

2 DISASSEMBLY
The addition of disassembly into the recovery process incurs extra costs. Specifically, disassembly contributes to the investment and labour overheads [9].

The cost involved in these operations are then constraints upon the methods employed and minimizing the cost of some or all of them can lead to finding the optimal choice, a preferred sequence that maximizes the reclaimed value.
Optimal disassembly sequence generation – in terms of minimal cost, maximum benefit and the degree of disassembly – is a non-trivial task. It is often done by analysing design characteristics of the assembly [2]:

- Geometrical relationships;
- Characteristics of operations such as tooling or accessibility and how much they overlap;
- Clustering of materials;
- Concurrent operations and the amount of material recovered.

To find the optimal sequence it is usually considered that all of the feasible paths – called Disassembly Process Plans (DPPs) or Disassembly Sequence Plans (DSPs) – must be generated and assessed, leading to a two-stage process [2, 10]. Generating the total sequences often involves some sort of complex searching methodology and a significant amount of research has been put into extracting the DPPs from assembly diagrams and representing these plans in graphical form [11].

Most disassembly problems are variations of the same basic model structure, which is described as a list of possible disassembly operations or subassemblies [11]. From these feasible subassemblies and feasible actions, hierarchical disassembly graphs can be built by either representing the geometric/relational properties (AND/OR graphs), or via the precedence knowledge alone (precedence graphs) [12]. In both cases this information can then be used to find the optimal sequence based on one or more criteria.

Although there is no general procedure for converting one type into the other [13], often assemblies can be described using both precedence and AND/OR information. The latter is the more flexible of the representations and there are a number of products that can be formulated using AND/OR graphs alone.

Precedence graphs are constructed by determining the parts directly obstructing the removal for each component and hence which other components must be extracted prior to their removal. By adhering to the structure – the existence of precedence relations stops every combinatorially possible permutation of parts resulting in a feasible subassembly [14] – the graphs hold all of the information on the possible solution paths.

Thus precedence graphs are conceptually simple and, as they consist of conditions that must be satisfied by the sequences, they are compact and have few nodes.

A major disadvantage with using precedence graphs to represent an assembly is that, although it is possible to concatenate "OR" sequences using dummy operations, for most products, no single graph can encompass every sequence [15].

Precedence relationships themselves, which represent individual operations or logical combinations of operations, can encompass all plans; however, this set of operations must be fixed – only the order can change – and can only be executed serially (parallel operations cannot be represented).

In fact, six types of disassembly relationships exist [16]:

- No precedents;
- No antecedents;
- AND relationships;
- OR relationships;
- AND relationships within an OR;
- OR relationships within an AND.

A representation methodology should be able to cope with these types, as well as have the ability to handle parallel disassembly operations.

AND/OR graphs map the tasks that can be performed on the assembly and thus are capable of depicting all technically feasible sequences. They are therefore explicit and it can be seen when operations can be executed in parallel [13, 17].

The drawback of this completeness is that the growth of AND/OR graphs is exponential and representing all DPPs using AND/OR graphs is computationally expensive [18]. As a result it has, in general, proved impractical to use AND/OR data in order to find optimal disassembly paths and the search for all solutions leading to combinatorial explosion.

These problems are reduced by organizing AND/OR information expressing the DDPs into the format utilized by HGA and then optimizing. This is especially true in the case of complete disassembly.

3 HIERARCHICAL GENETIC ALGORITHM

Many (complex) systems have a hierarchical structure. In conventional GAs and other optimization methods such hierarchical systems are transformed into a one-dimensional data array. However, these arrays are not suitable for expressing problems having hierarchical structures as lower level genotype variables depend upon upper level genotype variables.

Although there are several methods for dealing with hierarchical structures [5], for concurrent optimization of variables in different levels of the structural hierarchy, other techniques must carry out separate optimizations and the structure is not accurately represented. This is especially true as problems grow in size and finding the optimal solution becomes more complicated.

Such issues are avoided by using a genetic algorithm in which the hierarchical genotype coding exactly expresses the structure and detail of the hierarchical system. This uses the idea that when the upper level genotype variables change, the lower level genotype variables must also change. As the length of the genes may additionally vary, new crossover and mutation operators for treating the hierarchical genotype representations have also been required to be defined [5].

Crossover operations between individuals are conducted by: selecting another individual as the crossover partner and then exchanging the corresponding genes of the individuals, where to preserve consistency, all corresponding lower substructures are also swapped. Mutation operators are applied to the set of genes at the highest level of the hierarchical structural system, and then recursively applied to their child genes in the same manner as the crossover operator.

To represent a hierarchical structural system:

When a substructure has \( n \) lower substructures, each of which has \( a \) alternatives \((i = 1, 2, \ldots, n)\), this is denoted \((a_1, a_2, \ldots, a_n)\). When the substructure corresponds to the \( s \)th alternative for the \( t \)th substructure at the upper substructures level, the prefix symbols \( s, t \) are added to the original notation \((a_1, a_2, \ldots, a_n)\). If there are upper level substructures, these procedures are repeated, and further prefix symbols are added to the notation.

Hence, using this method, each of the substructures is independently described. The positions of all substructures in the hierarchical structure are denoted by nodes, where a node located higher than the node being considered is called a "parent" node and one located in a lower position a "child" node.
Antecedents of the subassemblies are used as they remove ambiguity from the possible solution paths. Precedents can cause situations where operation \( x \) is preceded by operation \( y \), only if \( y \) is not proceeded by operation \( z \); this further complication is eliminated using antecedents.

The methodology is outlined as follows (this procedure has been automated):

1. First enumerating the subassemblies and operations, using these labels, the antecedents for each subassembly are found. These are written in terms "&" and "\(/", which are used to express AND and OR operators respectively. The operations associated with the antecedents are also listed, as well as any other optimization criteria, e.g. the profit gained in releasing the subassemblies.

2. If full disassembly is required, now begins a process of amalgamating the subassemblies, during which the number of alternative paths is diminished (for incomplete disassembly only stage 1 is performed):
   1. Single components are initially removed from the list of subassemblies, as they have no antecedents, operations or profit associated with them.
   2. Subassemblies with a single AND antecedent consisting of purely of components are identified. Removing these, the antecedents of the remaining subassemblies are checked to see if they contain only the removed subassemblies. If the answer is "yes", the label of the removed subassembly is replaced with its antecedent. For example, if the only antecedent of subassembly 18 is 20&21, where 20 and 21 are components, then if subassembly 12 has antecedents 15&18 / 13&16, this is replaced by: 15&20&21 / 13&16.
   3. In a similar manner, all subassemblies with a single AND antecedent are found and removed. Again, if the antecedents of the remaining assemblies contain any of the removed subassemblies, these are replaced with the associated antecedent.

(Remark: 3. may need to be done iteratively as the removed subassemblies may contain other removed subassemblies.)
4. In doing 2. and 3., when the antecedents are replaced, the operations and other values linked to the subassemblies are added to those already present, leading to a partial ordering of the operations and the costs/benefits carrying these out.

- With only OR operations remaining, the data is in a form that can then be optimized using HGA, such that the path that minimizes/maximizes the optimization criteria can be located.
- The flexibility of the AND/OR representation in showing parallel operations can be maintained whilst doing procedures 1–4. However, it is more simple to do a post-analysis of the output sequence to highlight whether any concurrency to the operations exists and a routine is added to achieve this.

The approach is now illustrated by way of two examples. The first of these is the famous "Bourjault's Pen" [19], a simple example to highlight the details of the solution method.

4.1 Example 1: Bourjault's Pen

The benefit of using HGA is that, not only is the structure of the assembly maintained, but it is also used at the core of the optimization process. In the remainder of this section it is shown how, by using the subassembly antecedents generated from the AND/OR information, HGA can be implemented such that optimized solutions are found.

Figure 2: Pen studied by Bourjault [19]. (After Lambert and Gupta [11])

To outline the technicalities in the application of HGA in practice a deliberately simple case study is first considered. This is the well-known disassembly of a ballpoint pen, as introduced and studied extensively by Bourjault [19]. The assembly for this pen is shown in Figure 2.

Table 1: Subassemblies for Bourjault's pen.

<table>
<thead>
<tr>
<th>Label</th>
<th>Subassembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABCDEF</td>
</tr>
<tr>
<td>2</td>
<td>ABCDE</td>
</tr>
<tr>
<td>3</td>
<td>ABCDF</td>
</tr>
<tr>
<td>4</td>
<td>ABCD</td>
</tr>
<tr>
<td>5</td>
<td>ABF</td>
</tr>
<tr>
<td>6</td>
<td>BCD</td>
</tr>
<tr>
<td>7</td>
<td>AB</td>
</tr>
<tr>
<td>8</td>
<td>AE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operation</th>
<th>Parent</th>
<th>Children</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>1</td>
<td>2, 15</td>
<td>4</td>
</tr>
<tr>
<td>(2)</td>
<td>1</td>
<td>3, 14</td>
<td>5</td>
</tr>
<tr>
<td>(3)</td>
<td>2</td>
<td>8, 6</td>
<td>4</td>
</tr>
<tr>
<td>(4)</td>
<td>2</td>
<td>4, 14</td>
<td>8</td>
</tr>
<tr>
<td>(5)</td>
<td>3</td>
<td>4, 15</td>
<td>9</td>
</tr>
<tr>
<td>(6)</td>
<td>3</td>
<td>5, 9</td>
<td>1</td>
</tr>
<tr>
<td>(7)</td>
<td>4</td>
<td>10, 6</td>
<td>3</td>
</tr>
<tr>
<td>(8)</td>
<td>4</td>
<td>7, 9</td>
<td>3</td>
</tr>
<tr>
<td>(9)</td>
<td>6</td>
<td>11, 9</td>
<td>5</td>
</tr>
<tr>
<td>(10)</td>
<td>5</td>
<td>7, 15</td>
<td>9</td>
</tr>
<tr>
<td>(11)</td>
<td>9</td>
<td>12, 13</td>
<td>5</td>
</tr>
<tr>
<td>(12)</td>
<td>8</td>
<td>10, 14</td>
<td>6</td>
</tr>
<tr>
<td>(13)</td>
<td>7</td>
<td>10, 11</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2: Permissible operations upon the subassemblies and the profit recovered in performing them.

Once the operations have been defined it becomes a fairly trivial task to identify the antecedents for each subassembly. As a result, using Table 2, the information in Table 3 is generated, where "&" is used to specify that the two antecedent assemblies are released in parallel (AND operator) and "/" represents a choice between antecedents (OR operator). A graph showing the relationships between the subassemblies is given in Figure 3.

Table 3: Antecedence information for the subassemblies.

If a solution path consists of only AND operators, i.e. there are no alternative paths and only leaves of the hierarchy are involved, when optimizing, HGA merely combines any associated values. This evaluation gives the same result at all stages of resolving the system and must be done for every iteration. This redundancy is removed by combining AND operators and hence the amalgamation procedure described in 1–4 is done.

The first stage involves removing the components from Table 3. Subassemblies 10–15 are single components and these are therefore eliminated. It can also be seen that subassemblies 7–9 have solitary antecedents which consist of only components and these are also subtracted from Table. Studying the remaining subassemblies, it is ascertained that 2–6 all have antecedents containing the removed subassemblies and hence for these antecedents, "7" is replaced with "10&11", "8" with "10&14" and "9" with "12&13". Additionally the profits and operations
associated with 7–9 are added to those of the antecedents altered, leading to Table 4.

Figure 3: Antecedence graph of the subassemblies.

From Table 4, it is subsequently observed that subassemblies 5 and 6 have lone AND antecedents and so these are removed. As before, it is recognized that subassemblies 2–4 have antecedents which include those deleted and, as a result, in these antecedents “5” and “6” are replaced with “10&11&15” and “12&13&11” respectively and the related profits and operations are again incorporated into those associated with the antecedent. Finally Table 5 is obtained.

This data is in the correct form to be solved by HGA as it can be represented by the hierarchical structure given in Figure 4. In this diagram, the labels within the solid boxes are the number of choices available at each level of the hierarchy.

When this label is “0”, it represents a leaf in the tree and attached to each leaf are the operations that lead to parts of the assembly being released. The respective profit obtained by performing the operations is then shown by the numbers on the hyperarcs.

By examining Figure 4, for total disassembly of the pen, it is obvious that the maximum profit is found when the sequence of operations is either:

(2) – (5) – (7) – (9) – (11) or (2) – (5) – (8) – (13) & (11).

Optimizing the problem using HGA confirms this.

4.2 Example 2: Kang et al. Photocopier

Figure 5: BOM for a photocopier (After Kang et al. [20]).

Table 5: Antecedence information after steps 3 and 4 have been performed.

Table 4: Antecedence information after steps 1 and 2 have been performed.

Table 5: Antecedence information after steps 3 and 4 have been performed.
Attention is turned to the more complex and realistic industrial example of dismantling a photocopier in order to obtain the materials for recycling. This was first examined by Kang et al. [20] and, following that article, the Bill of Materials (BOM) is shown in Figure 5.

This example is used to illustrate how the possibility of incomplete disassembly is included as a solution and how extra constraints are added to the problem structure. Whilst the former can increase the number of levels in the hierarchy, when compared to complete disassembly, the benefits of HGA to deal with both wide structures and sub-hierarchies means that the optimal path can still be obtained efficiently.

<table>
<thead>
<tr>
<th>Label</th>
<th>Sub</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABE-L</td>
<td>965</td>
</tr>
<tr>
<td>2</td>
<td>AE-L</td>
<td>949</td>
</tr>
<tr>
<td>3</td>
<td>BE-L</td>
<td>915</td>
</tr>
<tr>
<td>4</td>
<td>ABG-L</td>
<td>880</td>
</tr>
<tr>
<td>5</td>
<td>E-L</td>
<td>829</td>
</tr>
<tr>
<td>6</td>
<td>AG-L</td>
<td>817</td>
</tr>
<tr>
<td>7</td>
<td>BG-L</td>
<td>806</td>
</tr>
<tr>
<td>8</td>
<td>G-L</td>
<td>739</td>
</tr>
<tr>
<td>9</td>
<td>ABEF</td>
<td>721</td>
</tr>
<tr>
<td>10</td>
<td>GHIJ</td>
<td>653</td>
</tr>
<tr>
<td>11</td>
<td>GHKL</td>
<td>507</td>
</tr>
<tr>
<td>12</td>
<td>IJKL</td>
<td>482</td>
</tr>
<tr>
<td>13</td>
<td>AEF</td>
<td>374</td>
</tr>
<tr>
<td>14</td>
<td>BEF</td>
<td>362</td>
</tr>
</tbody>
</table>

Table 6: Photocopier subassemblies and respective disposal/recycling costs.

The available combinations of materials A – L are given in Table 6. To set up the optimization problem, a cost is associated with each combination and singular material and these are also given in Table 6.

Often it possible to recycle an assembly consisting of more than one material, if the material types are compatible. However, none of the groupings in Table 6 involve only metals or plastics and, as a result, a disposal cost is assigned to all assemblies of more than one substance.

For the individual materials, value is gained by recycling and this is reflected in Table 6, by a negative cost. The exception is the toner (material A), which is considered a hazardous substance having a harmful environmental impact, thus it must be disposed of safely.

From Kang et al. [20], the set of allowable operations to release the materials for recycling, and the cost of performing these operations, are given in Table 7. From this table the relationship between the subassemblies is depicted in Figure 5.

As in Section 4.1, the optimal disassembly path is found by maximizing the total profit recovered, where the profit is found using the formula:

\[ \text{Recovered Profit} = \text{Cost(Child 1)} - \text{Cost(Child 2)} + \text{Cost(Operation)} \]  \hspace{1cm} (1)

Figure 6 shows that, similarly to Bourjault’s pen, if complete disassembly is required, converting the AND/OR information to a format solvable using HGA notably reduces the problem size.
It can also be seen from Figure 6 that a regular occurrence in DDPs is duplication of at least one part of the hierarchy. It is particularly advantageous for HGA to recognize these recurring sub-hierarchies as they can be optimized as a pre-process and therefore do not have to be repeatedly resolved, thus increasing the efficiency.

In fact, only the data pertaining to subassemblies 1, 2, 3, 4, 8 and 9 remain and, optimizing with the HGA routine, the series of operations leading to the maximum recovered profit is quickly obtained:

\[(28) \rightarrow [(8) \rightarrow (10) \rightarrow (12)] \& [(23) \rightarrow (13) \& (27) \rightarrow (14) \& (15)].\]

**Adding further constraints**

As HGA uses the problem structure directly, extra constraints upon the ordering of the operations are added without difficulty. This is done by simply eradicating the paths that become infeasible through the introduction of the constraints.

By way of example, suppose that it is necessary to remove the toner (subassembly 20) as early as possible within the disassembly sequence. Looking at Table 7, there are four initial operations: (1), (2), (16) and (28); of these only operation (2) involves the toner. Hence all paths that do not follow directly from operation (2) can be ignored.

As a result, the new optimal path is:

\[(2) \rightarrow (30) \rightarrow [(10) \rightarrow (12)] \& [(23) \rightarrow (13) \& (27) \rightarrow (14) \& (15)].\]

**Incomplete Disassembly**

Determining the depth of disassembly, the extent to which components should be subtracted, is also a vital part of creating an optimal process plan.

Depending on the motives for disassembly, the “best” cause of action maybe to halt the plan partway through its execution and leave the rest of the assembly as it is. This is particularly true when balancing the value of the recovered elements with the cost of obtaining them and when performing maintenance.

To this end, a robust solution method for discovering the optimal ordering of removal of parts must incorporate the option of incomplete disassembly. Formulating the problems such that they are optimized using HGA admits this possibility.

This is done by inserting an extra choice when each “OR” function is encountered. This choice uses a dummy operation – operation (0) – and is a potential dead-end, allowing the resolution routine to stop at this point if it is deemed to be the optimal result.

Figure 7 illustrates the new solution hierarchy structure for the leftmost branch of the graph associated with the photocopier when applying this methodology. As previously, an algorithm has been constructed to automate the generation of this arrangement of the operation data.

The value associated with selecting operation (0) depends on the stage of disassembly and often corresponds to cost of disposing of what remains of the assemblage. From Figure 7 it can be seen that the full structure of the problem must be used; it is no longer practical to condense the representation. However, although supplementary levels can be added to the hierarchy in doing this; including partial disassembly predominantly increases the width of the graph. HGA is specifically designed to cope with wide hierarchies, hence the speed and efficiency of optimization is not impaired.

Furthermore, in this situation, the benefit of finding the optimal paths for repeated/shared sub-hierarchies as a precursor to solving the main body of the problem becomes more pertinent.

By way of an example, regard the cost of the operations as being zero in all cases. The recovered profit is then a comparison between the cost of disposing of the remaining subassembly and of disposing/recycling its potential children.
The optimization criterion is taken to be that disassembling any subassembly must be advantageous. Thus disassembly only continues as long as the disposal/recycling cost of two children is less than the disposal of their parent (i.e. the profit is positive). To do this, the operation (0) is given zero value.

The optimal sequence of operations, as can be seen in Figure 7, is: (1) – (3), i.e. by performing operations (1) and (3) a profit is achieved; after this the cost of disposal of the rest of the subassembly is less than the disposal of its children.

5 SUMMARY AND CONCLUSIONS

This paper introduces a Hierarchical Genetic Algorithm, which has previously been used to optimize a number of engineering problems and successfully utilizes it to find the optimal disassembly path from the group of all feasible process plans derived from AND/OR information.

Products are often considered as modular systems, with the resulting AND/OR data that represents them having a hierarchical configuration. This structure can be exploited by solving using HGA.

The advantages of using HGA to optimize problems best depicted by AND/OR graphs is highlighted through the solution of several assemblies from the literature. In all cases, the optimal disassembly path can be found using the methodology. Additionally, by using the antecedents to convert the problem definition into that required for HGA, the size of the structure is condensed.

However, with the possibility of combinatorial explosion with AND/OR graphs, even with a reduction in the number of solution paths, there is still the difficulty that the problem may get unmanageable. For very large examples, by combining HGA with “limping” – where the product is first separated into principle modules – and “branch and bound” techniques [18], an efficient strategy may be built to find the optimal disassembly path for even highly complex systems.

6 ACKNOWLEDGMENTS

This study was supported in part by a Grant-in-Aid for Forming Strategic Research Infrastructure from Ministry of Education, Culture, Sport, Science, and Technology, Japan (MEXT), 2008-2012 (S0801058).

7 REFERENCES

Integrated Design for Solving Imaginary Complexity in Design

K. Pimapunsri¹, S. Tichkiewitch²,
¹ Department of Industrial Engineering, Faculty of Engineering, KMUTNB, Bangkok, 10800, Thailand
² G-SCOP laboratory, Grenoble Institute of Technology, Grenoble, 38031 Cedex 1, France
kusolp@kmutnb.ac.th

Abstract
Complexity in Design is defined by N. Suh as the measure of uncertainty in achieving the functional requirements (FRs) of a system within their specified design range. According to his complexity theory, here is described how integrated design is a way to get a non complex solution for an imaginary time independent complex problem. Two case studies in the field of furniture design and building design demonstrate how the integration of the actors concerned by the life cycle of the product can achieve this goal, using a cooperative design modeller and working with the just need concept.

Keywords:
Integrated Design, Imaginary Complexity, Just Need, Furniture, Passive House

1 INTRODUCTION
The environment of global manufacturing and global economy brings us into the world of complexity. Products are always changed to meet today’s requirements from both internal and external companies. These requirements challenge the designer to develop products and systems that are able to satisfy the customer and relative individuals through the end of its life.

In the design keynote at CIRP 2005 General Assembly, Suh [1] presented the fact that “engineered systems in the future will become more complicated since the number of the functional requirements (FRs) will continue to increase requiring many layers of decomposition, unless fundamental principles for reducing complexity can be devised. Complexity of these systems will depend on our ability to successfully synthesize and operate large systems without making them complex”. In fact, in this approach, the complexity does not always depend on the number of FRs we have to take into account. "When there are many FRs that a system must satisfy at the same time, the complexity of the system is determined by whether or not the design parameters (DPs) chosen satisfy the FRs."

The next part describes the concept of complexity and axiomatic design proposed by N. Suh in [1-3]. In the third part, the concept of integrated design, as defined at the G-SCOP laboratory [4-6] is reviewed. In the fourth and the fifth parts, two case studies are presented and show how integrated design can solve imaginary time-independent complexity problems, one being in furniture design and the other in building design. The final part gives the conclusion and discussion.

2 COMPLEXITY IN ENGINEERING DESIGN

2.1 A brief introduction to complexity
The term “complexity” is commonly found in use throughout all fields of science including physics, biology, sociology, etc. It can be described as: distinct components that are joined and mutually entangled, a change in one component will propagate through a tissue of interactions to other components which in turn will affect even further components, including the one that initially started the process [7].

In engineering design, a complicated object may be decomposed into elements which can be defined and recomposed as simple objects, while a complex object can be decomposed into elements as well but they may not be defined independently. Suh [3] defined complexity as:

“Complexity is a measure of uncertainty in understanding what it is we want to know or in achieving a functional requirement (FR).”

Suh further mentioned that complexity arises when we cannot give a complete description to a product or a system. Complexity is classified into two kinds: time-dependent complexity and time-independent complexity. Time-independent complexity is further divided into time-independent real complexity and time-independent imaginary complexity, depending on its root cause and does not require a time dimension. On the other hand, time-dependent complexity involves time as one of its determinants. It is also divided into two different types: time-dependent combinatorial complexity and time-dependent periodic complexity.

This study focuses on the imaginary time-independent complexity occurring in the design process. This complexity is defined as complexity that is not real complexity but arises because of the designer’s lack of knowledge an understanding of a specific design [3]. The imaginary complexity can exist even though the system range is inside the design range. It often exists when there are many FRs that must be satisfied in particular sequence i.e. decoupled system.

2.2 Axiomatic design
Axiomatic design is a systems design methodology using matrix methods to systematically analyze the transformation of customer needs into functional requirements, design parameters, and process variables [8]. The axiomatic design theory has been developed by N. P. Suh [2] since the 1990s.
In Axiomatic Design, there are two axioms that govern the design process. 

Axiom 1 – the independence axiom

The first axiom is to maintain the independence of the FRs. In designing engineered systems, an FR is a function to be achieved by the designed system. FRs are defined as a minimum set of independent requirements that completely characterize the functional needs of the product in the functional domain.

The independence axiom decomposes the design process into hierarchical levels as branches. These branches are mapped one to another by zigzagging between the four domains until the design is complete.

The mapping process generates design equations and design matrices which describe the relationships between the characteristic vectors of the domains. The design matrix will identify the system which can be uncoupled, decoupled, or coupled design. In order to maintain the independence of the FRs, designers must develop a design matrix in either uncoupled or decoupled design (diagonal or triangular) form.

The information axiom is to minimize the information content of the design. There could be many different designs that satisfy a given set of FRs. However, the best design is the design with the highest probability of success that requires the least amount of information to satisfy the FRs. Designers must propose the system that has least variance in order to make the proposed system range lie inside the design range specified by the FRs as much as possible. Designers must reduce the information content by eliminating the bias or reducing the variance of the design range.

Engineered systems today have become more and more complicated since the number of FRs is increasing in order to satisfy the today customization environment. This requires many layers of decomposition and easily leads the designers to meet complexity even though the design is decoupled. To solve this kind of complexity, an integrated design approach is required. The next part describes a method for solving imaginary complexity.

3 METHOD FOR SOLVING COMPLEX DESIGN

3.1 “Just need” notion

Suh [1] states that imaginary complexity arises because of the designer’s lack of knowledge and understanding of a specific design itself. When there are many FRs that a system must satisfy at the same time, the quality of the design in terms of the independence of the FRs affects the uncertainty of satisfying the FRs. An uncoupled design is likely to be least coupled. However, the complexity of a decoupled design can be high due to imaginary complexity. If we do not understand the system – it is not really complex, but appears to be complex due to our lack of understanding. Imaginary uncertainty can still exist in a good design when we are ignorant of what we have.

This study proposes a method to solve the imaginary complexity of design problem, in other words, when the unknown design matrices can be in triangular form.

This method called Integrated Design is based on a participation of all of the actors, concerned with one aspect or another of the life of the product during the design process. Integrated Design is firstly cooperative design, in the fullest sense of this concept: the actors are brought together to make group decisions.

In Integrated Design and in order to perform the design tasks with less problems and contradictios, the actors must have a coincident notion of design, which is called the “just need concept”, defined by Brissaud et al. in [9], and described as follows:

- Each actor has to give constraints as soon as possible. This notion enhances other actors to have further information to evaluate the design and to define the product more precisely.
- But each actor has to give the constraints that he is able to prove. To emphasize the previous notion, this notion permits the actors to contribute only with justified constraints, not just because of personal preference. The actor must be able to prove why it is necessary to take into account such constraints.

These notions help the designers, all people who take part in any stage of the product life cycle, to realize what the constraints of the others are. The given constraints will be propagated to one another by the integrated design system, which is described in the next section.

3.2 Design system to support collaborative environment

This section presents the integrated design system named “Cooperative Design Modeler, CoDeMo” [4-6], developed at the G-SCOP laboratory. CoDeMo is used to bring the designers into a virtual collaborative environment which allows them to access the relative data and information of the design problem which are stored in a shared database. It supposes that the designers have their own expertises on the design problem.

A collaborative environment helps the design team to communicate with each other, to exchange information, or to present their constraints into a virtual meeting room. However, the design process is often a conceptualization, which is not easy to share and is seldom documented formally. Otherwise, some intent information may be lost.

As a result, a complex design is often carried out through collaborative work. Furthermore, different disciplines are concerned with different objectives but must be integrated to achieve the common goal. This is a reason why the integration takes an important role in this part of the design process.

CoDeMo allows the designers to work together in a virtual collaborative environment. It supports the designers as they create and share data, constraints, and knowledge with one another. The concepts of CoDeMo (multidisciplinary concept: multi-actor, multi-view, multi-representation; models for integration: knowledge model, data model; knowledge management: data propagation, data translation, substitution method) had been well described in [4], [10] and [16].

Gaucheron [15] proposed that designers have to prepare and discuss common problems before they begin the design process. The interactions of this preliminary discussion can be expressed in the form of a production rule [17], which is an element of knowledge. It allows the designers to create temporal knowledge which are elements of features (factual knowledge). The created features and production rules of individuals are stored as knowledge modules in a feature-based engine of CoDeMo. CoDeMo supports the interactions among the individuals during the design process, and the interactions enrich the knowledge modules that are used for solving design problems.

To use CoDeMo for solving the imaginary complexity of a design problem, the designers first retrieve initial information and constraints from the shared database provided by all individuals in the design team.

As soon as an individual has sufficient information to be able to define a DP that satisfies a given FR, the proposed DP will be propagated to the shared database.
The designers who may be concerned will be notified and receive the translated information by the design system. On the other hand, if the proposed DP does not satisfy the FR, the proposer must adjust the DP values to satisfy the FR. Otherwise, if the proposed DP conflicts with the other existing DPs, any concerned designers will be asked to negotiate about the problem to revoke or adjust that conflict, as shown in Figure 1.

![Figure 1: Interactions and evaluation process of DP.](image)

4 A CASE STUDY OF FURNITURE DESIGN

4.1 Complexity in furniture design

At the beginning of the furniture design process, the designer who is concerned about the global form and the aesthetic of the product e.g. shape, form, color, texture, etc., must propose a conceptual design which satisfies the primary functionalities [17]. The conceptual product model is usually handled by a CAD system and should be manipulated with the primary functional requirements. It initially provides the shape and the propositions that articulate the functionalities and the style of the product.

Before kicking-off the collaborative environment, the conceptual product must be transformed into the design system, CoDeMo, as the initial information. To accomplish this task, the designer converts the CAD file of the conceptual product into a universal file format, such as the STEP format. This universal file that contains data model of the conceptual product consequently is imported into CoDeMo and is transformed into entities of the product [18]. The transformation process permits the designers to recognize the product model in different views. The data translation translates the product model into different points of view while the data propagation propagates the corresponding information to the corresponding views, trade views and common views.

However, the results of this initial design phase do not provide any assembly solution, neither mechanical behavior, nor manufacturing method. What kind of fastener should be applied for assembling the parts? What type of material and what thickness of the parts should be defined to resist a given load? What is the manufacturing method that should be applied for producing the parts? These are questions that the designers have to answer.

At this time, all designers must bring their information and constraints into the design system in order to determine the design parameters. The collaborative environment provided by CoDeMo allows the designers to share their information to solve the design problems that occurred during the design process. This section presents how to deal with imaginary complexity due to the number of FRs from multiple stakeholders, including assembly aspect, manufacturing aspect, and mechanical aspect.

**Assembly aspect**

In the assembly view, the assembler must provide assembly solutions to fasten the parts. To fasten any two parts, there is usually more than one solution. Therefore, the assembler possesses a feature library that stores available assembly features and their characteristics as presented by example in Table 1.

<table>
<thead>
<tr>
<th>Assembly features and characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dowel</td>
</tr>
<tr>
<td>Type (Strand, Groove)</td>
</tr>
<tr>
<td>Diameter</td>
</tr>
<tr>
<td>Material (Wood, Metal)</td>
</tr>
<tr>
<td>Maximum load</td>
</tr>
<tr>
<td>Screw</td>
</tr>
<tr>
<td>Type (Tapping, Confirmat...)</td>
</tr>
<tr>
<td>Diameter</td>
</tr>
<tr>
<td>Length</td>
</tr>
<tr>
<td>Maximum load</td>
</tr>
<tr>
<td>Connector Joints</td>
</tr>
<tr>
<td>Type (Minifix, KD fitting...)</td>
</tr>
<tr>
<td>Diameter of housing</td>
</tr>
<tr>
<td>Length of housing</td>
</tr>
<tr>
<td>Diameter of bolt</td>
</tr>
<tr>
<td>Length of bolt</td>
</tr>
<tr>
<td>Maximum load</td>
</tr>
</tbody>
</table>

Table 1: Examples of assembly solutions.

Choosing different assembly feature may satisfy different FRs. One may satisfy a given load resistant while another one satisfies the minimum cost. For example, using the feature Dowel to fasten two parts, the designer must take into account its mechanical properties if the FR concerns the resistibility. Or else, the length and the diameter of the dowel, as well the thickness of parts, must be determined if the FR concerns manufacturability. Whenever an assembly feature is chosen, its production rules must be propagated to those concerned trade views. Table 2 shows by example the production rules of the feature Dowel.

If a dowel is applied to fix a pair of parts

Then those two parts must be drilled

If the thickness of the horizontal part is \( T \) mm

Then the diameter of the dowel is not more than \( T/2 \) mm

If the diameter of the dowel is \( D \) mm

Then those two parts must be drilled with diameter \( D \) mm

If the length of the dowel is \( L \) mm

Then the horizontal part is drilled \( 2L/3 \) mm while the vertical part is drilled \( L/3 \) mm

Table 2: Production rules of feature 'Dowel'.

**Mechanical aspect**

In this view, the mechanic's task is to define the most appropriate material type and dimension of the parts that satisfy the given FRs. The mechanic possesses a database of materials and its properties e.g. modulus of rupture, modulus of elasticity, density, etc. The criteria of standard requirements are included in the database as well. For example, a table must be able to resist a load of 100 kg. on the top, 5 accidental drops by tipping it over for testing the strength and 10,000 lateral thrusts of 300 N for
testing the durability. The mechanic must take into account these values while evaluating the design.

Regarding the criteria of standard requirements and the default dimension of parts (if they exist), the mechanic could accept the default material and dimension of the parts if it satisfies the given FR. Otherwise, a negotiation may be required if it does not satisfy the FR due to the dependency of other DPs, such as assembly solutions.

**Manufacturing aspect**

In the manufacturing view, the manufacturer must know information and constraints e.g. available resources: machines, tools, available manufacturing technology, in order to propose the manufacturing method that satisfies the corresponding FRs. The manufacturer focuses primarily on the manufacturability of parts, manages the process plan, and also the estimation of manufacturing cost. To perform this task, Pimapanusi [17] developed a specific tool named DAPP, Database Application for Production Planning, in this manufacturing view to support the evaluation task of the manufacturer.

Sometimes other DPs may influence the choice of manufacturing method e.g. assembly solution, material type, part's dimension, etc. In order to optimize the design, the manufacturer may request the concerned designers to adjust or revoke their DPs.

### 4.2 Imaginary complexity in furniture design

The integrated design system allows the designers to solve the problem of decoupled or weak-coupled design. If a coupled design is introduced at the highest level, it cannot be overcome by lower level design decisions. The system is also coupled and difficult to improve. We present here an example of designing a ready-to-assemble computer desk. The highest level of FRs for the computer desk may be stated as follows:

- FR$_1$ = Support a monitor
- FR$_2$ = Keep a computer case hidden
- FR$_3$ = Support a keyboard
- FR$_4$ = Keep cables and wires tidy
- FR$_5$ = Support feet to reduce fatigue

In addition, the constraints (Cs) for the computer desk are:

- C$_1$ = Maximize load resistant
- C$_2$ = Minimize weight ≤ 20 kg.
- C$_3$ = Minimize number of assembly
- C$_4$ = Minimize space requirement of packing ≤ 120 cm. of length and ≤ 80 cm. of width.

As proposed in the integrated design methodology, a designer has to take into account first these functionalities to propose a first view of the furniture in terms of geometry (forms and relative proportions), colour, texture, etc. The designer uses a CAD system to deliver an initial model as illustrated in Figure 2.

![Figure 2: The initial CAD model provided by the designer.](image)

With such a model, we get the functional surfaces and, as shown by Belloy [19], we can start the second step of the final design with the cooperation of all of the actors concerned about the life cycle of this product. The CAD model is transferred to CoDeMo via a STEP file, and a technologist, a mechanic, and two experts in manufacturing and in assembly have access to this model and to the specifications.

The technologist can understand the CAD model and transform it in a set of boards. The analysis of this set of boards and their relative positions permits the expert in assembly to propose the choice of the fasteners, but without knowledge about the thickness of the boards, this expert cannot define the size of these fasteners.

The mechanic is in charge of fixing the thicknesses for the boards, depending on the norms he has to respect and the quality requested for the furniture, but he needs to know the type of the fasteners in order to choose the behaviour of the connection and so to fix the limit conditions for the computation. With his results, the expert in assembly can dimension the fasteners and fix the surfaces that have to be manufactured before the assembly.

Then the expert in manufacturing can analyse the manufacturing process and evaluate the cost and delay to realize such operations.

This design loop can be done with different variants of fasteners and give in fine a set of solutions with their costs, delays and qualities.

We saw in this example that the process cannot be a linear one, going from the technologist to the expert in manufacturing though the expert in assembly and the mechanic, each actor giving a global work to the next, taking some decisions out of the just need. We saw also that the just need process permits here each actor to bring a piece of information that permits the global design process to be achieved and to deliver a global non complex solution.

### 5 A CASE STUDY OF PASSIVE HOUSE

Sustainable design in architecture describes environmentally conscious design techniques, minimizing the negative environmental impact of buildings by enhancing efficiency and exhibiting moderation in the use of materials, energy, water resources, etc. Sustainable integrated design of passive house consists of asking various trade associations to work in constant connection to be able to carry out the best compromises on the choice of materials and the technological solutions. This makes it possible to think of comfort of dwelling at the same time as saving in consumable resources for the future user. It is quite important to think of heating and air-conditioning as well as to think of the techniques of the shell.

#### 5.1 The request for passive house

The passive house standard for central Europe requires that the house fulfills the following requirements [20]:

- The house must be designed to have an annual heating demand of maximum 15 kWh/m² per year in heating and 15 kWh/m² per year in cooling energy.
- Total primary energy (source energy for electricity and etc.) consumption (primary energy for heating, hot water and electricity) must not be more than 120 kWh/m² per year.
- The building must not have air leakage of more than 0.6 times of the house volume per hour at 50 Pa as tested by a blower door.
Saving such energy relative to a standard house (today around 80%) permits the designer to consider an initial cost for the passive house that is a little more expensive than a normal one (15%), due to the specific solutions to be adopted.

Such passive houses need to take into account the multiple aspects of superinsulation, advanced window technology, airtightness, ventilation, natural energy for heating and cooling, lighting, etc. during the design process.

5.2 The shell of a passive house

According to the area and local natural resources, the traditional materials of construction are stone, brick, cob, concrete, wood, etc. All of these materials are controlled technically perfectly from the point of view of the solidity of construction, all not having therefore same environmental qualities [21].

Wood is today, without context, the material privileged in modern construction due to its many qualities. When building with wood, an architect chooses a material which can be completely recycled and which will have stored CO₂ during its development, while other traditional construction materials require a considerable amount of energy (Figure 3).

Left visible on a wall, wood can offer its best asset of comfort: its weak effusivity. Thus, a covered wood part is heated quickly, requiring less energy, and offers a particular comfort, from the weak difference between the temperature of the walls and that of the ambient air (Figure 4).

Equipped with good thermal resistance, wood takes part in the insulation of the building, in addition to its structural role. Six to eight times more insulator that breeze block or cooked brick, it competes with alveolar brick at a much lower costs of implementation (Figure 5).

Concerned also with a good heat capacity, wood plays a central role in the comfort of the habitat: the resistance to the temperature variations. Associated with an external insulator, it is indeed able to store heat and to restore it in time (Figure 6).

Considering the previous characteristics, passive houses take advantage to be in sawn timber isolated by outside. Two technologies are used for massive wood walls: glued timber (Figure 7) or KLH (Figure 8). If we consider the airtightness constraint, using 110 mm of KLH board and 200 mm of insulation in two cross layers is the best. Such a solution is also very advantageous concerning the time needed to build the wood shell: typically 10 hours for the KLH house and 1 week for the glued timber one. Both houses have large picture windows in the south and a wall in almost blind north.
With the goal of keeping a perfect airtight envelope, it is necessary to pay attention of stopping all of the junctions between the boards with a specific adhesive tape (grey band in the lower right corner) as illustrated in Figure 9, before starting the insulation (Figure 10).

5.3 Control of ventilation and heating

The airtightness of the house is an advantage for the heating but is a disadvantage for the quality of the air, keeping inside chemical pollutants emitted by certain materials (paintings and varnish, adhesives of coatings of the ground, CO$_2$,...) or biological pollutants (moulds, bacteria, dust mite...).

The use of double flow ventilation (Figure 11) makes it possible to permanently maintain an air healthy and pure. The polluted air is extracted from the water rooms, such as kitchen, bathroom, toilets, laundry... While crossing the plate heat exchanger, it deposits part of its heat before being rejected outside. The new introduced air is filtered then heated by the contact of the plates in the exchanger, before being distributed in a controlled way into the rooms to live.

A geothermic exchanger of air increases the effectiveness of the group of ventilation by recovering the energy of the basement. A Canadian well is an ecological and economic solution. Its advantages are:

- Air is preheated in the cold season without additional energy expense,
- No icing of the plate heat exchanger,
- The air in hot season is refreshed,
- Limiting the reheating of the new air to the periods of very low temperatures is allowed.

The geothermic exchanger of air uses the characteristic of the basement, according to which the temperature starting from a certain depth remains about constant any days of the year. The surrounding air is not brought directly in the house, but passes by a collector buried in the ground to a depth higher than 1.20 m.

The installation of a double flow ventilation system is only possible if all of the collectors and distributors of air are envisaged during the design of the house (Figure 12). This requires the integration of a heat engineer and a specialist in ventilation into the design team.

The principle of the sustainable integrated design is also to save energy for heating-cooling and natural resources.
One difficulty experienced by passive houses is the ability to consume the energy of heating produced by the system during summer. The quantity of heating energy produced can exceed the need for domestic water. In this case, it is essential to cool the warm water tank by voluntarily turning off the circuit of the sensors during the night.

5.4 The need of integration

The previous description of the technical problems we have to solve for maintaining the comfort of dwelling while saving consumable resources shows how much the complexity of designing a new house has increased.

5.4 The need of integration

The previous description of the technical problems we have to solve for maintaining the comfort of dwelling while saving consumable resources shows how much the complexity of designing a new house has increased.

6 CONCLUSION

Integrated design allows all actors concerned by the life cycle of the product or system that we have to design to be presented and active in a cooperative way. Asking them to react to the just need concept helps them to avoid the addition of late constraints that often create multiple contradictions and usually limit a number of negotiations.

In the design process, it is the fact that only one person could not provide the solutions that answer to all required functionalities of various specific domains. It thus could seem as an opening toward complexity. In this case, the integrated design allows the different actors to provide some bricks to picture an answer without engaging of the hypothetical solutions. This methodology was also tested in other fields such as the design of extrusion dies or the design of a car taking into account its recycling.
7 REFERENCES


Obstacles and New Opportunities for Integrated Design

L. B. Jensen, T. R. Nielsen
Department of Civil Engineering, Technical University of Denmark
Anker Engelundsvej 1, Kgs. Lyngby, 2800/Denmark
LBJ@byg.dtu.dk

Abstract
Recent developments in integrated design in Civil Engineering are outlined by describing how projects can now be informed with engineering knowledge at a conceptual level. The well-defined methods of integrated design are challenged by young engineers and architects. The development of simulation programs has been massive and engineering educational programmes can no longer teach just one design method with its related programs, but must instead give room for a continuous experimental design laboratory.

Keywords:
Integrated Design in Civil Engineering, Digital Modelling, Energy Design, Architectural Engineering, Sustainable Design

1 INTRODUCTION
Design theory and design processes in civil engineering traditionally focus on the final stages – the documentation and certification phase of the design process. Architects in the building industry have had the privilege of working with the first conceptual stages. This pattern is changing and the following will describe recent developments where an architectural perspective is applied to civil engineering with a view to sharing information across the disciplinary boundaries.

The term ‘Integrated design’ captures this effort to integrate technical scientific knowledge from the conceptual design phases onwards. The result of an EU research project 10 years ago, Task 23, (integrated design in civil engineering in a Scandinavian context) is an approach based on a fully developed theory and the facilitating software to go with it [1]. Essentially, Task 23 suggests a method for structuring the design process as first defining a space of solutions and then a series of iterations where parameters are varied, which results in information for project decisions.

Research on Task 23 integrated design clearly showed that the greatest reductions in energy consumption are to be obtained in the earliest design phases. Operating buildings accounts for 40% of the energy consumption of European societies [2]. The most effective way to reduce this figure is to ensure that building design is informed from the start with the technical-scientific knowledge needed to construct potentially net-zero-energy buildings (an EU requirement for new publicly funded buildings from 2020).

Integrated design has now been widely implemented and is leading to a perception of the engineer as a member of a team in which the engineer contributes with his knowledge in parallel with other professionals right from the very earliest design stages.

A number of European countries (including Denmark) have the goal of being CO2 neutral from 2050, so a quantum leap is needed with regard to the information level of the design process, if this goal is to be reached.

There is consensus on the need for a more informed design process if the construction industry is to reach the goal of low-energy buildings and sustainability. However, there are obstacles to this process, as described in what follows.

1.1 Integrated design in Civil Engineering
For several years, design engineers educated at the DTU (Technical University of Denmark) Department of Civil Engineering have received teaching influenced by Task 23 Integrated Design Process, which as mentioned was originally conceived under the auspices of the EU. In the traditional design process, architects control the earliest design stages. The perspective of architects is to create coherence between the use and cultural qualities of a building. But the value of the building – for the owner and society – includes many other aspects.

In a traditional design process, the engineer contributes at the end of the design process by redesigning if necessary for safety and buildability. The documentation is another traditional task of the engineer in the final phases of the design process.

Integrated design, e.g. Task 23 and other well-developed methods, challenges this traditional consultancy process and proposes an engineering involvement in the design process from the earliest conceptual design phases, at which stage operating costs and other engineering aspects are (or should be) important design parameters.

In this model, the initial stage is an exploration of the needs and requirements of the various stakeholders of a building project in order to define a multidisciplinary team in defining the space of solutions.

Freeware, Multi-criteria design method tools (MCDM-tools) were part of the Task 23 effort to facilitate a multidisciplinary team in defining the space of solutions.
The MCDM approach has not been further developed in Civil Engineering for final design phases, although it is described in other integrated design methods [3]. Software to facilitate the basic iteration and parameter variations was also developed as freeware in a further development attached to Task 23 (www.Idbuild.dk). Idbuild is fundamentally an approach, in which a room in a future building is modified (height, width, window orientation, u-values and g-values, window size, etc. are varied) and the effect on the energy balance is graphically illustrated as a deviation from a reference. Subsequently, the early space of solutions as defined by the criteria and objectives is systematically scanned for solutions. Solutions are then provisionally calculated and compared with the criteria and aims. On this basis, it should be possible to make a design decision. In the original theory of integrated design Task 23, the iterations were to be facilitated by still more and more advanced and detail-focused software [4].

1.2 Integrated design – a creative tool or a just a checklist?

Most kinds of integrated design processes include an initial outline of a space of solutions. This is often perceived as a checklist: the achievement of certification, etc. But viewed in another way, it is a demarcation of design ideas, and in this sense should play an important role in creative idea generation. Theories of integrated design do not profoundly address how the initial ideas that start the iterations are generated or how the synthesis of options is made [5] but when students work with integrated design, new creative options arise that were not foreseen in the initial theory and methods of integrated design. Experienced design educators recognize a tendency for students to generate many more ideas in a tightly defined framework than when working in a mentally completely open space, and in this sense integrated design could actually focus the students' creativity. Progressively more and more advanced software is involved in the iterations and used in yet new ways and in other design phases than it was originally designed for. Based on the supervision of numerous projects made by students in DTU civil engineering in collaboration with architectural studios, this paper outlines the latest developments and current challenges in integrated design in Civil Engineering.

2 OBSTACLES AND NEW OPPORTUNITIES FOR INTEGRATED DESIGN

2.1 Simulations as an investigative and creative activity

Simulations are an important part of the informed design process. There are numerous simulation programs for almost every engineering discipline and many universities have developed their own simulation tools – also DTU. This simulation software allows for the handling of enormous amounts of information and makes possible the quantum leap in the level of information in the design process that is needed to reach the ideal of CO2 neutrality and low energy buildings. The perspective is promising, especially with regard to building physics. Structures are calculated with respect to maximum impact, that is, fixed factors. Indoor climate and energy performance, however, are dynamic, because they vary under the influence of the outdoor climate on the building, changing by the hour and season. Simulation tools developed to inform an integrated design process have in this sense cleared the way for new architectural expressions seen in recent climate-dynamic architecture. The performance of climate-dynamic buildings is complex and can only be handled digitally. Since the energy crisis in 1973, the focus has been on the reduction of energy used for operating buildings. At first, this was done by reducing heat loss by using more insulation and reducing glazed areas. But by the 1990s, focus shifted to the building's potential energy gain from the outdoor climate. Optimizing the building's symbiotic relationship with the exterior climate is the essence of intelligent net-zero-energy buildings. Simulation programs are the admission card for non-traditional architecture to explore energy gains from the natural environment. In 2006, Danish legislation became based on a government-approved energy balance simulation program developed by DTU and SBI: Be06 [6]. To obtain a building permit, a building's energy balance must now be documented using this program.

As mentioned above, it is in the initial conceptual design stages that the most important reductions in the energy consumption of a building can be obtained. Therefore it is decisive to have simulation software that can be used for informing this level of the design process. The early design phases are characterized by a wide search of the entire space of solutions. In other words, there is a need for both speed and precision, which creates a serious dilemma. Speed is necessary to inform an on-going creative design process. Precision is mandatory because inadequate simulations risk misinforming important design decisions. There are two basic categories of energy-balance/indoor-climate programs suitable for early design processes. The first are simplified hour-based simulation programs limited to fragments of a building (simulations at room level). Adding rooms together to match the scale of the building gives an indication of how the entire building will perform. This is the principle behind e.g. Idbuild, developed at DTU Civil Engineering to facilitate the early, conceptual design processes [7]. In the second category, calculations are based on monthly values, which is the principle behind e.g. the official Danish BE06 program. As described above, the theory of integrated design processes often prescribes the use of yet more advanced simulation software as the iterations develop. The software developers explicitly aim at different levels in the design process as described above, but there is a trend among engineering students to transcend the prescribed uses of the programs. Simulation programs developed for the final documentation phases are used by students in the early conceptual phases. Students of design engineering display a new equilibrism tendency to use the various simulation tools dynamically and freely depending on the specific design challenge in question. Students switch between different simulation programs according to what they believe will give the required information fast and precisely for the specific building and location in ways that the software developers and theorists of integrated design did not foresee. For example, the advanced final-stage oriented daylight simulation program Radiance is used by DTU students for the conceptual design phase in urban scale projects. The concept of a future low-energy building in an urban plan is informed by a simulation of sun- and daylight conditions for an abstract volume with a specific location and orientation. This ‘conceptual simulation’ is made in the Radiance program because it can be arranged to simulate daylight impact on façades. This means it can be utilized for a design process that creates the best conditions for naturally lit and not overheated future buildings in an urban plan. Radiance is normally used for
detailed calculations of daylight conditions for end-phase designs.

The same applies to the documentation software Be06, which was created to document the energy balance of finalized design projects. Its monthly-based calculations have proved fast and easily applied in an on-going design process. For instance, it can give information with regard to basic geometries – which generally accounts for a conceptual design level in building projects. The basic geometrical dispositions define the future energy consumption of a building. For example, Be06 can give information that suggests a reduction in south-facing façades, etc.

The choice of structural system and materials can also be informed at a conceptual level by Be06. E.g. if the architectural expression consists of heavy, archaic structures with a lot of thermal mass, this can affect the energy balance positively or negatively. Information on such effects is important at the conceptual level.

These simplified, fast programs, however, risk missing good design solutions due to imprecision. For instance, a design with an innovative sun-shading device risks being turned down because a fast and superficial simulation shows inadequate daylight penetration, while a more advanced daylight simulation might show something else. There is a tendency in the early conceptual phases for students to create advanced hourly-based simulations based on luminance distribution on a realistic sky (e.g. Perez sky instead of CIE sky). This is a natural development connected to a germinating craftsmanship in digital modelling and simulation, and increases in computer power [8].

To inquire into the potential of new multidisciplinary collaboration in highly-informed early design phases, DTU Architectural Engineering and the Royal Academy of Art, School of Architecture organised joint workshops. At these, students have an informal platform for exploring the interface between design and digital modelling and simulation. In a similar way, final thesis projects become an Exploratorium. At DTU Architectural Engineering, they are often carried out in collaboration with architectural studios. It is clear that digital modelling and simulation integrated in a design process create new opportunities for architectural expression and slowly open up for climate-dynamic architecture.

But there are also obstacles to these developments. One lies in the way classical engineers are traditionally educated, where they have their first teaching in decision-making late, e.g. in relation to so-called capstone projects. An endless variation of parameters does not lead to design solutions and ideas. Integrated design informed by results from simulations is an engineering specialty. Simulation can be a waste of time if it does not match the immediate context of the design process.

In order for the engineer to perform an adequate simulation, he must have knowledge and experience of design processes; he must have tried to design on his own and in groups. Because of the many possible combinations of parameter variations, a qualified choice of concept as a guiding principle for the process is essential. These kinds of open-ended situations must be introduced early in the engineer's education to create an ambitious and constructive attitude with the motivation of keeping being continuously to respond to open-ended challenges in projects. Moreover, engineers who have worked so much on design are able to perceive their engineering knowledge as concepts, e.g. the ability to synthesize extensive amounts of information on thermal behaviour into potential design concepts that can be matched with architectural concepts. The engineer specialist in integrated design also needs to know something about architecture in order to work in the interface between engineering concepts and architectural concepts.

One often overlooked part of integrated design is the graphical translation or communication of simulations. Some simulation programs already have good graphic features, but almost all simulation results still need extensive graphical reworking to parallel and match the architectural language of façades and plans (Figure 1).

2.2 Intuitive design methods and integrated design

During the early, conceptual phases, few architects design by systematically searching solutions within a multidisciplinary defined space of solutions. This is not due to a lack of capacity, but is rooted in tradition and deliberate choices. In Denmark it constitutes an important standpoint because the intuitive design method is an explicit foundation for the pedagogical programme of the Royal Academy of Art, School of Architecture in Copenhagen [9].

The logically structured systems and methods meant to give an overview of design options based on scans of all relevant concepts is bypassed by intuition. Architects educated in the intuitive design method are the main collaborators of civil engineering design engineers.

The Bauhaus School of Architecture in Dessau was one of the first schools of architecture to work explicitly with intuition, possibly inspired by modern psychological theories [10]. This was a revolutionary initiative which produced a quantum leap in form and design because it was a decisive break with the imitation-based design methods of the Beaux Arts Academies [11]. The Bauhaus impulse is, of course, still a source of inspiration.

The explicitly intuitive method is a well-defined tradition and indirectly takes a critical stand with regard to integrated design methods. It poses an important obstacle for the implementation of integrated design because transparency in the design process is necessary when working in multidisciplinary design teams. This transparency, expressed in decision charts etc., is of course a core in classic engineering design methods [12].

On the other hand, Task 23 presupposes some architectural ideas, but as mentioned their generation is not directly addressed. So the intuitive design method of the architects is a professional approach to the 'black holes' in the integrated design process and has value [13].

In the Copenhagen area, very few architectural studios work with integrated design. But intuitive design methods are challenged by requirements for documented and strictly sustainable buildings.

Engineers have taken a completely opposite viewpoint to architects in recent decades. The development of indoor-climate and energy balance simulations can be construed as a popularization of advanced specialist knowledge, which means enhanced transparency.

The improved design of program interfaces and illustrating graphics potentially gives laymen opportunities for addressing this specialist knowledge. The German architect, Dietmar Eberle argues that in this respect engineers could be obsolete as partners in the design process because architects can get the information more freely and individually by means of computer programs [14]. This is a logical consequence of the development, but a first-hand encounter with energy simulation programs will demonstrate that it takes engineering knowledge to control and evaluate simulations. As mentioned above, a defective simulation will risk misinforming the design process, and wrong information can be worse than no information. However, simulation
programs can form common ground for collaboration and integrated design when used by both architects and engineers. Similarly, drawing programs used by both architects and engineers, such as Google SketchUp, can be productive in a multidisciplinary design process.

2.3 BIM and Integrated Design in Civil Engineering

Drawing programs used in the early design phases by architects and engineers alike, and which can be imported in simulation programs, will entail a leap in the development of integrated design. It will mean new options for collaboration between architects and engineers if a special model does not need to be constructed for the simulation program. It will be a step towards the elimination of the dilemma of speed versus precision in the integrated design process, because it will be possible to test architectural ideas directly and continuously in the simulation program.

The integration of information from different simulation programs can also be revolutionized. Software developers can create platforms for plug-ins from simulation tools from different engineering subject areas. But because they were not thought out in the same way, it is still a demanding task to model and evaluate simulations.

The state of play of these promising visions was investigated in three afternoon workshops for engineering students at DTU. The students were all experienced users of energy simulation programs and various drawing programs, but had not worked with the option of importing drawings directly into the simulation program before. Two drawing programs and an energy simulation program were chosen.

IESVE contains a model tool called modelIT which can model basic geometries. There is also an option to import a gbXML file (Green Building XML) which is a common open source format created to exchange geometries between BIM programs (Revit, ArchiCAD, etc.) for simulation programs such as IESVE. IESVE has its own export tool, which functions directly in Revit. As a third and very new possibility, the import of geometries from Google SketchUp has been established. Google SketchUp Pro has a function which also allows import of CAD drawings.

The current situation (2010/2011) outlined by the reports and evaluations made by students at the three workshops is that it is possible to import a drawing into the simulation program with some difficulty. After the simulation, the information obtained must still be manually exported to the drawing program. Conclusions were that the process does not function ideally, but all students could see and experience a great potential according to their reports and evaluations.

The traditional roles of architects and engineers are dissolving, opening up for new design processes. Apart from the changes caused by severe requirements for a high level of information in the design process due to sustainability issues, new project design tools are pushing developments as well.

Project designing with BIM (Building Information Modelling) has been legally mandatory for publicly funded large construction work in Denmark since 2009. BIM holds potential in relation to multidisciplinary integrated design because all groups can place information in different layers in the same digital model.

However, BIM also poses a line of obstacles to integrated design in the early design phases. BIM models do not work dynamically in that you cannot move freely from conceptual design to project engineering and back.

3 CONCLUSION

Theories and methods of integrated design are challenged by well-consolidated architectural design traditions. These traditions have quality and the potential for being developed in the context of integrated design. In Intuitive design methods there is a reservoir of tacit knowledge and experience in addressing the ‘black holes’ in the framework of integrated design. There is the challenge that decision charts do not unambiguously suggest solutions and the synthesis necessary for creating the actual form is not directly addressed.

But if work is made on the interface between architectural design and engineering simulations and attention is paid to the graphical communication of simulation programs, the goal of sustainable buildings is within reach. The use of simulation programs as common ground for architects and engineers can give the programs a new role in the design process. Informing the early design phases has become a creative engineering activity, in which increased computer power and student familiarity with a wide range of drawing programs and simulation programs change the prescribed methods of integrated design. The dilemma of speed versus precision in simulations used to inform an on-going design process is being solved by technical development. The effect of informing the early design phases is well-demonstrated with regard to reductions in energy consumption for operating buildings.

The way in which the conceptual design phase is informed is changing rapidly and involves new ways of doing both engineering and architectural design. Integrated design is evolving in ways not foreseen in the theory behind the methods.

4 REFERENCES

[13] Tovey, M, 1997, Styling and design: intuition and Development.
[14] Lecture by Eberle D., 15 June 2009, RASA DK.
Figure 1: Example of graphical translation of results from a daylight and energy-balance simulation program. Seventh semester student project. The representations explain the changes caused by altering the design of the balcony.
Design of an Integrated Process Chain to Manufacture Titanium Components with Micro Features

A. V. Vargas¹, U. Berger¹, H. N. Hansen²

¹ Chair of Automation Technology, Brandenburg University of Technology Cottbus
Siemens-Halske-Ring 14, 03046 Cottbus, Germany
² Department of Mechanical Engineering, Technical University of Denmark
Bygning 427S, 2800 Kgs. Lyngby, Denmark
vargaalg@tu-cottbus.de, ulrich.berger@tu-cottbus.de, hnha@mek.dtu.dk

Abstract

More and more micro components and micro features have been introduced into diverse types of industrial products and market sectors. Micro features are used in the aerospace and automotive industry and for medical and biomedical applications. There are several micro manufacturing processes but they are quite limited to machine 3D free-form micro structures in a wide range of materials, mainly hard metals. Thus, the integration of processes into a continuous process chain is required. In this paper, the design of an integrated process chain to support the manufacturing of micro features on titanium components using a formal approach is presented.

Keywords:
Process chain design, axiomatic system design, technology data catalogue

1 INTRODUCTION

In the last decade, the use of micro products has significantly increased [1]. A market analysis for the years 2004-2009 shows a clear indication of the scope of the economic sectors that are directly affected by micro technologies with their current investment trends. Investments are expected to keep on growing rapidly with the potential of the market reaching 20 billion € in 2010, with a growth rate of around 20% in micro technology products [2].

Micro components have been introduced into diverse types of industrial products in diverse industries and applications, e.g. the automotive and aerospace industry, and for medical and biomedical applications [1, 3-4].

Micro holes, the most basic micro features of micromachining, with diameters between 200-300µm are commonly found in several products; for example, mechanically drilled in printed circuit boards. Fuel injection nozzles are also mechanically drilled or machined by electric discharge machining (EDM). Holes with a diameter smaller than 100µm have become the new machining challenge [3].

Figure 1: Instrumentation holes on blisk [5].

In the aerospace industry, micro holes with the functionality of decreasing the temperature of the component surface are designed and machined on combustors and in high pressure turbines. Recently, micro features were also incorporated on stator vanes and rotor blades. This is the most significant application due to the fact that in aero engine manufacturing blades are the only components manufactured in relatively large quantities [4]. Moreover, micro holes are found on blade trailing edges, as micro instrumentation holes on all parts of the engine (see Figure 1), and as oil whizzer holes on compressor discs [5]. There are several technologies to drill cooling holes but their biggest drawback is their achievable machining speed. EDM and electro chemical drilling (ECD) have been the main manufacturing methods because they can make multiple holes at the same time using multiple electrodes [6], all at the same angle with high levels of roundness and taper [7]. ECD and EDM have typical drilling speeds of 1-10mm/min [6].

Although they are slow processes and they can not be used if the blades and vanes are coated with a thermal barrier coating, they are used when the aspect ratio of a hole is so large that laser drilling will not produce the quality level desired [4]. Electron Beam Drilling (EBD) is a fast process, but it is not an attractive technology because it needs a vacuum chamber and is more expensive than a YAG laser [6-7].

According to [1], the integration of processes into a continuous process chain is necessary for the manufacturing of new micro components and micro features; and therefore, the usage of systematic design methods is of relevance. Moreover, the importance of process chains thrust the need to characterise processes not only for their individual capability but also for their suitability for the integration into process chains to satisfy specific functional and technical requirements [8].

Toolboxes are highlighted as solutions to link the design of products and design of processes and process chains to fulfill specific functional and technical requirements [8].

An integrated methodology to generate process chains for the manufacturing of micro features on titanium components was developed using a formal approach [9].
A process chain developer, based on the axiomatic system design, was created to define timed sequence sub-processes. On the other hand, in order to retrieve, to characterise and to share relevant engineering data generated during the manufacturing process, a technology data catalogue (TDC) was developed. While the process chain developer provides the structure of the process chain, the TDC provides manufacturing process scenarios with all of the data needed for the manufacturing of the micro features on the titanium component. A semantic net, part of the TDC, will be utilised as an effective tool to share the final process chain and manufacturing information.

2 AXIOMATIC SYSTEM DESIGN FRAMEWORK

Axiomatic design is a methodology created by N.P. Suh that endows designers with the scientific basis for the design of engineering systems. Additionally, axiomatic design enhances creativity, minimises the iterative trial and error process, and determines the best design, among other advantages. Suh defined design as an activity that involves interplay between what we want to achieve and how we choose to satisfy the need (the what) and four domains that delineate four different design activities [10]: the customer domain, the functional domain, the physical domain, and the process domain (see Figure 2). The customer domain is characterized by attributes (CAs) or the needs that the customer seeks in a product, or a process or a system. In the functional domain the needs are defined based on functional requirements (FRs) and constraints (Cs). In the physical domain the design parameter (DPs) that satisfy the specified FRs are described. Finally, in the process domain manufacturing process variables (PVs) are characterized and a process based on the PVs that can produce the DPs is developed. Constraints (Cs) provide the limits on the acceptable design. The difference between Cs and FRs is that Cs do not have to be independent as the FRs do. Axiomatic design starts by identifying and defining the customer attributes or needs, and translating them into functional requirements. That involves a mapping process from the customer domain to the functional domain. Then a mapping process between functional domain and the physical domain follows to satisfy the customer needs. This process is also called a zigzagging method. This method allows hierarchies for FRs, DPs, and PVs to be created in each domain.

![Figure 2: Schematic representation of domains](image)

During the mapping process, many possible DPs can be found; the key DPs are selected for the design according to two design axioms: the independence axiom and the information axiom. The mapping process can be expressed mathematically in terms of characteristic vectors. A vector of FRs can be related to a vector of DPs according to the following equation:

\[ \{FR\} = [AD]\{DP\} \tag{1} \]

The elements of the matrix are represented with a “0” if there is no effect and with an “X” if there is an effect, and later on are substituted by other values. Moreover, when all Aij are equal to zero except those where i=j then the design matrix is defined as diagonal and the design is called an uncoupled design, where each of the FRs can be satisfied independently by means of one DP. And when the upper triangular elements are equal to zero then the design matrix is defined as a lower triangular matrix and the design is called a decoupled design, where the independence of FRs can be ensured only if the DPs are defined in the right sequence. In any other case, the design matrix is defined as a full matrix and the design is called coupled, which is the most undesired design.

\[
\begin{align*}
\text{FR1} & = \begin{bmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{bmatrix} \\
\text{FR2} & = \begin{bmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{bmatrix} \\
\text{FR3} & = \begin{bmatrix} X & X & X \\ X & X & X \\ X & X & X \end{bmatrix}
\end{align*}
\]

Equation (2) shows a diagonal matrix/ uncoupled design, equation (3) shows a triangular matrix/decoupled design, and equation (4) shows a full matrix/ coupled design.

The independence axiom declares that the independence of the functional requirements must be maintained; the design solution must satisfy each FR without influencing the other FRs. The information axiom defines that the better design is the one with the minimum information content to fulfill the design. The information content is calculated as follows.

The information content \( I_i \) for a given FR can be defined in terms of probability \( P_i \) of satisfying FRi

\[
I_i = \log_2 \frac{1}{P_i} = -\log_2 P_i \tag{5}
\]

where the information is given in units of bits.

As defined in [10], the logarithmic function is chosen so that the information content will be additive in the case of many functional requirements that must be satisfied at the same time. Either the logarithm based on 2 or the natural logarithm may be used. And the total information content or information content of the system (Isys) is calculated as follows.

\[
\text{Isys} = \sum_{i=1}^{m} I_i = -\sum_{i=1}^{m} \log_2 P_i \tag{6}
\]

where \( P(m) \) is the joint probability that all m FRs are satisfied. When all FRs are statistically independent, as in the case of an uncoupled design, \( P(m) \) is defined as follows.

\[
P(m) = \prod_{i=1}^{m} P_i \tag{7}
\]

Axiomatic design presents many advantages compared with other methodologies such as TRIZ, the Taguchi method, engineering design and systematic design. It provides a good structural foundation based on functional analysis and information minimisation, which lead to
robust design [10]. The information axiom provides the criterion for making the decision without the arbitrary weighting factors used in other decision making theories. As stated in [10], when the design range is specifically defined for all functional requirements, the design goal is fully specified, and any design that satisfies the functional requirements within their specified design ranges fully satisfies the design aim. Thus, there is no need for weighting factors.

3 INTEGRATED DESIGN METHODOLOGY

The integrated methodology proposed for the design of process chains for the manufacturing of titanium components is shown in Figure 3 [9]. The inputs in the schematic representation are the requested component design, generally provided as 3D-CAD design, and the respective workpiece material, e.g. Ti6Al4V alloy. The output is the optimised machining information. This information is represented in a net in order to guarantee that it is easy to understand and communicate, and to navigate through its information dynamically. The machining information includes the adequate technology, optimal process resources (machine tool, tools and fixtures), and optimised kinematic and cutting parameters to manufacture the component. With them, the tool path and the NC programme can be further generated by the CAM operator. The process chain developer, based on the axiomatic system design, generates timed sequence sub-processes with the aim of enhancing productivity in the manufacturing domain. It translates customer needs into functional requirements which are fulfilled by design requirements. In an acceptable process chain design, the design parameters can be adjusted to satisfy their corresponding functional requirements without affecting other functional requirements. The timed sequence sub-processes provide the necessary guidelines for modelling and planning the manufacturing process. The technology data catalogue is a manufacturing knowledge repository where information (which corresponds to design parameters from the process chain) is characterised and shared in a dynamic way. Moreover, the catalogue includes ontology and a semantic net, for retrieving, structuring, processing and communicating optimal engineering data, such as:

- Properties and limitations of manufacturing processes,
- Properties and limitations of machine tools,
- Material characteristics,
- Geometric data of components,
- Machining performance and productivity parameters.

The development of the ontology assures that the knowledge is extracted (even that which otherwise would remain hidden as implicit knowledge, e.g. operational knowledge) from different sources, and stored in a single database to support the manufacturing process planning. The process chain developer communicates with the technology data catalogue to gain design parameters for the machine setup such as tools, fixtures, coolants and lubricants and the relevant process parameters, such as feed rates and cutting speeds for specific features and machining operations. These process parameters were previously extracted from successful used cases of machined parts and stored in the technology data catalogue after being characterised. The main advantage of the integrated methodology is that precise information exchange is enabled and consequently the time for process planning, machine setup time and cycle time are minimised. The process chain designer and the technology data catalogue will be detailed in the following sections.

3.1 Process chain developer

The process chain developer is used to generate a process chain for the machining of a Ti6Al4V alloy with dynamic cutting conditions; this means higher cutting speed and higher axial tool engagement. As an example, the process chain development for the manufacturing of blisks, where large axial depth of cuts result in chatter vibrations and increase in dynamic cutting forces, is presented.

The process chain developer defines the first level functional requirement (FR) and the respective design parameter (DP) as follows:

- FR1: Enhance productivity
- DP1: Reduction of manufacturing costs

The target cost for manufacturing the required component is identified. Then, it is translated into the first constraints of the design. Constraints provide bounds on the acceptable design solutions and they do not have to be independent as the FRs [10].

C:= Manufacturing target cost

Blisk manufacturing costs can be split into three main categories: the material costs, the airfoiling process cost, and other manufacturing and quality assurance costs. Thus, the process chain developer determines the next decomposition as follows:

- FR1:= Minimise airfoiling process costs
- DP1:= Integrated process chain design
- FR2:= Minimise quality assurance costs
- DP2:= Steady process to target design specifications
- FR3:= Minimise material costs
- DP3:= Optimum material utilization

![Figure 3: Schematic representation of the methodology.](image-url)
The design equation representing the interaction between the FRs and DPs is as follows.

\[
\begin{bmatrix}
FR_{11} \\
FR_{12} \\
FR_{13}
\end{bmatrix} = 
\begin{bmatrix}
X & 0 & 0 \\
X & X & 0 \\
X & X & X
\end{bmatrix} 
\begin{bmatrix}
DP_{11} \\
DP_{12} \\
DP_{13}
\end{bmatrix}
\]  

(7)

where \([A]\) is a triangular matrix, thus a decoupled design.

The integrated process chain is designed to reduce the manufacturing process costs, to achieve a steady process and optimise the use of material. On the other hand, a process without readjustments to the target design specifications achieves the minimisation of material wasted. Since the manufacturing of the disk involves processes where a sufficient amount of experience has been obtained through the manufacturing of the conventional blade assembled disk (e.g. conventional cutting, surface compactness and finishing), the process chain is specifically designed for minimising blisk airfoiling process costs. Three machining processes may be used for manufacturing airfoils: milling, linear friction welding (LFW) or electrochemical machining (ECM).

The selection of the process which better fulfils the design requirements can be selected by applying the second axiom, the information axiom. The following are the most important functional requirements that the machining process must initially satisfy: dimensional accuracy, surface quality and machining productivity.

- FR1 = Machine blisk with (number) airfoils
- FR2 = Machine blisk of diameter... mm
- FR3 = Accomplish airfoil thickness at the inner annulus in the range of ... mm and at the blade tip of ... mm
- FR4 = Obtain a surface roughness in the range of... \(\mu m\)
- FR5 = Achieve the minimum machining time per airfoil

While the above mentioned FRs represent the design range, the current capabilities of the processes represent the system design. Once the design and system ranges are defined, the areas with a common range between each value from the design and each of the values from the system are identified and the probabilities of satisfying each functional requirement are analysed. Then, the information content for each FR is calculated using the equation (5). The information content of the system is calculated using equation (6).

The information axiom states that the design that has the smallest information content (I) is the best design, as it requires the least amount of information to achieve the design goals.

The process decomposition will be further focused on the integrated process chain for milling, since this process is utilised before the electro-chemical machining of the airfoil for rough machining to remove material up to an envelope of 2mm to final shape; and after the linear friction welding of the airfoil to remove the clamp shoulder. Thus it is considered as a representative process to demonstrate the effectiveness of the process chain developer.

The process chain design must integrate a strategy, tools and machine tools to make possible the reduction of manufacturing throughput time and thus, costs. The process chain developer defines \(FR_i/DP_j\) as follows:

- FR11 = Retrieve surface and dimensional requirements
- DP11 = Characterised feature from TDC
- FR12 = Generate the machining strategy
- DP12 = Machining range and operation definition
- FR13 = Determine machine tool performance and limitations
- DP13 = Machine tool selection from TDC and process constraints determination
- FR14 = Determine type of milling cutter for machining strategy
- DP14 = Milling cutter type selection
- FR15 = Maintain cutting zone at stable temperature and lubricated to extend tool life
- DP15 = Cooling lubricant selection
- FR16 = Assure that the workpiece is firmly held in place for the machining
- DP16 = Fixture selection
- FR17 = Maxime surface finish quality while maximising machining productivity
- DP17 = Kinematic and cutting parameters

In accordance with the following constraints:

\[C_0 = \text{Manufacturing throughput time}\]

The design equation representing the interaction between the FRs and DPs is as follows.

\[
\begin{bmatrix}
FR_{11} \\
FR_{12} \\
FR_{13} \\
FR_{14} \\
FR_{15} \\
FR_{16} \\
FR_{17}
\end{bmatrix} = 
\begin{bmatrix}
X & 0 & 0 & 0 & 0 & 0 & 0 \\
X & X & 0 & 0 & 0 & 0 & 0 \\
X & X & X & X & 0 & 0 & 0 \\
X & X & X & X & 0 & 0 & 0 \\
X & X & X & X & 0 & 0 & 0 \\
X & X & X & X & X & 0 & 0 \\
X & X & X & X & X & X & 0
\end{bmatrix} 
\begin{bmatrix}
DP_{11} \\
DP_{12} \\
DP_{13} \\
DP_{14} \\
DP_{15} \\
DP_{16} \\
DP_{17}
\end{bmatrix}
\]  

(8)

where \([A]\) is a triangular matrix, thus it is a decoupled design.

From the 3D-CAD design, the manufacturing features are manually retrieved and characterised. After their characterisation, they are stored in the technology data catalogue. The characterised data is further utilised in the design of the process chain in the form of constraints, where workpiece material, raw material dimensions, part dimensional requirements (shape and size), tolerances, and surface requirements are specified.

Once manufacturing features are identified, the machining strategy must be generated. Here, the machining range is defined as rough machining, or semi-finishing or finishing. Then, a machining operation is identified which will assure the generation of the required part shape by removing material from the workpiece in the form of chips. The machining operation can be classified in several categories, e.g. face milling, end milling, slotting, form milling, micro milling, etc. The machining strategy provides the general information for the design of the manufacturing process and the definition of process resources (tool, machine tool, fixtures) and process parameters (speed and feed). On the other hand, the machining strategy is important for determining the tool path, and the tool responsible for tool time, geometry accuracy and surface finishing.

The process chain developer selects a milling machine in accordance with the part dimensional requirements, and determines the machine performance by retrieving data characterised in the TDC, e.g. positioning accuracy (resolution and repeatability), maximum spindle speed, power, machine tilt angles (rotational axes limitation), kinematics limitations in X, Y, and Z axes, physical ability to run simultaneously 5 axis tool paths, and tool holder.
The characterised machine data need to be taken into consideration to define the process resources (tool) and process parameters (spindle speed, power required for the cutting operation, etc.), as well as, to generate the tool path (collision free with respect to the cutter and holder) in the CAM system. Thus, these data are further handled in the form of constraints.

The milling cutter type must be selected in accordance with the machining range and operation, workpiece material, initial block dimensions, the characteristics of the machine tool (tool holder, max. spindle speed). The manufacturing through put time (where a short through put time may required faster cutting speeds, thus a better tool performance) and the manufacturing target cost.

In order to select the appropriate milling cutter geometry, the fact that a higher axial depth of cut generates chatter vibration must be taken in account. Thus, the process chain developer defines FR114/DP114 as follows.

\[ FR_{1141} = \text{Assure that the tool is strong and rigid enough to resist cutting forces} \]
\[ DP_{1141} = \text{Cutting tool material for machining high strength part} \]
\[ FR_{1142} = \text{Minimise thermal load on the tool cutting edge and welding between tool and chip (BUE formations) to increase machining accuracy} \]
\[ DP_{1142} = \text{Tool coating or insert type which can resist heat and adhesion} \]
\[ FR_{1143} = \text{Assure dimensional requirements can be achieved} \]
\[ DP_{1143} = \text{Tool diameter and length (L/D ratio)} \]
\[ FR_{1144} = \text{Assure a uniform chip formation/chip breakability} \]
\[ DP_{1144} = \text{Number of teeth and tool helix angle} \]
\[ FR_{1145} = \text{Minimise self exited chatter vibration} \]
\[ DP_{1145} = \text{Irregular tooth spacing (variable pitch angle)} \]

In accordance with the following constraints:

\[ C1 = \text{Manufacturing target cost} \]
\[ C2 = \text{Manufacturing throughput time} \]
\[ C3 = \text{Workpiece material} \]
\[ C4 = \text{Dimensions of initial block (raw material feature)} \]
\[ C5 = \text{Tool holder} \]

The design equation representing the interaction between the FRs and DPs is as follows.

\[
\begin{bmatrix}
FR_{1141} \\
FR_{1142} \\
FR_{1143} \\
FR_{1144} \\
FR_{1145}
\end{bmatrix} = \begin{bmatrix}
X & 0 & 0 & 0 & X \\
X & X & 0 & 0 & 0 \\
0 & 0 & X & 0 & 0 \\
0 & 0 & 0 & X & 0 \\
0 & 0 & 0 & 0 & X
\end{bmatrix} \begin{bmatrix}
DP_{1141} \\
DP_{1142} \\
DP_{1143} \\
DP_{1144} \\
DP_{1145}
\end{bmatrix}
\]

(9)

where \([A]\) is a triangular matrix, thus it is a decoupled design.

From the design equations one can realise that the material selected for the tool may also minimise the thermal load on the cutting tool edge and the welding between the tool and chip. Thus, an extra tool coating may be not required.

The fixture must be selected in accordance with the dimensions of initial block (raw material feature). The fixture must anchor the workpiece effectively for the machining operation to avoid vibration due to the generated cutting forces.

The maximisation of the productivity and the surface finish is possible by suppressing chatter vibration and minimising airfoil deflections during the machining process. The tool geometry previously selected is influencing the minimisation of chatter, but the complete suppression will be achieved by selecting the correct spindle speed value. Thus, \( FR_{1171}/DP_{117} \) is decomposed as follows.

\[ FR_{1171} = \text{Minimise chatter intensity} \]
\[ DP_{1171} = \text{Spindle speed definition (S)} \]
\[ FR_{1172} = \text{Assure a constant cutting force to minimise airfoil deflection} \]
\[ DP_{1172} = \text{Feed rate definition (V/\text{f} \times S)} \]

The design equation representing the interaction between the FRs and DPs is as follows.

\[
\begin{bmatrix}
FR_{1171} \\
FR_{1172}
\end{bmatrix} = \begin{bmatrix}
X & 0 \\
0 & X
\end{bmatrix} \begin{bmatrix}
DP_{1171} \\
DP_{1172}
\end{bmatrix}
\]

(10)

where \([A]\) is a diagonal matrix, thus it is a uncoupled design.

The spindle speed can be defined based on a stability lobes diagram with variable depth of cut. Once the process is stabilized, milling force modelling can be used to determine the feed rate corresponding to a reference force value (maximum value). The selected speed for the highest depth of cut and the optimised feed rate will determine the maximum material removal rate, while avoiding the effects of chatter vibration like poor surface finish, and dimensional errors on the part.

The tool path must be generated to fulfil dimensional and shape requirements which include airfoil thickness and milling crest geometry tolerances; and in accordance with the machine limitations and the milling cutter type and geometry. On the other hand, the tool path must be determined considering the kinematics of the tool exits from the workpiece to avoid burr formations by following these criteria: avoiding exits of inserts, controlling exit order sequence and sequencing process steps to create any burrs on a less significant edge, etc.

\( FR_{119}/DP_{12} \) is specified as follows:

\[ FR_{119} = \text{Minimise quality assurance cost} \]
\[ DP_{12} = \text{Steady process to target design specifications} \]

The manufacturing process stability has influence on the cost and delivery time of compressors and turbines. Thus, stability of the machining process, where a poor surface quality may indicate tool wear, burr formations, chatter vibrations, etc. The integrated process chain determines the required process resources and parameters to assure the surface quality and minimise further process adjustments, thus assuring that the part is delivered on time and meets the design specifications.

Finally, \( FR_{13}/DP_{13} \) is defined as follows:

\[ FR_{13} = \text{Minimise material costs} \]
\[ DP_{13} = \text{Optimal material utilisation} \]

The characterisation of titanium alloys allows the material data quality and machining productivity to be increased. The material data will be characterised in the technology data catalogue (TDC) to facilitate their retrieval when designing the integrated process chain. The structuring of knowledge and precise knowledge exchange is enhancing effective optimization of a process chain for manufacturing Ti6Al4V components.

The first timed sequence sub-processes for machining the macro component have been designed. Further sub-processes for machining micro features can be designed in a similar way; since tool wear and machining accuracy were already defined as relevant functional requirements. On the other hand, the material data are identified as design parameters, thus the fact that the material for
micro machining is heterogeneous will be taken in account during the design of the process chain. The design of a process chain for micro features is illustrated in the case study presented in section 4.

3.2 Technology data catalogue
It is proposed that the design and development of a technology data catalogue (TDC) is a key component of the integrated methodology. The main aims of the TDC are gathering, retrieving, structuring, processing and sharing relevant engineering knowledge in an intelligent way. The TDC will also provide structured information about the best practice settings of the manufacturing system. Data sources are defined and a suitable knowledge ontology is created in order to store the relevant implicit and explicit knowledge generated at the different levels in the manufacturing domain. The definition and use of ontologies will eliminate conceptual and terminological confusion and lead to a shared understanding [11], solving problems of semantics; while the use of semantic nets will ensure the effective sharing of the final process chain and manufacturing information. Furthermore, the development of semantic nets provides external data access to the TDC. Thus, the use of semantic nets will ensure the effective sharing of information through formal procedures. Moreover, the advantages of the TDC are:

- Facilitate the extraction of information selectively about process resources (machine tool, tools, etc.), process parameters (feeds and speeds) and process constraints (spindle power);
- Improve and increase information by recording relevant operational knowledge avoiding that this knowledge remains as personal hidden knowledge; and by documenting information from successful used cases.
- Ensure consistency of the information and avoid redundancies.

The technology data catalogue, defined as a knowledge platform, consists of an ontological system and a database. In the ontological system, the manufacturing ontology and the semantic nets are built. The ontology development follows the design criteria defined by [12], and the design process suggested by [13]. The semantic nets provide external data access to the TDC. Thus, the fast heterogeneous information access for handling process requirements is enabled. Moreover, the development of semantic nets further the use of semantic nets will ensure the effective sharing of information through formal procedures. The shared terms collected in the database include:

- Manufacturing operations (applicability of parameter settings for manufacturing setups and relevant measuring parameters),
- Machining processes or technologies,
- Process resources (material properties, machine tools, tools, fixtures, etc.),
- Process kinematics (tool access direction),
- Process constraints,
- Manufacturing productivity parameters,
- Manufacturing performance parameters.

As a pull mechanism to fulfil the TDC, interviews of shop floor operators will be part of a formal procedure to avoid the possibility that operational knowledge will remain personal hidden knowledge. Another way of filling up the TDC is by extracting and modelling the relevant information from successful used cases. To enhance the applicability and dissemination of the gathered information, the exchange of data from TDC via web portal is functional.

4 CASE STUDY: MANUFACTURING OF MICRO FEATURES
Micro machining experiments on a T6A4HV alloy were conducted. The machining information obtained from the experiments was characterised in the TDC. Then, the characterised data was retrieved and used for the determination of the suitable technology to fulfil multiple functional requirements simultaneously. Three experiments were carried out on the T6A4HV alloy utilising three different technologies: drilling, laser, and electric discharge machining (EDM). The necessary machining operations are defined to generate micro holes of three different diameters Ø=1, 0.5 and 0.3mm, and depth range from ≈ 1 to 10mm. Micro holes were chosen as these micro features have applicability for complex components manufacturing in several industries, e.g. the automotive and aero space industries. The diameter accuracy, the roughness of the hole-walls and the roughness of the part surfaces were measured. From the measurements, it can be highlighted that the best diameter accuracy was obtained from the laser drilling experiments; dimensional deviation in the EDMed holes was mainly produced by the spark gap, while deviation in the drilled holes was generated by the tool diameter. Laser was the fastest technology compared with milling and EDM, as it may machine the three holes diameters at the same time; also, less non productive time was observed due to no tool requirements, and thus no tool changes. In contrast, drilling requires the use of four different tools to pre-drill and drill three different diameters. Tool wear of 7-8% was observed when machining the holes with a diameter of Ø=1mm. Moreover, big aspect ratios (e.g. aspect ratio=16) had higher cutter bending. This bending causes an eccentric movement of the tool and tool breakage. However, this technology provided better surface quality on the walls of the holes (e.g., Ra pin 1mm= 0.231µm) compared with the surface quality obtained from EDM (e.g., Ra pin 1mm= 0.460µm). This may be of crucial importance when drilling for instance cooling holes on turbine blades where the roughness may influence the air flow. On the other hand, drilling and EDM machining do not affect the surface around the machined holes as lasers do; the surface roughness increased from 0.2 to 0.6µm. This requires extra processes where a kind of protection (e.g. wax) is applied around the zone where holes will be machined, and then it is cleaned after the machining. In summary, the advantages and disadvantages of the three technologies have been highlighted; while laser experiments provide better accuracy, drilling experiments presented better surface quality on the hole-walls. When the machining information generated from the three technologies is characterised in the technology data catalogue, different scenarios are created. If only one functional requirement was defined by the process chain developer, e.g. maximising dimensional accuracy, then the machining information from the laser process must be retrieved from the TDC. If the functional requirement was defined as the minimisation of the surface roughness on the hole-walls, then the machining information from drilling experiment must be retrieved from the TDC.

But if two functional requirements must be simultaneously fulfilled by the machining process, e.g. dimensional accuracy and surface quality when machining micro holes of different aspect ratios, then a tool to facilitate the selection of the adequate machining process is required.
Such a tool is provided by the process chain developer, through the information axiom.
Assume that the functional requirements for the machining of features of Ø= 1mm were defined as follows.
FR1= Diameter deviation from nominal<0.01mm
FR2= Hole-wall roughness<0.30microns
FR3= Change on surface roughness around holes<0.01microns
While these functional requirements specify the design range, the capabilities of the technologies summarized in Table 1, represent the system design.
Once the design and system ranges are defined, the areas with a common range between each value from the design and each of the values from the system are identified, and the probabilities of satisfying each functional requirement are analysed. Then, the information content for each FR is calculated using the equation (5), as shown in Tables 2-4.
The information content of the system is calculated using equation (6), as shown in Table 5.
It can be concluded from tables 2-5, that the laser machining satisfies the first functional requirement, achievement of minimum diameter deviation from nominal, when the diameter of the hole is 1mm. Drilling satisfies adequately the second functional requirement, achievement of good surface quality on hole-wall, when the diameter of the hole is 1mm.

<table>
<thead>
<tr>
<th>Machining technology</th>
<th>Diameter deviation (mm)</th>
<th>Hole –wall roughness (µm)</th>
<th>Surface roughness (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drilling</td>
<td>0.013</td>
<td>0.23</td>
<td>0.011</td>
</tr>
<tr>
<td>Laser</td>
<td>0</td>
<td>No information available (NA)</td>
<td>0.245</td>
</tr>
<tr>
<td>EDM</td>
<td>0.03</td>
<td>0.46</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Table 1: Technology capabilities obtained from experiments for micro holes of Ø= 1mm.

<table>
<thead>
<tr>
<th>Option</th>
<th>Design range</th>
<th>System range</th>
<th>Area with common range</th>
<th>Prob. of satisfying FR$_1$ (P$_1$)</th>
<th>Information content (I$_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drilling</td>
<td>&lt;0.01</td>
<td>0.013</td>
<td>0.01</td>
<td>0.77</td>
<td>0.38</td>
</tr>
<tr>
<td>Laser</td>
<td>&lt;0.01</td>
<td>0</td>
<td>0.01</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>EDM</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.33</td>
<td>1.60</td>
</tr>
</tbody>
</table>

Table 2: Calculation of the information content for FR$_1$.

<table>
<thead>
<tr>
<th>Option</th>
<th>Design range</th>
<th>System range</th>
<th>Area with common range</th>
<th>Prob. of satisfying FR$_2$ (P$_2$)</th>
<th>Information content (I$_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drilling</td>
<td>&lt;0.3</td>
<td>0.23</td>
<td>0.23</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Laser</td>
<td>&lt;0.3</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>*Infinite</td>
</tr>
<tr>
<td>EDM</td>
<td>&lt;0.3</td>
<td>0.46</td>
<td>0.3</td>
<td>0.65</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Table 3: Calculation of the information content for FR$_2$.

<table>
<thead>
<tr>
<th>Option</th>
<th>Design range</th>
<th>System range</th>
<th>Area with common range</th>
<th>Prob. of satisfying FR$_3$ (P$_3$)</th>
<th>Information content (I$_3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drilling</td>
<td>&lt;0.01</td>
<td>0.011</td>
<td>0.01</td>
<td>0.91</td>
<td>0.14</td>
</tr>
<tr>
<td>Laser</td>
<td>&lt;0.01</td>
<td>0.245</td>
<td>0.01</td>
<td>0.04</td>
<td>4.61</td>
</tr>
<tr>
<td>EDM</td>
<td>&lt;0.01</td>
<td>0.008</td>
<td>0.008</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Calculation of the information content for FR$_3$.

<table>
<thead>
<tr>
<th>Option</th>
<th>I$_1$ (bits)</th>
<th>I$_2$ (bits)</th>
<th>I$_3$ (bits)</th>
<th>I$_{sys}$ (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drilling</td>
<td>0.38</td>
<td>0</td>
<td>0.14</td>
<td>0.52</td>
</tr>
<tr>
<td>Laser</td>
<td>0</td>
<td>Infinite</td>
<td>4.61</td>
<td>*Infinite</td>
</tr>
<tr>
<td>EDM</td>
<td>1.60</td>
<td>0.62</td>
<td>0</td>
<td>2.22</td>
</tr>
</tbody>
</table>

Table 5: Calculation of the total information content.
*Infinite= it cannot satisfy FR / The design range and the system range do not overlap.
EDM satisfies the third functional requirement, minimisation of the change in surface roughness around holes before and after the machining. But, drilling is selected as the best technology since it has the highest probability for satisfying all functional requirements simultaneously, with the least amount of information. The semantic net retrieved from the TDC for drilling the micro holes is shown in Figure 3. It provides structured information about the best practice settings of the manufacturing system (tool, machine, parameters) to be used by the machine operator at the workshop.

5 SUMMARY
The integrated methodology, proposed in the previous sections, was successfully validated. It was demonstrated that the integration of the process chain developer and the TDC assures that the adequate technology is used for a specific micro machining operation. Micro features were machined on a Ti6Al4V alloy utilizing three different technologies: drilling, laser drilling, and electric discharge machining (EDM). The necessary machining operations were defined to generate micro holes of three different diameters $\Phi=1, 0.5$ and $0.3\text{mm}$, and depth range from $\approx 1$ to $10\text{mm}$. It was observed in micro machining that the tools have greater influence on the features accuracy than in conventional size machining. The three technologies present advantages and disadvantages; e.g., while laser experiments provide better accuracy, drilling experiments presented better surface quality. Different technologies with different parameters fulfill different requirements. If these parameters are characterised in the technology data catalogue many scenarios will be obtained which will not facilitate making the decision of which technology is the best for the process chain. Therefore, the technology data catalogue requires an approach to define systematically the process requirements (design range) and the actual process performance (system range) and find out methodically the best process without arbitrary weighting factors. The process chain developer provides such a decision tool, by applying the information axiom.

6 REFERENCES
Abstract

Product-Service System (PSS) provides an enhanced view about the functional aspects of value creation. Manufacturers are increasingly transforming their whole systems including processes in order to support the provision of PSS throughout its lifecycle. Substantial changes in terms of methods, tools and software is required to develop PSS in a structured manner due to its peculiar properties. This research aims to develop a software platform which will aid the development process of PSS. In this paper, merits and limitations of Service CAD (a non-commercial software) as a PSS development tool are discussed. A demonstration of PSS development is done through the use of a laser systems case study. In so doing, necessary modules which are required to enhance the Service CAD are identified and one of the modules is programmed as an enhancement to the software.

Keywords:
Product-Service System, Design, Software

1 INTRODUCTION

The functional aspects of value creation are gaining considerable importance in today's competitive business to business environment. The emphasis is on the 'sale of use' rather than the 'sale of product'. Lower costs and long term relationships with the customers are considered as the primary factors for sustained business. In this context, an integrated product and service offering that delivers value in use provides this competitive advantage which has been commonly termed as Product-Service System (PSS). Goedkoop et al. [1] state that a product service-system is a system of products, services, networks of players and supporting infrastructure that continuously strives to be competitive, satisfy customer needs and have a lower environmental impact than traditional business models. Meier et al. [2] define an 'Industrial Product Service Systems' (IPS²), as an integrated product and service offering that delivers value in industrial applications. It has also been defined as a 'self-learning' system, one of whose goals is continual improvement [3]. A major perspective of this concept is to consider the system as a whole, rather than just physical products [4]. It aims to provide required customer value through reduced cost, optimized resources and sustained production. The major merits for the manufacturer of this approach are increased revenue, prolonged and strategic relationship with the customer and product/service improvements based on the improved understanding of customer usage and requirements.

Even though PSS provides substantial benefits to the stakeholders involved in the business, generally manufacturers are holding risks which were previously held by customers. Therefore PSS needs to be designed carefully considering all the possible scenarios to avoid pitfalls. Also the manufacturer's core competences are moving away from manufacturing to systems design and integration. The manufacturer needs substantial support in this process of designing PSS concepts in terms of methods, tools, techniques and software. Even though industries are offering PSS solutions for many years, the process of designing these concepts are ad hoc. In academia, various methodologies and software are proposed in the recent years. But these propositions are in a preliminary stage and require substantial validation to achieve significant benefits to the industries in structuring the process of PSS development. Besides, it is important to note that the aims and objectives of these proposed design methodologies and tools differ extensively.

This research aims to develop a software platform to assist industries in developing PSS concepts in well-ordered fashion. In this paper, features required in an ideal software platform are identified and discussed. To identify the ideal software platform, currently available non-commercial software and methodologies proposed in literature are reviewed and analysed accordingly. To facilitate an in-depth understanding of the required features, an example problem is formulated and applied to Service CAD software developed by Komoto and Tomiyama [5]. Explanation of PSS concepts development for laser systems case study has been detailed through Service CAD. These illustrations help to propose necessary enhancement modules required in the PSS concept development software. One of the modules is programmed in Service CAD and demonstrated.

The rest of the paper is structured as follows: Section 2 discusses and analyses various PSS design methodologies proposed in literature and software platform developed to support PSS concepts development, Section 3 details the example laser case study problem and application to Service CAD, Section 4 elaborates the enhancement modules required to augment the features of Service CAD and describes an implemented module and Section 5 concludes with conclusion and future work to be carried out.
2 LITERATURE REVIEW

In this section, PSS design methodologies proposed in literature are analysed and discussed. Emphasis has been provided to the methodologies which have been implemented through software to generate PSS concepts. By doing a literature review, we aim to understand the required features highlighted in the methodologies and incorporated in the software platforms. Studying across the methodologies helps to summarize the important features needed in the PSS concepts development. This section concludes with the gaps identified to improve the overall development process.

Komoto and Tomiyama [5] proposed Service CAD which supports designers to generate conceptual design of PSSs. They argue that in PSS design process, designers define activity to meet specified goal and quality, and define environment, under which the activity is realized. The elements used in Service CAD are service environment, provider, receiver, channel, contents, activity, and aim of the service receiver's activity, target, promised goal, realised service, quality and value added. They also developed ISCL (Integrating Service CAD with a life cycle simulator) which has functions to support quantitative and probabilistic PSS design using life cycle simulation. Figure 1 illustrates the architecture of Service CAD.

Shimomura et al. [6] aim to propose a method for designing service activity and product concurrently and collaboratively during the early phase of product design. To enable this designing, a unified representation scheme of human process and physical process in service activity is proposed. They expressed a state change of a customer by parameters called Receiver State Parameters (RSPs), which represent customer value. They propose a view model which handles functions and attributes to represent RSPs. They include three phases in service design process: identifying customer value, design of service contents and design of service activity. They also developed a method to evaluate these processes with Quality Function Deployment. Sakao et al. [7] developed a service model consisting of four sub-models: flow model, scope model, scenario model, and view model. They emphasize that the critical concept is not the function of a product, but rather the state change of the receiver. The state change can be fulfilled either by products or by service activities.

Sakao et al. [7] have implemented these models through prototype software named Service Explorer. Figure 2 illustrates the conceptual structure of Service Explorer. Welp et al. [8] argue that an Industrial PSS (IPS²) can be made up of any combination of product and service mix. They propose that the IPS² concept development is responsible for generating principle solutions that meet customer specific requirements. They present a model based approach to support an IPS² designer generating heterogeneous IPS² concept models in the early phase of IPS² development. They frame three planes for systematic conceptual development: IPS² function plane, IPS² object plane and IPS² process plane. The combination of all three modelling planes constitutes a heterogeneous IPS² concept model. Three different types of model elements are defined: system elements, disturbance elements and context elements. The combination of all types of model elements and their respective relations constitutes a heterogeneous IPS² concept model.

Apart from these three methodologies implemented in software, there are many methodologies proposed in literature for PSS development. The rest of this section discusses these proposed methodologies to highlight the features mentioned. Maussang et al. [9] consider the whole system and detail the physical objects and service units necessary to develop a successful PSS. They argue that this methodology can support the design of PSSs from the design of the architecture to the identification of physical object's (products) specifications. They used operational scenarios to elaborate the system description once main elements of the system (physical objects and service units) have been identified. External functional analysis is used to list the functions that the customer and the actors involved in the product lifecycle expect from the product without considering elements available to provide them. They argue that a specific external analysis must be carried out for each step of the product lifecycle (use, manufacture, maintenance, recycling, etc.). They characterize each function or constraint by criteria, level and allowance. They argue that this characterisation leads to the elaboration of specifications and product performance expected by the customer.

![Figure 1: Architecture of Service CAD [5].](image-url)
Aurich et al. [10] introduce a process for the systematic design of product related technical services based upon its modularization to link with corresponding product design processes. They propose an Object oriented technical service model to support the specification of technical services during their actual designing. The service components mentioned in the model are: the component description provides a general overview of a technical service both verbally and graphically; the component reference covers the description of the products, product components or users’ profiles addressed by the technical service along with the intended effects on them. The component function describes the measures for realizing the service functions and the component resources covers both physical and nonphysical resources necessary for realizing a service. They developed a systematic service design process to specify technical services according to the presented service model. They suggest that adapting already existing product design processes to account for the special characteristics of technical services would lead to maximum acceptance for application within the enterprise.

Alonso-Rasgado et al. [11] described a design process for Total Care Product (TCP) creation that integrates hardware and service support by providing a robust design methodology. The fast-track design process consists of a methodology that breaks down the iterative process between customer and supplier into a number of distinct stages necessary for the creation of the TCP. Fast-track design process is framed as: business ambitions of the client, potential business solutions, enhanced definition of the potential TCP, business case risk analysis of options, business case validation and evaluation of alternatives and contract. They consider two main variables of the system in simulations: (1) time taken to perform the service and (2) the quality and flow of information within the system.

Muller et al. [12] proposed a method for the development of PSS called PSS Layer method. This method intends to apply in early development phases comprising the clarification of the design task and the conceptual design phase. It defines a meta-model of nine main element classes for a PSS. The classes are: needs, values, deliverables, actors, lifecycle activities, core products, periphery, contract and finance. All classes are graphically layered to simplify the representation. They argue that this model provides the user to get a structured outline and the big picture of PSS idea or concept.

From the summary of literature, it has been noted that many factors are intertwined and influencing each other in the PSS concept development. These features need to be appropriately modelled to visualize and understand the impact of these factors to each other and to the overall performance as a whole. The important features necessary in developing PSS concepts that are stressed in the above discussion are summarized below:

- Understanding integration and impact between products and services are highlighted in most of the proposed methodologies.
- The support system provides modelling environment to describe activities, environmental, disturbances and contextual elements.
- Service CAD intends to support PSS designers through rule specification and generation.
- The modelling platforms stress the importance of specification through needed values and describe state change of the customers.
- Importance of reasoning the specified specifications has been emphasised.
- Modelling stakeholders and their relationships has been noted.
- Intention to model from the architecture level to specific product and service features is observed.
- Functional analyses to map between physical products and services are specified.
- Modelling through scenario development is emphasized.
- Graphical representation is emphasised in the modelling environment which includes qualitative and quantitative descriptions.
- Generation and evaluation of PSS offerings are mostly considered in the methodologies. Especially simulations through quantitative and probabilistic approaches are stressed.
- The approaches majorly based on system perspective.
Table 1: Comparison of key features to support representation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Defined ontology – Clear syntax and semantics</td>
<td>Ambiguous</td>
<td>Ambiguous</td>
<td>Clear (not specific to PSS)</td>
<td>Clear (not specific to PSS)</td>
</tr>
<tr>
<td>Check syntax and semantics</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Complexity handling through abstraction layers and managing subsystems interactions</td>
<td>Limited</td>
<td>Limited</td>
<td>In-zooming and out-zooming, Folding and unfolding, State expressing and suppressing</td>
<td>Nest the subprocesses</td>
</tr>
<tr>
<td>Creating views for different stakeholders</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Integrates function (what the system does or designed to do), behaviour (how the system changes over time) and structure (how the system is constructed) aspects of the system</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Some of the gaps identified from this literature summary are as follows:

- It is seen that the driving factors (risks and uncertainties) of PSS are not properly modelled.
- Only few methodologies stress the importance of co-creation between stakeholders and feedback loops between the steps involved in the process.
- Roles of the stakeholders involved in designing PSS offerings are not clearly defined in the methodologies. Especially capabilities of the stakeholders are not considered during design stage.
- In-depth solution description to represent PSS concepts is not elaborated. This solution description should inform contract formulation which will be the final step to establish the link between the stakeholders.
- Sensitivities of the resources involved in the PSS concepts are not appropriately represented.
- Even though evaluation approaches are proposed, detail cost modelling to evaluate PSS concepts is not described.
- Incorporation of lifecycle activities for adaptation of the PSS concepts are not modelled in the proposed methodologies.
- Representation techniques used in the modelling are not yet adequate to show the complexity involved in PSS.

These points illustrate that in order to develop a robust PSS concept considering the emerging risks and uncertainties, a sophisticated platform is needed which will incorporate various additional features and that platform is lacking in the extant literature. To understand some of the issues highlighted here, the next section demonstrates the application of a software platform through a laser case study example.

2.1 Representation comparison

Current software packages involve complex network structure to analyse the whole system. Representation should be structured and layer based approach can be used to avoid complexity. Co-creation between the customer, the manufacturer and the suppliers plays an important role in the design and provision of PSS. PSS models should represent stakeholders, products, services, business elements, work flow, business processes and interactions amongst them.

To aid the process of co-creation, the representation of PSS should be unambiguous, consistent, simple, complete, extensible, intuitive, easy to interpret and easy to maintain. In this section, comparison between various software (Service CAD [5], Service Explorer [6], OPCAT [13] and CAM [14]) which could be potentially used in PSS domain is carried out. OPCAT is commercially available software aims to turn concepts, cross-system processes, functions and requirements into a coherent, consistent blue print.

Cambridge Advanced Modeller (CAM) is a software tool for modelling and analysing the dependencies and flows in complex systems - such as products, processes and organisations. Even though OPCAT and CAM are not intended to support developing PSS concepts, they help to understand current state-of-the-art in representation techniques. Table 1 illustrates the comparison of key features to support representation in the PSS development. It highlights that commercially available software (OPCAT) addresses important issues in the representation. These state-of-the-art techniques need to be incorporated in the PSS development platforms. In specific, the challenges involved are: modelling the capabilities of stakeholders, the representation of qualitative factors, the integration of logical, operational, behavioural, temporal and physical processes involved in PSS. Certainly, the models should represent the value created by the developed offering and depict the risks associated with it.

3 AN APPLICATION ILLUSTRATION

To understand the merits and limitations of the software in detail, an example problem is framed and implemented in Service CAD [5]. The rest of this section is structured by describing the exemplar problem, implementation through the software and the results obtained.

3.1 Problem definition

Based on the spectrum of product and service mix in the offerings, PSS has been commonly classified into three types: Product Oriented, Use Oriented and Result Oriented. This classification is based on product
ownership and functionality, business models and product and service substitution. The example illustrated in this section falls in the type of Use Oriented PSS.

The problem description for PSS could be focused from different perspectives based on the respective stakeholders. In this paper, the focus is from the manufacturer who integrates and provides functionality of laser system as a complete unit to the laser job shop which then produces machined parts through laser cutting for the end product manufacturer. Figure 3 illustrates this value chain in the laser industries.

![Figure 3: Value chain in the laser industries.](image)

Rather than procuring the laser system from the manufacturer, currently the laser job shop intends to move to another business model of Use-Oriented type. In this functional oriented type, the laser job shop will specify the required level of Overall Equipment Effectiveness (OEE) for the laser system. Equation 1 illustrates the measurable elements used to calculate OEE.

\[
OEE = \text{Availability} \times \text{Performance} \times \text{Quality} \tag{1}
\]

The priorities for these measurable elements could be different for different laser job shops. In this example, the priorities are kept equal for all the elements. Equations 2-4 describe the measurement of individual element.

\[
\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} \tag{2}
\]

\[
\text{Performance} = \frac{\text{Working speed}}{\text{Designed speed}} \tag{3}
\]

\[
\text{Quality} = \frac{\text{Good units produced}}{\text{Total units produced}} \tag{4}
\]

The expectation level of OEE for the laser job shops is usually above 75%. The PSS offering requires the manufacturer to satisfy an OEE level well above 75% for five years in fixed cost. This is a challenge which needs to be addressed at the development stage of PSS. OEE is influenced by many factors involved in the manufacturing operations.

In this problem, it has been assumed that OEE is influenced by maintenance activities conducted by the manufacturer.

Planned, preventative and repair are the maintenance activities considered in this formulation. Planning these maintenance activities is an important task which influences total costs incurred by the manufacturer to provide the functional benefit to the laser job shop by maintaining OEE greater than 75%.

In the PSS design problem, for supplying 100 laser systems to the laser job shops through this business model, the manufacturer needs to identify answers for the following questions:

- How planned, preventative and repair maintenance activities influence OEE?
- How much maintenance man-days are required to support this business model?
- What will be the total maintenance costs incurred in this model?
- What will be the best fixed price / month for this business model?

Answering these questions through Service CAD [5] is illustrated in the next section. It should be noted that Service CAD is used as a design support tool rather than doing design itself.

### 3.2 Implementation in Service CAD

In this section, modelling of the illustrated problem in Service CAD is done and simulation carried out to answer the framed questions is described. As mentioned in Section 2, the first step carried out in modelling is to define activities to meet specified goal and quality, and define environment, under which the activity is realized. Figure 4 illustrates the modelling environment developed for the problem described in Service CAD. Emphasis has been provided to the manufacturer and the laser system. The activities considered are produce, sell, use, planned, preventative and repair maintenance. The environment is limited to the manufacturer, the laser job shop and laser system. The goals are represented through total costs and OEE for the manufacturer and the laser system respectively. In order to track the measurable elements in the laser system; availability, performance and quality are individually modelled through quantitative levels of 90, 70 and 50 respectively. Table 2 illustrates conditions and consequences for each activity to describe the proposed problem. The conditions specify the rules that are necessary for the particular activity to be executed. The consequences describe the impact of that particular activity on the environment after execution.

![Figure 4: Modelling environment developed for the problem in Service CAD [5].](image)
<table>
<thead>
<tr>
<th>Activity</th>
<th>Conditions</th>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Produce</td>
<td>[Manufacturer, production facility, product design]</td>
<td>[Generate(Laser system [100]), Availability, Performance, Quality: 100, Owner (Manufacturer, Laser system), set(placement(Laser system), Manufacturer]</td>
</tr>
<tr>
<td>Sell</td>
<td>[Manufacturer, Laser job shop, Laser system, Owner (Manufacturer, Laser system)]</td>
<td>[Owner(Manufacturer, Laser system), set(placement(Laser system), Laser job shop) [100]</td>
</tr>
<tr>
<td>Use</td>
<td>[Laser job shop, Laser system, state(Laser system, Operational)]</td>
<td>[Random(Availability: -9, Performance: -8, Quality: -8, set(OEE(Laser system)), Running time: +1]</td>
</tr>
<tr>
<td>Planned maintenance</td>
<td>[Manufacturer, Laser job shop, Laser system, Running time (Laser system, Once in six months)]</td>
<td>[Man-days:+2, Random(Availability: +5, Performance: +10, Quality: +10), set(OEE(Laser system)), Total cost: +£500]</td>
</tr>
<tr>
<td>Preventive maintenance</td>
<td>[Manufacturer, Laser job shop, Laser system, (Random (Running time) and OEE &lt; 90)]</td>
<td>[Man-days:+1, Random(Availability: +5, Performance: +5, Quality: +5), set(OEE(Laser system)), Total cost: +£700]</td>
</tr>
<tr>
<td>Repair maintenance</td>
<td>[Manufacturer, Laser job shop, Laser system, Random (Running time)]</td>
<td>[Random (Man-days: +2-4), (Availability: +5, Performance: +10, Quality: +5), set(OEE(Laser system)), Total cost: +£300-800]</td>
</tr>
</tbody>
</table>

Table 2: Conditions and consequences for each activity.

Table 2 consists of short explanations but further explanation is not provided due to the word limitations. Activity sequence has been defined in the order of produce, sell, use, planned, preventative and repair maintenance activities. Modelling through the goals, activities and environment provides a clear picture about the focus of the system. Specifying conditions and consequences of activities help to define and understand the possible state changes happening within the environment.

It has been assumed in this problem that the laser system is at the initial release, so past performance data is unavailable. Randomness is used to generate the possible variation to represent the consequences of use, planned, preventative and repair activities. If sophisticated mathematical functions are available to represent the conditions and consequences, it could be easily incorporated in Service CAD [5]. Also it is assumed that the laser system is produced with 100% OEE.

3.3 Generated results

The simulation has been carried out for 100 times with each simulation accounts for one month of use. It has been noted in the complete simulation times that all laser systems are operational. Figure 5 plots the man-days required to carry planned, preventative and repair maintenance. The pointed line illustrates that after 50th simulation (i.e. after fourth year) substantial human resources (more than 80 people) are required to carry out the activities. Figure 6 represents the gradual depreciation of OEE values across simulation cycle. It shows that until the simulation cycle of 70, OEE has been maintained above 75% for all the 100 installed laser systems. It means that the current set-up is sufficient for the five years contract in fixed cost terms. Figure 7 illustrates the fluctuation of laser system availability values of 50, 70 and 90 respectively.
Figure 8 depicts the linear progression of total costs incurred in these maintenance activities. This cost suggests that approximating £32K to each laser system could be the best price for carrying out these activities in the Use oriented PSS model. This cost does not include the capital and other costs incurred during the business operations. These outcomes are obtained for one set of conditions and consequences of the activities specified. By varying the conditions and consequences in terms of availability, performance, quality, maintenance time interval, man-days required, incurred total cost and influences of running time, various PSS design solutions can be generated.

Plotting these curves through basic experiences would help companies to understand PSS business solutions and respective systems required to support these solutions. Various parameters such as maintenance interval time, availability of spare parts, time to diagnose the problems and adequacy of labour could be varied and their impacts on the overall system could be modelled and studied. For example, Figure 9 shows the influences between availability and maintenance interval period. It is possible to achieve certain level of OEE either by varying maintenance time interval or increasing the time interval between MTBF. The next section details the limitations in this approach and suggests enhancement modules to increase the capabilities of this software.

Figure 9: Influences between availability and maintenance period

4 SUGGESTION FOR ENHANCEMENT MODULES

Based on the literature review and in-depth analyses of Service CAD through application of a sample problem, various required enhancement modules are identified for a more systematic PSS development. In this section, these identified enhancement modules are listed and some of the modules are detailed:

- Problem definition

Currently in the existing PSS software platform, PSS problem description and solution representation are merged together. Even though Service CAD provides goal and quality to define the PSS problem, they are not separated with the design objects (activities, constituents, environment and parameters). Developing this problem definition module should support designers to get the information from the customers through structured format and could be translated into requirements and constraints. Understanding these constraints is vital in developing the solutions. The structured format could be in terms of reliability, maintainability, availability, flexibility, cost, time scale, performance and risk. Analysing currently available industrial service tender documents in business to business environment could provide more elaborate structure to this module.

Explicit and clear definition of PSS problems, help customers to articulate their needs and desires in a relatively more structured form. It also aids manufacturers to identify and evaluate superior solutions through which customer's actual needs are satisfied. It is commonly noted in product and service design research that 'needs' and 'desires' are not appropriately transformed into problem statements and requirements. Often problem statements are incomplete, inconsistent, imprecise and ambiguous. Even though problem statements will contain these characteristics during the initial stages of design, a holistic approach which considers the whole lifecycle is usually missing.

Based on our understanding through literature and analysing tender documents used in the industries to formulate the requirements, we have identified seven major factors that should be incorporated in the PSS problem definition.

![Problem definition module developed in Service CAD](image)
These factors are detailed below with the laser case study example.

- Map the needs behind the needs: Customer needs and their respective rationale should be captured by mapping the needs behind the needs. For example, in laser job shop the needs behind the needs can be; minimum possible time to cut laser parts, to drawing errors and to end product manufacturer needs.
- Key Performance Indicators (KPIs): Measurable factors are important to develop well-built PSS solutions. Reliability, maintainability, availability, flexibility, performance, quality, cost, time scale, performance and risk could be some of the KPIs.
- Variation of usage patterns for the intended period: Functional aspects are largely related with the product usage activity and duration. Figure 10 represents this factor through matrix between years and size of laser machined parts.
- Experiences of the customer: Experienced customers would have the ability to answer the questions what they want and how they want it. In laser system case study, the laser job shop owners are very experienced in laser science and its application.
- Past occurrences in the business: Laser job shop owners could share the problems they face when using laser equipment such as time lost in fixing laser heads, cleaning optical parts etc which could help design better PSS solutions.
- Currently established relationships with suppliers and their customers: Since PSS covers a system level solution, understanding relationship between various stakeholders is vial. For example, the laser job shop has fixed contract relationship with gas supplier and pay per laser cut contract with their customers.
- Existing resources: Since the PSS solutions will be rest on the resources available with the customers, mapping existing resources is crucial. Resources could include skilled and unskilled labour, work space available, availability of consumables, shop location and financial strengths.

The first three factors help to understand the customer’s needs and other factors help to understand the rationale behind these needs. Figure 10 illustrates implementation of this module through these factors in Service CAD. Explicitly specifying these factors at the initial stage would help to design better PSS business models and to develop systems to support these solutions. Among these factors, specifying the needs behind the needs, usage patterns and KPIs need better support. The validation of this module is currently ongoing with many testing PSS examples.

- Co-Creation modelling - Resource modelling, Stakeholder modelling, Supply network modelling

Co-creation process between the customer, the manufacturer and the suppliers plays a vital role in developing adaptable PSS solutions in a systematic manner. Capabilities of retaining, enhancement, and relinquishment between the stakeholders are the primary outcomes from the co-creation process. The main emphasis in this process is to bring in the customer's resources along with the manufacturer and suppliers to achieve beneficial outcomes. This brings resource modelling into the part of co-creation process.

It should be noted that the additional resources required to perform an activity needs to be extended. But in Service CAD, the activity will be terminated if the condition to match the resources fails. Partial matching of the resources and adjusting to the required performance will be the key elements to be modeled. Also sensitivity of the system should be evaluated by understanding the key influencing factors. This modelling will then continue go on to assign responsibilities between the stakeholders. This part also includes supply network assessment to assess the capabilities to support the developed solution.

- Product and Operational modelling

Since integration and impact between products and services are continuously discussed in literature, these characteristics need to be modelled structurally. Capability shifts between products and services need to be modelled. Failure scenarios and product environment need to be properly constructed. Generally functional, life cycle and failure frequency are modelled for products; and for services; types, combination, response time, technician's skills and frequency are modelled. Currently available software describes the processes, interactions, precedence and causality relations and conditions and consequences of the activities. But the complexity of influences between the activities is not appropriately addressed. This is linked to the representation structure and understanding of the impact generated by these linkages.

- Solution representation - Cost modelling

This module should demonstrate the complete solution generated for the stated problem. Current version of software platform lacks in defining the final solution. The solution should explain product and service characteristics, business model elements, activities to be carried out, required resources and interactions among them. The solution should be detailed enough such that it could be alternative for bid document which is a current industrial practice consists of solutions for the released tender documents. Also this module should represent variety of solutions and provides comprehensive comparison between them in terms of costs and other influencing parameters. This module should cover detailed cost modelling for the solutions through responsibilities and resources allocated between the stakeholders. Outcomes could be in the form of matrix of combinations between products and services (solutions) represented by cost for each solution which could lead to better evaluation and to take informed decision.

5 CONCLUSION AND FUTURE WORK

In this paper, merits and limitations of currently available non-commercial software are discussed and in-depth explanation of PSS concepts development for laser systems has been detailed through Service CAD. In the course of these analyses and comprehensive literature review on PSS design methodologies, various necessary modules required to enhance Service CAD are proposed. The proposed modules illustrate that there is a huge scope for advancement in development of PSS solutions. One of the modules, problem definition is programmed in Service CAD and its importance is demonstrated. Our future work aims to develop a software platform through conceptual design framework that addresses the capabilities and requirements of the service network and customer using a co-creation process.

6 REFERENCES


Developing an Integrated Design Strategy for Chip Layout Optimization

W. W. Wits, J. M. Jauregui-Becker, F. E. van Vliet, G. J. te Riele

1 Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands
w.w.wits@utwente.nl
2 Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands
f.e.vanvliet@utwente.nl
3 Thales Nederland B.V., Hengelo, The Netherlands
gertjan.teriele@nl.thalesgroup.com

Abstract

This paper presents an integrated design strategy for chip layout optimization. The strategy couples both electric and thermal aspects during the conceptual design phase to improve chip performances; thermal management being one of the major topics. The layout of the chip circuitry is optimized according to the proposed design rules. This offers chip layout designers an intuitive way to optimize the layout for multiple performance indicators, such as temperature, RF power output or amplifier gain. In a case study, the strategy proposed a chip redesign, boosting overall chip performance without compromising the current cooling infrastructure. The developed integrated design strategy presents a new and time-efficient approach to chip layout optimization and electronics cooling in general.

Keywords: Thermal management, Electronics cooling, Chip layout optimization, Integrated design strategy, Layout design

1 INTRODUCTION

Thermal management has become one of the major challenges for electronic product design. Electronic component dimensions, for instance chip and resistor sizes, decrease continuously, whereas the demand for more power and functionality increases. The net result is a significant gain in component power (heat) density that needs to be controlled. As an example, the increasing power density of Intel microprocessors [1] is shown in Figure 1.

Figure 1: Increasing power density and scaling of heat sink devices for Intel microprocessors [1].

The main strategies to thermally manage the increasing power densities are focusing on the cooling equipment itself (increasing heat sink sizes, increasing air flow performance, etc.) or a change in cooling technology (from air cooling to fluid cooling or two-phase principles). Some of these cooling technologies are also shown in Figure 1. There are many examples of research primarily focusing on increasing cooling capacity [2-5]. A similar trend where this modular approach can also be observed is in publications on the design of LED modules. Researchers take the power specifications of the LED module as an input for their design process and design various cooling solutions to support the given LEDs [6-7].

Another trend in electronics design has been to focus on component layout management. By positioning and spreading high power components, cooling of the entire product can be better managed. Cheng et al. [8] use a sequential meta-modeling approach of analysis techniques to optimize board layout. Other approaches, amongst many others, are based on simulated annealing [9] or statistical parametric optimization [10]. Similar to the previous examples, these techniques also take the initial felon for granted and try to work around them following a modular approach.

As Figure 1 also indicates, future power densities will rise to absorbent levels, also referred to as the economic meltdown of Moore's law [11]. This calls for a more integral approach for the design of not only the cooling system, but the electronic components as well. This study focuses on the optimization of the dissipating electronic element itself.

1.1 Traditional electronics design process

Pahl and Beitz [12] presented in their model of the design process that during the conceptual design phase, a concept should be developed that fulfills the primary functions of the product. Whereas during the embodiment and detail design phase this concept is given its detailed description. The latter phase is illustrated more elaborately in Figure 2.
According to Figure 2, through a synthesis process the concept is developed into a distinct embodiment, which incorporates specific product form and dimensions. This detailed description is required for the analysis process, which predicts product performance. Finally, in an evaluation process these performances are compared to the initial specifications. To optimize product performance, the embodiment may be modified by returning to the synthesis process, after which the altered version is analyzed and evaluated again. Depending on the number of iterations, an optimal embodiment (the solution) of a (single) certain concept is reached.

Traditionally the conceptual design process for the electronic component encompasses mainly knowledge from the electrical domain combined with production knowhow. To manufacture an electronic product, usually various electronics components are combined and assembled on a Printed Circuit Board (PCB) structure. Once an embodiment is formed in the detailing phase, analyses can be performed. Generally this is also the time that thermal analyses are performed.

1.2 Cut and run engineering

Industry generally prefers to stick to tried and tested concepts, as this gives a short time-to-market at minimal development cost. Pugh [13] refers to this as “cut and run”, as engineering is started before the conceptual design phase is thoroughly examined. This applies also to the electronics industry, where today’s product is dispensable in 2 to 3 years time. Hence, they concentrate on the embodiment and detail design phase for their product development. The iterative loop shown in Figure 2 is sometimes also referred to as a trial-and-error approach.

When requirements are updated, proven concepts might not be able to fulfill these specifications and the design process will fail to converge to an acceptable solution. Subsequently, to reach an acceptable solution, one has to return to the conceptual design phase to develop new concepts. This is indicated by the dashed line in Figure 2. Needless to say, this is more time consuming and costly; hence, the industry tries to avoid this.

1.3 Thermal design critical

As thermal management aspects, until recently, scarcely impeded an acceptable product design, thermal analyses were addressed toward the end of the design process or not at all. Cooling was considered a support function, as it supports the primary functions. Consequently, cooling should not impede primary functions, as this leads to an unrealizable concept design. When thermal limits became increasingly critical, issues were initially addressed only in the embodiment and detail design phase. As the concepts themselves did not evolve, product changes were relatively small; hence, the ever-growing heat sinks.

Many reviews indicate that thermal control has already become a critical factor in the design of electronic equipment. For instance for mobile phones, thermal management has been a critical issue since the 1990s, especially in the case of lengthy call durations. With respect to battery usage, the loss of reliability due to increased internal temperatures was significant. According to Yeh [14] in 1995 more than fifty percent of all electronics failures were caused by undesirable temperature control.

1.4 Integrated design approach

Continuing to focus on thermal issues as a support function eventually will lead to conflicting requirements. The persistent growth of thermal management systems, such as heat sinks, contradicts the development of – in particular small and compact – electronic products. Especially for high power electronic products this means product dimensions and the positioning of components are no longer determined by primary functions; instead the embodiment will be affected by thermal criteria, a support function, as well. Moreover, as mentioned, product design may even fail to converge due to thermal constraints.

To solve these issues, a more substantial thermal engineering effort, earlier in the design phase, is required to further integrate primary and support functions. The design of electronic products in a monodisciplinary manner, for instance by focusing on electrical aspects only, should be avoided. Instead, a multidisciplinary and integrative approach must be utilized to achieve an optimal system configuration.

Domain decomposition

A first step toward such a multidisciplinary and fully integrated design process is to decompose the system requirements according to the domains involved. In 2000, Papalambros and Wilde [15] presented a method for system decomposition: Multidisciplinary Design Optimization (MDO).

In the case of the electronic component, the decomposition is illustrated in Figure 3, where E, M and T refer to the electric, mechanical and thermal domain, respectively.
share a mutual engineering "language", even small changes in one domain can have profound effects in other domains. In the case of the design of electronic components, this is a first and important step.

**Full domain integration**

Ideally, the knowledge of the engineering disciplines involved is directly related to each other and a mutual language has been developed. Evidently, both cases require in-depth and coherent knowledge of the engineering fields involved. This full integration of knowledge disciplines is illustrated in Figure 4.

1.5 Outline

The outline of the paper is as follows. The next section describes the basic building block for an electronic component: the transistor. It also highlights the applicable design parameters and their major influences. Section 3 presents the proposed integrated design process for chip design and in Section 4 the implemented design process for layout design is described. Also, the optimization strategy for the integrated process is presented here. Section 5 describes the application of the integrated design process and optimization strategy to a case study example. Finally, conclusions and future recommendations are drawn.

2 TRANSISTOR BUILD-UP FOR LAYOUT DESIGN

2.1 Transistor background

Chips perform their electronic functions by a number of interconnected transistors. Each single transistor acts as a switch that can be turned on or off. For microprocessor applications each transistor is used to control a binary (digital) signal: a one (1) or a zero (0). One of the latest Intel chips actually combines more than one billion transistors for their computing power.

The typical layout of a transistor is shown in Figure 5. The transistor is formed by a highly conductive channel deposited on a semi-insulating substrate. The terminals (source and drain) are connected to the substrate with an ohmic metal contact. The gate is also connected to the substrate but has a metal-semiconductor junction in between. This potential barrier acts as a diode and is used to open and close the gate quickly. The length of the gate (denoted L in Figure 5) is defined as the distance between the source and drain. The gate width refers to the extension of the transistor in z-direction. Typically the width is larger than the length.

2.2 Transistor for power amplification

Next to digital applications, transistor technology is also used in analog or microwave applications. Here, the transistor is used as an amplifier that can be switched on and off. When the transistor is switched on, the incoming signal is amplified. To boost amplification, multiple gates can be positioned in parallel. In this case, all of the gates of the transistor combined are capable of controlling the transistor state. If the gates are opened by an applied voltage, electrons can flow from the source to the drain terminal. If no voltage is applied, the gates are closed and the electrons are blocked. The flow of electrons is affected by the size and shape of the transistor. Hence, the amplification characteristics of the transistor are also determined by its geometry.

On a typical High Power Amplifier (HPA) chip the transistors are arranged in amplifier stages. The HPA chip may consist of a single or multiple amplifier stages. Each amplifier stage consists of a single or multiple transistors and within each transistor multiple gates can be positioned in parallel. Parallel positioning of the gates forming a multiple-gate transistor is shown in Figure 6.

Figure 5: Basic single-gate transistor layout [16].

Figure 6: Transistor layout with a multiple-gate structure [16].
gate-to-gate spacing are important to determine the transistor performance.

The transistor is deposited on a semiconductor substrate and this semiconductor device (chip) is in turn soldered into or onto an electronic package. Hence, other geometrical design parameters for this case study are the substrate and solder thicknesses. Table 1 lists all geometrical parameters. To get a better understanding of the design process, their influence on the chip performance is also given.

2.3 Electric and thermal criteria

Next to the geometrical design parameters, other design parameters are those of the electric and thermal domains. As was mentioned in Table 1, the gate width determines the amplifier performance. Since the width is limited, to boost performance gates are positioned in parallel. Also, the operating frequency of the Radio Frequency (RF) input signal is an important parameter. Usually the frequency is demanded by the application. Other electric criteria are the pulse length and the chip's duty cycle. During the time of the pulse (pulse length) the RF signal is amplified by the chip. Due to the chip's limited efficiency during this time thermal dissipation, heating up the device, also takes place. The duty cycle refers to the percentage of time that the transistor is actually amplifying versus the total operation time. A low duty cycle means a short heat-dissipation time and longer cooling-only time; whereas at a 100% duty cycle the transistors are always dissipating heat.

From a thermal perspective, two additional design parameters are important: (1) the dissipated power and (2) the operating temperature. The power of the input and output signals, the chip DC power and the chip efficiency determine the total amount of power that is dissipated as heat by the transistors. This heat needs to be disposed of through the substrate and solder layers to the external environment.

Chip temperature has an influence on the conducting properties of the base material; hence, it influences the electric efficiency of the chip. This coupling or loop-back effect is usually ignored due to its complexity and to speed up the design process. The chip's operating temperature and material properties determine to a large extend the Mean Time To Failure (MTTF). Usually the application demands a certain MTTF, which limits the maximum operating temperature.

The electric and thermal criteria are summarized in Table 2. For this paper, only the electric criteria that have an impact on the thermal characteristics are presented. Other aspects, as for instance the influence of interstage matching on signal reflections and oscillations, fall outside the scope of this study.

2.4 High Power Amplifier (HPA) chip design

As mentioned in Section 2.2, on a typical HPA chip the transistors are arranged in amplifier stages. Figure 7 shows the chip layout of a 16W X-Band HPA chip. X-Band refers to the frequency range of the RF signal. As indicated in the figure, this chip has two amplifier stages. Also shown are the matching circuits, typical for analog chip design. They split the input signals and combine the amplified signals, like a piping system. Also, they control the signals' impedance. Its layout is dependent on the electric parameters.

Due to the limited conducting properties, power is also dissipated in the matching circuits. However this can be ignored compared to the heat dissipation in the amplifier stages, where the transistors are positioned in parallel.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Design criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating frequency</td>
<td>[Hz]</td>
<td>Frequency of the RF signal.</td>
</tr>
<tr>
<td>Pulse length</td>
<td>[ms]</td>
<td>Time of amplification of the RF signal.</td>
</tr>
<tr>
<td>Duty cycle</td>
<td>[-]</td>
<td>Percentage of time that the amplifier is amplifying.</td>
</tr>
<tr>
<td>Dissipated power</td>
<td>[W]</td>
<td>Amount of power (heat) that is dissipated by the transistors.</td>
</tr>
<tr>
<td>Operating temperature</td>
<td>[°C]</td>
<td>Maximum temperature at the gates of the transistors.</td>
</tr>
</tbody>
</table>

Table 2: Electric and thermal design criteria.
The total number of stages, transistors and gates determine the output power and the amplification gain. Hence, the design can be tailored to meet the required performance of the HPA chip.

3 CHIP DESIGN PROCESS

3.1 Current chip design process

Currently, the design of electronic components is mainly determined by electronic properties. Since until recently the chip temperature did not pose a serious threat, thermal criteria were not a part of the initial design requirements. The design of electronic components was verified thermally by checking the maximum temperature value of hotspots (the gate temperatures) on the surface during operation. The MTTF based on the maximum operating temperature of the components was used to estimate the component lifetime. When the temperature was rising too high, better cooling devices were applied, as was shown in Figure 1. Hence, thermal criteria were applied as a form of go/no go option at the very end of the design phase. Schematically, this is shown in Figure 8.

As shown in Figure 8, an initial chip layout is proposed as an input to determine the electric performance. The electric optimization loop represents the current strategy electronic engineers apply in their search for the optimal chip design based on the initial requirements. This includes, amongst others, finding optimal parameter values for the first four geometrical parameters of Table 1. Once a chip layout is (electrically) validated and optimized, thermal engineers validate the chip layout. Based on the initial requirements and the final chip layout, the power dissipations at the transistor junction (gate) are determined. Finally, if sufficient cooling capacity can be achieved into the electronic product such that the junction temperatures remain below the demanded MTTF temperature, the chip design is validated. Otherwise, a redesign of the whole layout is imminent.

In practice, the redesign loop is hardly ever practiced. Since thermal analyses are done at the very end, returning would be very time consuming and expensive. As chip design progresses, thermal engineers have to become more and more creative to handle the thermal design. However, they still have no say in the original chip layout, since this is outside of their knowledge domain. Hence, there is still ample space for optimization.

3.2 Proposed integrated design process

This paper proposes an integrated design process that focuses on both electronic and thermal criteria simultaneously. In this new approach, thermal criteria have more influence during the conceptual design phase of the electronic component, compared to the current design process discussed in the previous section.

For instance, with a different chip layout, local thermal hotspots on the chip surface that limit the demanded criteria might be avoided. As a consequence, the maximum chip temperature values can be lower for equal or better electric performances. Hence, overall better performances can be achieved. The process of this strategy is shown in Figure 9.

As illustrated in Figure 9, an initial chip layout is proposed as input to determine both the electric and thermal performances. To determine either performance values, knowledge of the other domain is required. The amount of heat dissipation is determined by electric criteria; whereas electric performances are temperature dependent. Hence, there is a strong coupling between the parameters.

In recent years some work has been done in this field. For instance, Codecasa et al. [18] and Vellvehi et al. [19] both describe a method for electro-thermal analysis. The former uses a network (or lumped) approach that mimics the physical properties of the original phenomenon, whereas the latter uses a two-step approach where electrical and thermal models are solved separately and exchanging performance parameters periodically. Multi-physical-domain analyses are usually done this way, for instance for integrated power devices [20] or bipolar transistors [21]. The latter also present a strategy to evaluate the temperature fields. This information could be used in the design process.

Currently, as illustrated, the direct relation between specific, low-level electric parameters and local structure temperatures is not well understood. For this study, the coupling is realized in a global descriptive way. The dissipated thermal power is coupled to the overall chip electric properties, such as gain, input power and power added efficiency and the total gate width. In terms of handling design knowledge this falls under the category of domain decomposition as described in Figure 3.

Ultimately, as the underlying fundamental principles are better understood a stronger coupling between the local electric parameters and local thermal influences can be modeled. Hence, full domain integration as described in Figure 4 can be accomplished.

4 CHIP LAYOUT OPTIMIZATION STRATEGY

4.1 Identifying the total design space

To find a uniform chip layout optimization strategy, one of the transistors in one of the amplifier stages is modeled and analyzed in a finite element analysis software tool.
With this finite element model the total design space is identified. The design parameters are the geometrical layout properties of the chip, as listed in Table 1; the model's input criteria are the parameters of Table 2.

Since some of the parameters are dependent, to start two independent parameters are chosen. The first independent design parameter is the number of gates. It was varied from one to a maximum value. The second independent parameter is the amplification gain, represented by the total dissipated power. This also determines the total gate width. The transistor technology (e.g. GaN, GaAs, etc.) determines substrate thicknesses and the available chip surface. The following iterative process scheme was used:

1. Adjust the total number of gates (add 1 each time).
2. Adjust single gate sizes (length and width) to the overall chip power requirements.
3. Determine the gate-to-gate spacing.
4. Model and mesh the current layout.
5. Set thermal power dissipation.
6. Determine local gate temperatures with a finite element analysis.
7. Store the input geometry with the generated performance values separately.
8. Return to Step 5, until maximum power is reached.
9. Return to Step 1, unless the maximum number of gates is reached.

For the numerical analysis Comsol Multiphysics was used. Comsol has a good interface with Matlab, which allowed us to automate the iterative process scheme. After the scheme was automated robustly, a high performance computer fully characterized the entire design space. This took several days, for one type of transistor technology.

One of the calculated layouts is presented in Figure 10. It shows a possible configuration for the HPA chip. Four gates are modeled – identified by the finer mesh – at the calculated gate-to-gate spacing (112.5µm). In this case, the gate width and length are 250µm and 1µm, respectively.

Figure 10: 3D mesh of GaN/SiC HPA chip.

Before identifying the total design space, the transistor finite element model was compared to an in-house benchmark reference dataset. The simulated temperature values at the gate fingers showed a maximum difference of 0.1°C. For a thermal analysis this is an acceptable value.

Data representation

The calculated data from the iterative process can be represented in many ways. Most parameters show an expected trend when the dissipated power is set to a constant value. For instance, for increasing gate width the maximum temperature will increase to a certain threshold value. This is explained by the fact that thermal saturation takes place in the gate finger. Another example is that for increasing gate-to-gate spacing the maximum temperature drops continuously, as can be expected since the power density also becomes lower.

Figure 11: Maximum gate temperature as a function of gate-to-gate spacing and gate width.

Figure 11 illustrates both described effects. For increasing gate width values, gate temperatures increase reaching a threshold. The threshold is caused by the thermal saturation of the gate width. When saturation is reached, changing the gate width does not affect the gate temperature anymore. However, by changing the gate-to-gate spacing still a temperature drop can be realized. In this specific example, a temperature reduction of approximately 25°C is still possible.

4.2 Building an optimization strategy

Further in-depth analysis of the total design space has led to the development of a transistor layout optimization strategy. The goal of the optimization steps is to get as much performance out of the transistor as possible without overheating. This strategy is illustrated in Figure 12.

Figure 12: Chip layout optimization strategy.

The optimization scheme centers around the chip temperature. If the temperature is too high and gate saturation is reached, the gate width must be decreased. When saturation is not yet reached the gate-to-gate spacing must be increased. A similar strategy but vice versa must be applied when the chip temperature is too low. Finally, at the desired (maximum MTTF) temperature optimal performance is achieved.

To make this optimization strategy manageable, in our implementation for both the gate width and the gate-to-gate spacing the smallest discretized step was 1µm. Using finer step sizes did not yield higher order effects, but the computational load increased tremendously.
5 CASE STUDY
To demonstrate the procedure and advantages of the proposed integrated design strategy, it has been applied to an industrial case study. The goal was to redesign a HPA chip, based on a GaN/SiC technology platform, used for naval radar systems.

In the design a two-stage chip design was used to reach the required RF power output. The second stage of the chip consisted of several 16-gate transistors. Analysis of this chip design shows a maximum gate temperature of 223.2°C. Figure 13 shows the initial layout of the second stage with the temperature distribution.

![Figure 13: Initial transistor layout (16 gates) overlaid with thermal analysis.](image)

Following the chip layout optimization strategy of Figure 12, it was discovered that the maximum gate temperature could be reduced to 186.1°C, almost 40 degrees lower than the initial design. In the final design, the number of gates was reduced to 8. At the same time, the gate-to-gate spacing and gate-width was optimized to compensate for the lower number of gates. The final design is shown in Figure 14.

![Figure 14: Final transistor layout (8 gates) overlaid with thermal analysis.](image)

With the transistor layout of Figure 14, the required RF power can still be realized. However the gate temperature is much lower (approximately 40°C). This layout performs much better thermally for equal global electric performance. Hence, in total a more overall optimized design is realized. Also, these analyses can now be performed during the very early stages of the conceptual design phase.

For this case study, this meant that radar performance could be boosted without changing the current cooling infrastructure. Obviously, this is a very desirable outcome, not only for this case study.

6 SUMMARY
By integrating electric and thermal criteria directly into the conceptual design phase of electronic components, better chip performances can be realized. A model of the transistor gate structure is presented, highlighting important parameters. The parameters are split up into geometrical design parameters, and electric and thermal input criteria. Finite element analyses were used to identify the entire design space. By analyzing this data, a new design strategy is formulated for chip layout optimization.

As a result of this study, chip performance can be improved with lower local hotspot temperatures. The local maximum temperature was lower due to a better geometrical layout of the chip circuitry. The proposed strategy was successfully applied to a case study: a redesign could be proposed that fulfilled the RF output requirement while reaching a lower chip temperature.

In the end, this study tries to give chip layout designers an intuitive and pragmatic way to optimize the layout for a multitude of performance indicators, such as temperature, RF power output or amplifier gain.

6.1 Future perspective
Further research is required to tackle further chip layout optimization. For instance, the number of amplifier stages is still chosen as an expert's educated choice. By further expanding the total design space, new strategies are expected to be found to solve these design issues as well.

More fundamental research might reveal the direct interactions between electric and thermal effect locally in the gate finger regions. If this knowledge can be captured in an integrated multi-physics model, the integrated chip design process can be further optimized.

7 ACKNOWLEDGMENTS
The authors like to specifically thank Niek Albers. His work [16], constructing the numerical model and also automating the analysis steps, was an essential part of this study.

8 REFERENCES


Cognitive Designers Activity Study, Formalization, Modelling, and Computation in the Inspirational Phase

C. Bouchard, J. E. Kim, J. F. Omhover, A. Aoussat
Products Design and Innovation Laboratory (LCPI), Arts et Metiers ParisTech,
151 boulevard de l'Hôpital, 75013, Paris, France
{carole.bouchard, jieun.kim, jean-francois.omhover, améziane.aoussat} @ paris.ensam.fr

Abstract
This paper refers to a research project that we are conducting about the formalization of the designer's cognitive activity in order to develop new computational tools to support the early design process. These tools are especially focused on the inspirational phase of design. We first formalized the cognitive processes of the designers dedicated to our specific phase, and identified some routine parts where computational tools could be useful in order to enrich the traditional design process. The computation of design rules in the early phases of design needs to establish specific formalizations that can be implemented by algorithms. After modeling designers' cognitive processes, we explored the main information systems they use and completed them by an investigation about Content-Based Image Retrieval systems (CBIR). Our research consisted then in establishing specific formalizations in order to cope with recent technologies that could improve the precision and efficiency with which designers can access inspirational images.

Keywords:
Inspirational process, Case study, Conjoint Trends Analysis, CTA method

1 INTRODUCTION
This paper describes a research project to model the cognitive processes of designers in order to develop computational tools for the earliest phases of design. The formalization and explicitation of designers' cognitive processes are becoming a strategic topic for many scientific communities including design science, cognitive psychology, computer science, and artificial intelligence. This growing interest is partly due to pressure from industry where the shortening of development delays and the increasing variability of the offerings expected by the consumer require a formalization and a digitalization of the earliest phases of the design process.

In this context, three research areas are now well established and tend to develop new models and tools that will help to progressively digitize the early design process:

- the formalization of the cognitive design process with the extraction of design knowledge, rules and skills;
- the translation of design rules into design algorithms;
- the development of software tools that will be used by the designers themselves and the other trades involved in the early collaborative design process.

Following this, we first investigated the cognitive activity of designers and focused on the inspirational phase. These cognitive processes were formalized as a design method named the Conjoint Trends Analysis (CTA) method. The CTA method [1] is a recent method which has been molded to the information gathering process in industrial design, taking into account the task-based requirements and the cognitive and affective processing of designers.

Our original work focused on the identification and use of various domains of influence (nature, arts, industrial sectors, sociological end values) in order to enrich the design solution space.

Finally, the CTA method enables the identification of formal trends in attributes (shape, color, textures) linked to particular environments in order to use them in the early design of new products. This makes it possible to enrich and to inspire the designers and the design team when designing products. It is positioned in the earliest phases of the design process.

2 COGNITIVE DESIGNERS ACTIVITY, FORMALIZATION AND MODELLING
2.1 The information phase in the early stages of design: the inspirational process

The design process reduces abstraction through the use of various successive levels of representation which integrate more and more constraints. It can be seen as an information processing activity that includes informative, generative, evaluative and deductive stages. The informative phase is a crucial. First, it enables the completion of design problems which are by nature ill-defined and ill-structured [2-3] and so refers to semantically impoverished tasks.

Designers use a large variety of sources coming from different areas such as comparable designs, other types of design, images of art, beings, objects and phenomena from nature and everyday life. Sources of inspiration are an essential base in design thinking, as definitions of context, triggers for idea generation [4], and anchors for structuring designers' mental representations of designs. In favorable contexts, designers built trend boards in order to structure their inspiration sources. Trend boards offer a visual and sensorial channel of inspiration and communication for design research and development, which could be considered to be more logical and empathic within a design context than only verb-centric approaches [5]. They are usually a collection of images compiled with the intention of communicating or provoking a trend or ambiance during the product design process.

CIRP Design Conference 2011

63
As a routine part of the creative process, product designers search for and collect materials that they find inspirational. They get their inspiration in their personal lives and through a more focused way in their professional lives, in various sources like specialized magazines, bibliographies, material from exhibitions and the web.

2.2 The Conjoint Trends Analysis (CTA) method

The information phase of the cognitive design activity was studied and formalised in order to define the Conjoint Trends Analysis (CTA) method [1]. As shown in Figure 1, the CTA method is composed of three steps: gathering/categorizing images and keywords; ambience definition; and pallets composition. The CTA method enables the identification of attributes linked to particular datasets (e.g., common properties of images in a database) so that they can be used to inspire designers in the early stages of design for new product.

Figure 1: The Conjoint Trends Analysis (CTA) method

The CTA results are trend boards that represent sociological, chromatic, textural, formal, ergonomic, and technological trends. The trend boards communicate identified homogeneity in terms of style and consumers' sociological values. They are mainly based on visual information, and come from the frequent occurrence of certain properties within a dataset. From this analysis, images and relevant words are selected and formalized under a form of ambiances. Ambiances are typical representations where the emotional impact is intended to be higher. Global and discrete design elements are then extracted from these ambiances under a form of pallets. These design elements are used for the generation of new design solutions.

Therefore, trend boards offer a relatively exhaustive representation of the references usually used by the designers for their composition and play an important role in stimulating idea generation while anchoring contextual matter [4]. They reinforce the link and semiotic coherence between the consumers end values, functionalities in any domains of influence, and product attributes such as form, color, texture, and usability principles. Another purpose of the trends analysis is to define user-convenient principles and solutions that can be integrated in future products. Indeed designers often have to provide new designs using insufficient information about consumers. Trend boards show ambiances including people in context. Contexts are decisive in the attribution of a signification to the object.

3 CONTENT-BASED IMAGE RETRIEVAL (CBIR) APPROACH

Content-Based Image Retrieval (CBIR) is a technology that in principle helps organize digital picture archives by their visual content (colors, shapes, or textures) [6]. User queries are mainly based on image example, region of interest, or concept keyword. The search process mainly consists of query formulation, specification of which images to retrieve from the system from the database (DB) in various ways: one by one browsing, keywords or image feature specification, automatic image feature extraction, or finally providing an example from which features will be compared by similarity to those of other images. The technologies used come from a range of scientific knowledge bases from artificial intelligence and computer vision applications, including statistics, pattern recognition, and signal processing. Searches must rely on metadata such as captions or keywords. These keywords can be generated by a human or automatically extracted from the web.

CBIR systems tend to turn towards specific categories of end-users working in areas where creative thinking is needed, like industrial design or architecture. On the other hand, specific databases are increasingly developed in those domains, which do not yet integrate the sophisticated technologies from CBIR. Multi-Dimensional Scaling seems to be a good way to interactively display sets of products arranged according to semantic axis. Besides, the use of a folksonomy for gathering progressively relevant expert material for building the ontology seems to be a promising way to bridge high and low level information.

3.1 Supporting design by Content Based Image Retrieval (CBIR)

CBIR systems are applied in the context of huge databases where the question arising is how to find the right image. This is particularly difficult when dealing with design information. Indeed, the design activity is specific in the way that design information covers various levels of abstraction of the information. However the main majority of commercial and academic systems still focus on low-level features. This difficulty is increased when dealing with images that have a multi-dimensional and subjective nature as is the case in the context of industrial design. One major problem in CBIR is to link visual content with semantic content to bridge the semantic gap between low-level content and higher-level concepts. The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in the same situation [6]. To close the problem of the semantic gap, some CBIR systems automatically generate additional search terms by conceptual similarity with the original terms, as identified through term co-occurrence in a corpus or through the use of an ontology. In the field of CBIR, many papers mention the reduction of the semantic gap, but they often do not refer to semantic adjectives and are more based on concepts like car. However their architecture is often interesting insofar as these systems could also integrate semantic adjectives.

A study [7] showed that indexing concepts by keywords in image indexing systems is much more appropriate with human similarity perception than low level features and provides good retrieval performance. Naphade et al [8] suggest that matching semantics might be achieved by employing top-down as well as bottom-up techniques. In order to support the top-down method, it will be important to develop semantic nets that link higher-level concepts that are evoked by visual images. Existing systems frequently use about ten adjectives and at the best about several thousand of images in the DB. About the system architecture, we encounter often the combination of a filter and of a direct re-classification to images. When several users enter the same information, the system performs a re-classification of the images. The feedback systems using a weighting system adjusting the importance between the features that seem interesting. A remaining problem is the changing nature of the user's
perception of a semantic adjective in the time. So the feedback system has to be well developed, with a simple and efficient interaction with the user. The system should also enable the user to add information by entering new terms and new evaluations of the results.

3.2 Towards Kansei-Based Image Retrieval (KBIR)
Future CBIR systems should correlate high-level dimensions like concepts, semantics and emotions with low-level dimensions. The connection of low-level and high-level dimensions is very subjective and variable from person to person. Consequently, the previous systems are often based on a strong interaction between the end-users and the system itself, using images and semantic adjectives. It is frequently done with the intervention of the end-users thanks to learning systems using neural networks [9-11], or genetic algorithms [12].

Some studies were already led in this way, but not dedicated to the field of design [8-9,12-16]. The advantage of Kansei Engineering methods is that they focus more on the viewer rather than on the image [17], and similarity measures derived from kansei indexing come from inner experience, rather than visual similarity. These methods enable the designer to assess evoked feelings on the basis of impression words including frequently semantic adjectives (urban, romantic, aggressive, and so on) or emotional adjectives (amused, astonished) describing the viewer in front of a specific image. It is, however, difficult to develop competitive systems able to search and classify images from semantic adjectives because their appreciation may be altered by an inter-individual subjectivity. Tanaka et al. investigated which regions and features of images are most attractive, contributing so to human Kansei [18] in [17]. An increasing attractiveness seems to be correlated with size effects (attractiveness increases with size) or color effects (warmer, highly chromatic and high values colours). Hayashi et al. [18] in [11] attempted to train a neural network to predict Kansei with impression words evoked by outdoor scenery images. The best results were obtained for visual words such as spring or clear. It was emphasized that the mapping of human impressions with physical features is not one to one and that any retrieval system must retrieve multiple images, allowing the user to choose the best one. The results of this study showed a statistically significant superiority of the Kansei Based Retrieval systems over the random retrieval.

4 A CASE STUDY: THE TRENDS SYSTEM
On the base of research and industrial projects, we started to formalize and structure design information into specific formats. Especially in the TRENDS European project, we aimed to raise a new formalization of design information which can improve designers' access to web-based resources, and help designers to find appropriate materials and identify design trends in those materials.

4.1 Towards a new formalization of design information
We expected that the TRENDS system would enable us to elaborate the field of image search, including content-based image retrieval and Kansei Based Image Retrieval (KBIR). Finally, the Trends Analysis (CTA) method would be partially digitalized and implemented by the computational tool which integrates semantic processing and image contents.

This computational tool should reduce retrieval time and provide a comprehensive overview of the retrieval results in expanding the corpus of images from different sectors of influence; enhancing creativity with more or less open image retrieval; and facilitating idea generations using key harmony rule of design.

In the TRENDS project, the data collection was carried out on the basis of fictional scenarios and extraction of design information from previous projects which have used the CTA method.

![Figure 2: Design knowledge extraction by manual annotation [19]](image)

First, we validated a list of sectors of influence in car design (see Table 1). This table shows sectors of influence identified in 1997 and in 2006. Interestingly, 70% of the sectors of influence have not been changed. This implies that we could integrate some routine parts of sectors of influence in the data database of the TRENDS system.

<table>
<thead>
<tr>
<th>Rank</th>
<th>1997 (40 designers)</th>
<th>2006 (30 designers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Car design</td>
<td>Car design</td>
</tr>
<tr>
<td>2</td>
<td>Aircrafts, aeronautics</td>
<td>Architecture</td>
</tr>
<tr>
<td>3</td>
<td>Architecture</td>
<td>Interior design &amp; furniture</td>
</tr>
<tr>
<td>4</td>
<td>Interior design &amp; furniture</td>
<td>Fashion</td>
</tr>
<tr>
<td>5</td>
<td>Hi-Fi</td>
<td>Boat</td>
</tr>
<tr>
<td>6</td>
<td>Product design</td>
<td>Aircraft</td>
</tr>
<tr>
<td>7</td>
<td>Fashion</td>
<td>Sport goods</td>
</tr>
<tr>
<td>8</td>
<td>Animals</td>
<td>Product design</td>
</tr>
<tr>
<td>9</td>
<td>Plants</td>
<td>Cinema &amp; commercials</td>
</tr>
<tr>
<td>10</td>
<td>Science Fiction</td>
<td>Nature&amp; urban ambiances</td>
</tr>
<tr>
<td>11</td>
<td>Virtual reality</td>
<td>Transportation</td>
</tr>
<tr>
<td>12</td>
<td>Fine arts</td>
<td>Music</td>
</tr>
<tr>
<td>13</td>
<td>Cinema</td>
<td>Fine arts</td>
</tr>
<tr>
<td>14</td>
<td>Music</td>
<td>Luxury brands</td>
</tr>
<tr>
<td>15</td>
<td>Travels</td>
<td>Animals</td>
</tr>
<tr>
<td>16</td>
<td>Food</td>
<td>Packaging &amp; advertising</td>
</tr>
</tbody>
</table>

Table 1: Sectors of influence in car design [20-21]

Second, we identified that designers employed different type of design information and they consists of different levels: high-level (values, semantic words, analogy, and style), middle-level (sector name, context, and function), and low-level information (color, form, and texture). These levels of information can be seen as the position of an axis going from abstract (high-level information) to concrete (low-level information). The structure of design information enabled the construction of a design domain ontology. Table 2 shows some examples in the field of image retrieval.
This combination yielded a more flexible structure of information and it was more suited to design information. Therefore, the developers of the TRENDS project used these data filtering models and rules to develop the design domain ontology and the Bag of Words [36]. Those two methods could be used in applications independently or complementary.

![Design domain ontology + Bag of Words (BoW)](image)

Figure 3: Examples of semantic queries 'Chic' under two conditions: without the BoW (left) and without the BoW (right)

Fourth, we formalized the cognitive processes of designers wherein designers mentally or explicitly categorize images sources. Those categorizations were initially based on the distribution of colors on the chromatic circle according components. Further, we integrated the aesthetic harmony rule of image which is a core expertise of designers. In order to support this phase, the TRENDS system is supposed to provide a new way to retrieve images according to similarities of images corresponding to harmony rule of design and categorize them in a smart way (pallet generation process).

### 4.2 Functionalities of the TRENDS system

Finally, the TRENDS system proposes two groups of functionalities: image retrieval and design advanced functionalities. Figure 4 shows a global interface of the TRENDS system.

Regarding the image retrieval functionalities, the system searches the information, but not from the overall web. The information from the web is filtered according to the listed sectors of influence of the designers (See Table 1): for instance car design, advertising, architecture, arts, etc.

First, the image retrieval functionalities consisted of:

- **random search**: an open search, providing serendipity and so favouring creativity;
- **semantic search**: the user inputs some keywords, semantic adjectives or concepts;
- **search with an image example**: the user selects an image as a query, similar images can be found in the database based on similarity rules between the global descriptors of the image content related to shape, colour or texture dimensions;
- **relevance feedback**: when the results are sometimes not so convincing after the initial query, we could give the relevance feedback (positive or negative) manually to refine the search;

In addition, some options enable the designer to select the size of the displayed images depending on the task.
that the designers have to manage. We can also apply interesting display functions, like a slideshow for instance. Second, design advanced functionalities consist of:

- **grouping**: to automatically display subsets of images grouped by specific harmony rules. This function is required because it would be difficult for the end-user to have a complete view of the content of the sphere, wherein search results are displayed;

- **pallets function**: to generate a pallet of specific harmonies linked to low-level information between colours, shapes, and textures;

- **semantic mapping**: to generate automatically a first version of semantic mapping with a text search;

This function enables the designer to link and arrange the words with the images. Some images do not possess semantic descriptors by default. Others are overlapped when they have the same description. The designers can also achieve the complete mapping manually.

- **statistics**: to apply statistics related to a word or an image. The system promises quantitative information about the representativeness level of a word or an image in the sectors;

- **lifestyle**: a bookmark oriented towards specific websites which propose information about sociological changes, values and lifestyles evolutions. This function was a request of the designers themselves.

![Figure 4: A global interface of the TRENDS system](image)

5 CONCLUSION

This paper presented a study we led about modelling the cognitive processes of the designers in order to develop computational tools for the earliest phases of design. We focused this study on the inspirational phase of the design process, where designers have to deal with a large amount of information. Cognitive models here gave birth to the Conjoint Trends Analysis (ATC).

After practicing this method in several projects in the field of automotive design, some computer support tools were foreseen in order to enrich the information gathering.

We intended to partly digitize the ATC method which corresponds to the informational part of design activity. Two ways were combined at the same time, which were presented in this paper previously: the elaboration of ontological data tables and bag of words which were related to more or less abstract concepts, which both were used to be implemented by algorithms. The design rules so elaborated are based on particular formalizations that enable the designer to link low-level descriptors with high-level descriptors.

Cognitive models enabled the designer to identify some routines that could partly be implemented by algorithms and bring an added value to the traditional process.

6 ACKNOWLEDGMENTS

The authors are grateful to the European Commission for funding TRENDS project (FP6 IST 2005-27916) and express their gratitude to all partners of TRENDS consortium for their contribution.

http://www.trendsproject.org

7 REFERENCES


[5] McDonagh, D., Denton, H., 2005, Exploring the degree to which individual students share a common perception of specific trend boards: observations relating to teaching, learning and team-based design, Design Studies, 26: 35-53.


A Process-Based Model of New Venture Creation: Toward Modelling a Practical Application of Extant Theory using SADT Diagrams

T. Hernandez Rico, P. DuBois, M. Le Coq
Product Design and Innovation Laboratory (LCPI), Arts et Métiers ParisTech,
151 Boulevard de l'Hôpital, 75013, Paris, France
tatiana.hernandez@paris.ensam.fr

Abstract
Although the field of entrepreneurship abounds in studies attempting to explain the creation of new ventures from an array of theoretical perspectives, the answer to the critical question regarding "how" the process unfolds over time remains unsolved. The main goal of this paper is to demonstrate the importance of integrating an engineering point of view with the new business creation process in order to find the answer. This study will dig deeper into the issue by proposing the use of SADT (Structured Analysis and Design Technique) for modelling the "road map" that could assist entrepreneurs in dealing with uncertainties in a systematic and comprehensive way.

Keywords:
Venture Creation Process, Theory, Models, Practical Focus, Modelling, SADT

1 INTRODUCTION
Although the entrepreneurship field abounds in studies attempting to explain the creation of new ventures from an array of theoretical perspectives [1], the answer to the critical question regarding "how" the process unfolds over time remains unsolved [2]. To date, the entrepreneurial process has lacked a "road map" that could assist would-be entrepreneurs going through the proverbial "black box" between required inputs and desired outcomes [3] and dealing with the uncertainties which surround any new business creation [4].

According to Tötterman [5], the different ways to describe entrepreneurial processes originated from specific fields and each one of them has broadly strived to answer a set to four questions regarding the entrepreneur: The functional approach (what) derive from economics, the approaches focusing on the individual (why and who) from human sciences and the approaches on the processes (how) from management and organization sciences (Figure 1).

Consequently, if we make a logical link between our first statement of the unanswered question of "how" and the field that is in charge of finding the right answer, in this case Management and Organization Sciences, the result will be that the discipline is failing to do its job. Given the multidisciplinary nature of entrepreneurship theory, this fact uncovers the urgent need to introduce a new approach that could bring greater clarity about the absolute fundamental issues of entrepreneurship: what goes in, what comes out and how the transformation takes place [6].

Traditional pioneering studies that explore venture creation processes have used different terminology to describe the temporal sequences of events or activities that occur as entrepreneurs create a new organization. For example, Reynolds and Miller [7] prefers to use the term "Gestation process" which he defines as the moment between the principals elect to initiate a new firm and the new firm participates in the economy. Ultimately, the extant literature represents a great number of heterogeneous models whose key components demonstrate little uniformity other than patterns related to the life cycle stages (such as pre-venture, birth, growth, death) and only a few of them aimed at providing practical implications that address the "how" of entrepreneurship [6].

As a result of the above statement, the biggest challenge now is to explore the process with a pragmatic focus and empirically theorize from the ground of practice, which implies that scholars need to stop drawing conceptual models that describe the different stages and major issues related to the venture creation process using variance theory methods and qualitative case studies [1]. In this sense, built on the assumption that researchers must re-engage in open minded efforts at laying a foundation upon which extant work in the field of entrepreneurship may be successfully integrated [8, 9...
cited in [6], it might be possible that the key path to fulfill today's challenge could come from a different field. A field that uses different methodologies and tools and that will help us to integrate propositions from previous studies into a synthetic and applicable process-based model. Which science could hold the key to developing a process-oriented model capable of being used by entrepreneurs as a “road map” in a real basis, formalized in a logical pathway that can be followed, and that additionally proposes tools and controls to minimize the risks of looping fundamental data at each step of process?

In sum, a serious research gap exists regarding the process, methodology and tools for new business creation. The purpose of this study is to evaluate published and peer-reviewed models of entrepreneurial processes in order to discover key components; establish clear links or relationships between them; determine tools or criteria to go from one activity to another; and find out whether the study states practical implications of venture creation evidence by using the proposed model. Hence, this paper synthesizes research from a practitioner perspective and uncovers the importance of understanding the interrelationships between activities and the need for a well structured process model that prevent entrepreneurs from finding themselves repeating actions that could lead them to lose time and resources through the process of starting their own business.

This paper is structured as follows. Section 2 outlines the scope of relevant literature and attempts to provide a comprehensive theoretical view on differences among the extant models of the venture creation process. Section 3 discusses the gaps in the literature and identifies engineering sciences as the holder of the answer to the “how” question in entrepreneurship theory. Section 3 also considers the modelization of the process using SADT (Structured Analysis and Design Technique) as an option to synthesize and operationalize the process. The paper concludes with suggestions for future research studies.

2 LITERATURE REVIEW ON NEW VENTURE CREATION PROCESS MODELS

We start by defining the field of entrepreneurship as a scholarly examination of how, by whom, and with what effects opportunities to create future goods and services are discovered, evaluated and exploited [10]. Based on this definition, we establish that the central activity in entrepreneurship is the formation of new organizations. In this sense, all the functions, activities and actions associated with perceiving opportunities and creating organizations to pursue them, are denominated as the entrepreneurial process [11].

Only recently, attention has been given to the events involved in new business creation [7, 12, 13]. Different studies have attempted to explain it from an array of theoretical perspectives, such as economics [14], psychology [15, 16], population ecology [17, 18], ethics [19], Strategic management [20], Marketing [21], among others.

During the past few years several researchers [2, 6, 22, 23, 24, 25] have called for more process-driven research in order to better understand dynamic organizational processes. However, extant process-based models are far from being homogenous and a variety of alternative classifications may be done depending on the variables taken into account. For example, Tötterman [5] makes a classification of 22 scholars and their models based on a two-perspective view: Process models focusing on entrepreneurial opportunities and those that are focused on new entrepreneurial behaviour (Table 1).

### Entrepreneurial opportunities process models:
These models are focused on the process by which new goods, services, raw materials, markets and organizing methods can be introduced through the formation of new means, ends, or means-ends relationships. Then, the author makes a further classification into three differing opportunity perspectives: Allocative process view, Discovery process view and the Creative process view.

### Entrepreneurial behavior process models:
They focus on entrepreneurs as individuals and the activities undertaken by those individuals. The behavioral approach argues that it is central for entrepreneurship research to study what entrepreneurial individuals actually do and what they actually create when they shape new venture ideas.

<table>
<thead>
<tr>
<th>Table 1. Taxonomy of entrepreneurial process models based on Tötterman [5].</th>
</tr>
</thead>
<tbody>
<tr>
<td>On the other hand, we have the recent work done by Hindle and Moroz [6], where they develop a taxonomic matrix of entrepreneurial process models; then, a classification of 32 peer-reviewed journal publications and scholarly books published in the last forty years was done. Different from the two perspectives used by Tötterman [5], this study builds up the classification on four main types of entrepreneurial process models: Stage Model, Static Framework, Process Dynamics and Quantification Sequences (Table 2).</td>
</tr>
</tbody>
</table>

#### Stage model: divide into a priori stages major tasks or phases; One major weakness is that they tend to narrow the scope of investigation and that temporal orders of events do not fit the proposed stages and/or often overlap.

#### Static framework: characterizes the overall process of venture creation without examining the sequence of activities, consists of a limited set of variables connected by speculative causal links; process oriented but do not capture sequence of dynamics

#### Process dynamics: employs qualitative methods to examine how and why variations in context and process shape outcomes; often interpretive, temporal and change-oriented

#### Quantification sequences: is a historical sequence-based approach of the new venture creation process; this approach does not allow researchers to understand the dynamics of how antecedent conditions shape the present and the emergent future within the process

<table>
<thead>
<tr>
<th>Table 2. Taxonomy of Entrepreneurial Process Models [6]</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition, the authors [6] point out that there were only 7 models that explicitly stated practical implications for the research conducted: Bygrave [26], Carter et al. [27], Corbett [28], Cuneen and Mankelow [29], Fayolle [30], Sarasvathy [31], Spinelli et al. [32]. Based on the overview of extant models indentified by Hindle and Moroz [6], we now focus on the special characteristics of the 7 studies (Table 3).</td>
</tr>
</tbody>
</table>
Table 3. Characteristics of the 7 models that explicitly stated practical implications [6].

<table>
<thead>
<tr>
<th>Author</th>
<th>Type of Model</th>
<th>Key Components/Events/Stages/Domains</th>
<th>Variables/Factors/Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>[26]</td>
<td>Stage Model (4)</td>
<td>Innovation, Triggering Event, Implementation, Growth</td>
<td>Personal, sociological, environment, organizational</td>
</tr>
<tr>
<td>[27]</td>
<td>Quantification sequence</td>
<td>Up and running, Still trying; Given up</td>
<td>Bought equipment, got financial support, developed prototypes, organized start-up team, devoted full time, asked for funding, invested own money, looked for facilities, equipment, applied license/patent, saved money to invest, prepared plan, formed legal entity, hire employees, rented facilities /equipment, had sales, positive cash flow, credit listing, EI, FICA, filed tax</td>
</tr>
<tr>
<td>[28]</td>
<td>Stage Model</td>
<td>Discovery, formation</td>
<td>Preparation (deliberate, unintended), Incubation, Insight (eureka, problem solved, idea shared), Evaluation (recursive), Elaboration</td>
</tr>
<tr>
<td>[29]</td>
<td>Stage Model (4)</td>
<td>Opportunity recognition; opportunity evaluation; opportunity development; opportunity commercialization</td>
<td>Creative activity, innovative activity, strategic activity; Preliminary evaluation (personal, commercial), detailed situational analysis, formulation of mission and objectives, entry strategy, feasibility analysis, and BP, resources search, operational plans, implementation plans, secure funding.</td>
</tr>
<tr>
<td>[30]</td>
<td>Stage Model (2)</td>
<td>TRIGGER PHASE Act of new venture creation not perceived, perceived, considered, desired, COMMITMENT PHASE started, completed, perceived, refused.</td>
<td>Displacements, perceptions of desirability (culture, family peers, colleagues, mentors), perceptions of feasibility (financial support, other support, demonstration effect, models, mentors, partners), commitment; resource acquisition, integrating networks, structuring emerging organizations</td>
</tr>
<tr>
<td>[31]</td>
<td>Process dynamic</td>
<td>Inputs, effectual strategy: outputs.</td>
<td>What I know, who I am, whom I know, environment, constraints, expectations</td>
</tr>
<tr>
<td>[33]</td>
<td>Static Frame work</td>
<td>Opportunity, Resources, Entrepreneurial Team</td>
<td>Creativity, Communication, Leadership, Founder, Business plan (fits and gaps)</td>
</tr>
</tbody>
</table>

Finally, from this section we can conclude that, even though the literature is extensive in information about the stages and key activities in the process of starting a new business, there is not yet a dynamic method that could integrate all the stages/event/activities and help the nascent entrepreneur in dealing with the concrete actions that must be done from the idea in their head to the consolidation and further evolution of the new business.

3 DISCUSSION

3.1 Gaps in the literature

It is clear that the majority of extant literature focuses only on the theorizing power of business creation process models and there is an imminent need to explore other perspectives and alternatives that researchers in entrepreneurship have been neglecting.

In that respect, the following statement made by Van de Ven [3] is unequivocal and indicative: "...An appreciation of the temporal sequence of activities in developing and implementing new ideas is fundamental to the management of entrepreneurship, because entrepreneurs need to know more than the input factors required to achieve desired outcomes. They are centrally responsible for directing the innovating process with the proverbial "black box" between inputs and outcomes. To do this, the entrepreneur needs a "road map" indicating how and why the innovating journey unfolds, and the paths that are likely to lead to success or failure..."

Given the above, researchers should give considerable interest to developing a venture creation model that integrates all dimensions of the process to become a real "road map" that fills the gaps left by scientific literature. On the other hand, we have more recent studies like those of Liao et al. [34] that conclude that firm gestation is a complex, nonlinear process, rather than a simple, unitary accumulation of sequential events in which the developmental stages are hardly identifiable. In this sense, what they are suggesting us is to think that the process is so complex and ambiguous that planned actions may not lead to desired responses.

Similarly, we can take for instance the innovation process and all its different theories, models, principles, methodologies, and so on. In a practical ground, what reality has shown, is that one of the major impediments in the innovation process is the belief that invention cannot be systematic and be based on scientific principles [35].

In addition, most of the scientific studies are fragmented, descriptive, and focused only on a few aspects of the new venture creation process. More importantly, most of the literature has not paid adequate attention to the needs of the entrepreneur - the main beneficiary of the model. Nor has the literature placed proposed methodologies of the models in practice.
3.2 Bridging the gaps in the literature: A call to action for engineering sciences and SADT modelling

According to Bygrave [33], entrepreneurship research has emerged by using methods and theories from other sciences, but in order to become a distinct discipline it needs to develop its own methods and theories. On the contrary, based on the gaps found in the literature — notably the need for pragmatic research and empirical evidence — and the recent studies that outline the urgent need to develop a harmonized model of entrepreneurial process capable of embracing the best of what is on offer and adding new theoretical arguments in areas where practice shows that they are lacking [6], this study believes that engineering sciences may hold the key to resolving the limitations of current process-based models.

Subsequently, due to the ineffectiveness of existing models the engineering perspective might respond — as the starting point — by engaging in the use of a modelling language that respects the following requirements on the basis of the literature review:

- It must be able to model complex and dynamic systems.
- It must be able to focus on elements of the system without losing the links to the whole model.
- It must provide a common language to describe and model all aspects of the system.
- It must be compatible with existing ideas and principles in economics, management and organizations and human sciences.
- The resulting model must provide a normative statement about the way in which the venture creation process should be structured and operate.
- It must allow the formalization of functional interactions and the identification of information flow.
- It must integrate rigor and control in the process of analysis and a method for using the modelling language.

As a result, based on these requirements, and a review of modelling methods and languages that have been developed so as to model business processes [36] - SADT, IDEF3, BPMN, FBSPRE-, we have chosen the use of SADT since it not only describes the tasks involved in a project and their interactions, but also describes the system that the project aims to explore, create or modify, highlighting the different parts which constitute the system, their purpose, their operation and the interfaces between the various parts which let us see the system as more than a mere collection of independent elements [37].

Each SADT diagram is composed of boxes (representing activities) connected by arrows (representing flows of materials, data, or information) and provides a robust structured method to model hierarchical systems [38].

SADT models are composed of Inputs (needed data), Outputs (produced data), Controls (commands that influence the execution of the activity) and Mechanisms (means, components or tools used to accomplish the activity), and uses several hierarchical blocks (Figure 2).

The A0 block is the top-level, which presents the overall system. This block can be broken down into lower levels in order to describe the subsystems, or in other words, the parts that make up the overall system.

3.3 The elements of the model

For instance, taking into account that the use of SADT lets us focus on specific parts of the process without losing their relationship to other parts, and at the same time we are able to make links to other field’s principles like management, organizations and economics, we cannot be deaf, dumb and blind to what the market has to offer regarding the best-selling phrase: “How to start your own business”.

In this sense, nascent founders have to face the challenge of searching among millions of books or internet links that offers the “magic recipe” on “How to start your own business”, and more certainly if they ask for advice on how to increase a venture success, a likely response is “Start planning”, given the fact that universities around the globe teach students in numerous entrepreneurship classes about the importance of preparing business plans and how to write them. Store bookshelves abound with books on how to prepare a business plan [39] and 10 million business plans are written each year worldwide [40]. But, what if that is not really a requirement or the “answer” for success? After all, some of our role models today, such as Bill Gates (Microsoft), Steve Jobs (Apple), Michael Dell (Dell), and Sergey Brin and Larry Page (Google), did NOT have business plans in hand when they started their companies.

On the other hand, regarding the content of the business plan, there is a plethora of books, consultancy services, do-it-yourself software, government support agencies and universities that explains how to write this document, typically suggesting that business planning is valuable and important for new firms [41-45]; but whose number of chapters and methods varies from one to another. However, and most important, all of them present a common factor that, as Gruen’s research findings clearly indicate: “Handbooks typically focus on the content of business plans (and neglect the process), and offer a fairly standard, “one-size-fits-all” notion of planning” [46].

In sum, what we can retain from the literature regarding the business plans is that it does not present a complete process for new venture creation and instead the different sections or chapters resemble bits or pieces of processes without neither a clear connection between them, nor a logical pathway to follow. Indeed, this is mainly the reason why entrepreneurs find themselves repeating actions and information through the business creation process.

According to Van de Ven and Poole [47] and Aldrich [18], process theories generally have distinct sequences and mechanisms which explain how and why various changes occur and why certain processes progress. In just the same way, entrepreneurs organize new firms through a series of actions and they are undertaken to different degrees, in different order, and at different points in time [48].
To date, Delmar and Shane [49] have been among the first researchers to emphasize that by engaging in different patterns of activities, firm founders will create variation in the firm formation process. Then, in order to understand how this variation occurs, we have to first focus on the characteristics defining the evolution of the process, and second on the purpose of the different activities in which firm founders can engage.

Moreover, there are four different characteristics that dictate how the patterns of activities evolve [49]:

- First, not all activities are necessary for the founding team to perform.
- Second, due to the firm founders' limited cognitive capacity, they lack the ability to undertake all organizing activities simultaneously.
- Third, the ability to undertake some activities is dependent on or will be enhanced by the completion of other activities.
- Fourth, some activities are more important early in the history of new venture, others are more important later in the life history of the new organization.

Most important, this study distinguishes two (2) different types of activities: Planning and Operational activities [49]. Planning activities refers to events that coordinate different activities at the early stage of venture creation; and Operating activities can be in turn divided into legitimacy building activities, resource transformation activities and market-related activities.

Finally, in line with the idea of developing a categorization of the different types of activities involved in the business creation process, this paper proposes a more comprehensive classification that basically distinguishes 3 types of activities: Product/Service; Market; and People and Operations activities.

Having reviewed the literature in section 2 and based on the discussion presented in Section 3, a descriptive model of the new venture creation process is proposed as a linkage between theory and practice in the entrepreneurship field. This model represents a concrete contribution from engineering science to help answering the question of how the process unfolds over time.

The detailed process-based model using SADT diagrams is not presented here, since it is the subject of an entire paper in progress at the present time; however, we present a partial graphical representation for illustrating our approach (Figure 4).

4 CONCLUSIONS AND PERSPECTIVES

With the ever growing infant mortality of Small and Medium Sized Enterprises (SMEs) around the world, whereas only 40% to 50% of firms created in a given year survive beyond the seventh year [50], the contributions made by economics, management and organizations and human sciences have shown their limits as approaches to help and assist the entrepreneurs in starting their own business. One of the main contributions of our study is to have highlighted the lack of practical implication of existing models of business creation and to have identified the potential of engineering sciences as a highly valuable contributor to the domain. To some extent, our analysis offers new perspectives for the normative literature and for practitioners.

We have also proposed SADT modelling to represent the “road map” in which the venture creation process should be structured and should operate. Moreover, the formalization of functional interactions and the identification of information flow inside the process will become the key element to avoid repetitive actions and/or analysis that lead entrepreneurs to lose time and resources through the process and that eventually affect the chances of success.

Finally, our model highlights major opportunities for future research to explore potential use of our 3-axis framework regarding activities related to the integration of products/services, markets, people and operations. Furthermore, in a short term, our model will be tested by a group of engineering students involved in a business creation course run by author’s laboratory in Arts et Métiers ParisTech to make a preliminary assessment to discover the strengths of our model and to point out uncovered areas.

5 REFERENCES


management. Application to ship equipment part families. CIRP Journal of Manufacturing Science and Technology 1 (2009) 185–190


Function Representations in Morphological Charts: An Experimental Study on Variety and Novelty of Means Generated

J. L. Richardson III, J. D. Summers, G. M. Mocko

Mechanical Engineering, Clemson University, Fluor Daniel Building, Clemson, SC, 29634-0921, USA
jenkinr, jsummer, gmocko@clemson.edu

Abstract
A systematic study of the introduction of a new function representation, function structures, into morphological charts is explored in this research. This is a first step in developing systematic guidelines to aid student designers in using the combinatorial idea generation tool in an effective and efficient manner. An experiment is conducted with senior mechanical engineering students to determine how novelty and variety of means generated from morphological charts is influenced by different function representations (function lists and function structures). The findings from this experiment suggest that the variety and novelty of means generated in a morphological chart is not dependent on the function representation. More significantly, this study provides a template on which future experimental studies on idea generation can be based, thereby increasing the potential for discovery and development of systematic guidelines to improve conceptual design for novice engineers.

Keywords:
Morphological chart, function structure, conceptual design

1 EXPERIMENTAL MOTIVATION
The objective of this research is to determine the relationship between function structures and morphological charts in the context of idea generation effectiveness with an emphasis on the novelty and variety of means generated. Combinatorial design tools, such as morphological charts, are prevalent in engineering design textbooks [1-4], yet there is no concrete evidence as to how effective these tools are with respect to novelty and variety of concepts actively pursued. Moreover, there is little guidance provided about how to represent and model the functions that are being satisfied by the various means that are generated. To begin to fill these noticeable gaps in the literature, a user study of senior mechanical engineering students using function structures and function lists incorporated into morphological charts is conducted.

The design process is often decomposed into four general phases: problem clarification, conceptual design, embodiment design, and detail design [1-2]. Sometimes the embodiment and detail design phases are combined into a single product development stage [3]. However, the conceptual design stage is recognized to be of special interest in which it is estimated that 70% of the life cycle cost of a product is determined during the conceptual design phase [5]. Further, the NRC identified several goals of engineering education including teaching students the basic tools of the design process. This aligns with more recent reports that call on the National Science Foundation (NSF) to concentrate on supporting research that explores early stage engineering design [6]. A subset of these basic tools used during the early design process phases includes idea generation methods [7-8].

The conceptual design stage can be decomposed further into four steps: decomposition, sub-solution generation, concept integration, and concept evaluation. Once the design problem is identified and understood or defined, function decomposition begins. Functional decomposition details the high level and lower level necessary functions of the product or artifact being designed. These functions are driven by the requirements as defined in the problem definition and clarification phase. After the sub-functions are identified, means are generated which satisfy each of the sub-functions individually. Numerous possible means are typically sought in order to more fully explore the feasible design space. Next, means are selected and combined based upon compatibility; along with some additional engineering synthesis, the combined means form integrated design concepts. Finally, multiple integrated design concepts are evaluated and a principal solution set is selected to be further explored through embodiment and detail design.

Idea generation methods are used typically within the means generation and combination stages of conceptual design. Morphological charts [1,4,9] are intuitive idea generation methods [7-8] which may be used by individuals or groups. Moreover, morphological charts are able to support both the means generation activity and the integration of the means to form solution concepts. Thus, this tool is recognized as powerful engineering design tool as evidenced by the fact that morphological charts are found in a plurality of the popular engineering design textbooks [1-4].

1.1 Morphological Charts
A morphological chart, also known as concept combination tables [10] or function-means tables [4], is a tool for systematic combination of solutions to a design problem [1]. A common organizational structure of a morphological chart is shown in Table 1.

<table>
<thead>
<tr>
<th>Function</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>F₁</td>
<td>M₁₁, M₁₂, M₁₃, ... M₁₄, M₁₅</td>
</tr>
<tr>
<td>F₂</td>
<td>M₂₁, M₂₂, M₂₃, ... M₂₄, M₂₅</td>
</tr>
<tr>
<td>F₃</td>
<td>M₃₁, M₃₂, M₃₃, ... M₃₄, M₃₅</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Fₙ</td>
<td>Mₙ₁, Mₙ₂, Mₙ₃, ... Mₙ₄, Mₙ₅</td>
</tr>
</tbody>
</table>

Table 1: A Morphological Chart

The morphological chart is represented as a table of decomposed sub-functions of the design problem and potential solution fragments for each sub-function.
Common convention lists the set of decomposed sub-functions of a problem in the first column of the table and the solution fragments (means [4], working principles [1], and design parameters [9]) to realize each of these sub-functions cells to the right of each. For this research, the solutions to each function will be referred to as means and the term concept will describe a set of means which collectively satisfy a majority of the sub-functions identified in the morphological chart.

In Table 1, the morphological chart is sized to be $n \times m$ where $n$ represents the number of sub-functions and $m$ represents the number of means. It is important to note that a morphological chart is not constrained to have the same number of sub-functions and means. The sub-functions, represented as $Fn$, that are used in the creation of the morphological chart should all be at the same level of detail [4]. For each sub-function, the means, listed by $M_{n,m}$, generated should typically be at the same level of abstraction, though some recent work suggests methods to accommodate varying levels of abstraction [11]. By combining one means for each function, a concept is created. Repeating this process with every possible combination contained in the morphological chart creates an exhaustive list of concepts. In this manner, morphological charts provide a sense of the size of the design space [4].

In an attempt to provide more guidance in how to control the design space to improve quality integrated concept exploration beyond a feasibility check, researchers have studied how the number of means and the number of functions influence concept quality [12-13]. The findings from this experimental study indicate that a chart with more means than functions produced higher quality concepts than a chart with more functions than means and that adding functions to a morphological chart did not improve the results [12]. Thus, rectangular morphological charts with more means than functions is preferred to a rectangular morphological chart with more function than means. This previous research appears to be one of the first to systematically investigating the construction and use of morphological charts. Therefore, it is the first step towards providing designers with specific guidelines for more effectively using a morphological chart to produce useful concepts.

In addition to the size of the morphological chart, the actual representation of the design problem may be influential on how easily the design space can be explored. This belief has yet been tested experimentally in the literature and specific designer guidelines are not yet available.

Further, this experimental study demonstrates the potential for refining and defining morphological chart guidelines. It is on this foundation that this paper explores additional approaches to refine morphological analysis methods for ideation improvement.

Advantages and Disadvantages

As with all design tools, morphological charts have several advantages and disadvantages which influence the context of use. Advantages of morphological charts include their ability to illustrate unexpected pairings of features [10], the potential creation of novel concepts not otherwise considered by the designer [4], and the capability to represent and explore large regions of the design space. Three specific limitations to morphological charts as design tools are the potential for the number of concepts to grow exponentially making exploration difficult [4], the reality that not all combinations of means will be feasible solutions to the design problem [4], and the absence of a set of guidelines to determine a useful way to choose the promising concepts for further evaluation.

The goal of this research is to improve first on the representation and exploration of the design space by increasing the quality of the means. Secondly, this improvement will be implemented through a set of specific guidelines for use with morphological charts. In this manner, some existing advantages will be enhanced and a current limitation will be addressed.

**Computation Automation of Morphological Charts**

Several examples of research into automating the exploration of the design space represented by morphological charts exist. Bryant, et. al. [14] describe an interactive, user interface driven approach which is the result of combining two previously developed and validated tools: an automated morphological search [15] and a computational concept generator [16]. The automated morphological search is a web-based tool that makes use of the information of the Design Repository [17] to populate a morphological chart. The computational concept generator takes a user-defined function block diagram of a product and converts it into a matrix which describes relationships between functions. Based off of information contained in the Design Repository, a function-combination matrix (FCM), and a design structure matrix (DSM), a list of possible solutions is created, filtered, and presented to the user. The proposed interactive morphological search is created by combining characteristics of each of these two tools. The hybrid technique has the connectivity information generated in the computational concept generator and the solution accessibility of the web-based morphological chart search. The major limitation of this method is the amount of design knowledge currently entered into the Design Repository.

Tiwari, et. al. [18] discuss using a genetic algorithm to combine means from a morphological chart into solutions. This method represents the means combination process as a combinatorial multi-objective optimization problem. The method allows for multiple criteria to be used in judging the combinations. The optimal combination will be a balance of high performance in each criterion. Advantages to this method include minimal computational effort, the utilization of a multitude of information from the designer, quick feedback to the designer from a large pool of potential solutions, and consistent results despite uncertainty in the inputs (i.e. a range of values instead of a single value). The limitations to this approach include the number of inputs from the designer, a lack of non-behavioral characteristics (i.e. aesthetics) as inputs, and the chance of variation among inputs (inconsistencies between multiple designers).

**1.2 Function Structures**

A function is defined as the intended input/output relationship of a system which performs a certain task [1] or the “desired output from a mechanism” [3]. The existence of a product is justified by its’ functions [2]. Therefore, when designing a product, it is often recommended to model the product using its functions during the conceptual stage [1-3]. Although there are several methods available to model the functions of a product (including function structures [1-2], the Function-Behavior-Structure model [19], the Function-Behavior-State [20], the Structure-Behavior-Function [21], and the affordance-based view of functionality [22]), this research will focus attention on function lists and function structures only.

Function lists and function structures are similar in that they are both form-neutral representations of a product and describe the functions that must be accomplished but
not how the detailed solution will be realized. The primary
differences between a function list and function structure
are: (1) a function structure is a graphical representation,
whereas the functional list is a textual representation and
(2) the function structure explicitly captures topological
connectivity between the functions whereas a function list
can only imply order. Based upon these similarities and
differences, it is recognized that a function list can be
derived from a function structure but the converse is not
true, therefore implying that a function structure is an
advanced representation.

A function structure consists of function blocks (verbs)
which are connected by flows (nouns). Previous research
has shown that function structures are an acceptable way
of modeling the functions and relationships between
functions of a product [1]. To create a function structure,
the first step is to identify the basic inputs and outputs of
the system, based on the customer needs or problem
statement. These inputs and outputs can be arranged into
a black box model [23]. A black box model is used to
define the relationship between the inputs and outputs of
the system [2]. An example of a black box model is shown
in Figure 1. Within this model (and all function structures)
the inputs and outputs can be categorized as material,
energy, or signal flows. Examples of each type of flows
include: gases, liquids, solids (material); mechanical,
electrical, thermal (energy); and magnitude, control, data
(signal).

As shown in the example, the main function of the
vacuum cleaner is “transport dirt off the floor”. This model
has four total inputs (electricity, hand, debris, air) and five
total outputs (noise, heat, hand, debris, and air). To go
from a black box model to a function structure, the inputs
and outputs remain the same but the function is
decomposed into verb-noun pairs (functions and flows). In
Figure 2, the vacuum cleaner’s black box model has been
expanded into a function structure. The single, original
function “transport dirt off the floor” has been decomposed
into fifteen sub-functions.

1.3 Summary
The motivation for this research began with an overall
look at the design process and selects the conceptual
design stage as a focus. Further investigation into
concept integration, specifically using morphological
charts, and function representations followed. The
literature reviewed in this motivation can be summarized
as follows:

- Four steps in the design process – problem
  clarification, conceptual design, embodiment design,
  and detail design
- Conceptual design has been identified as a can
  further be decomposed into decomposition, sub-
  solution generation, concept integration, and concept
evaluation
- Morphological charts are a intuitive design tool used
  in concept integration
- Common convention uses a function list (textual
  representation) to express the intended functions of a
  product in a morphological chart
- Function structures are graphical representations of
  a product which are not currently used in
  morphological charts

2 EXPERIMENT
If a design engineer is to use a morphological chart to
generate means and integrate concepts, it would be
useful to know how the representation of the design
problem impacts the exploration of the design space and
the quality of the concepts developed. To determine this,
an experiment is conducted in which two different
morphological charts are used: one using a function list to
represent the problem and one using a function structure.
The two morphological charts provide the participants the
opportunity to produce the same number of means. After
the participants generated means for the different charts,
the participants used the charts to form concepts with an
emphasis on identifying high quality concepts. These
means and concepts were then evaluated to determine
which configuration of morphological charts yielded the
higher quality means and concepts. Each of the aspects
of the experimental design as summarized is discussed.

Figure 1: Black Box Model of Vacuum Cleaner [2]
Figure 2: Function Structure of a Vacuum Cleaner [2]
Table 2: Experiment Summary

<table>
<thead>
<tr>
<th>Participants’ Class</th>
<th>Group #1</th>
<th>Group #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Statement</td>
<td>Automated burrito folder</td>
<td></td>
</tr>
<tr>
<td>Number of Participants</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Data Collection (Means)</td>
<td>Function list</td>
<td>Function structure</td>
</tr>
<tr>
<td>Data Collection (Concepts)</td>
<td>Table with three empty columns</td>
<td>Three blank function structures</td>
</tr>
</tbody>
</table>

2.1 Participants

The participants of this experiment were drawn from a homogeneous population based on educational background; they were all students enrolled in a required senior level mechanical engineering capstone design course at Clemson University. The course is in the last semester of their undergraduate program. Thus, they are within months of being practicing, albeit novice, professional engineers. The participants had been exposed to morphological charts during a common normalizing lecture. Further, only a few of the participants had previous experience from previous courses. Within this course, not all of the participants had used morphological charts in their semester long design projects. The participants’ previous experience with morphological charts varied from never having seen morphological charts (three of fifty participants) to having used a morphological chart at least once to generate concepts (forty of fifty participants).

All of the participants were in a single section of the class, so the experiment was able to be conducted in a common setting and time period. To mitigate discomfort to the participants in order to achieve unbiased results, the experiment was conducted in the normally scheduled classroom during the normal class time. Once all participants entered the room, they were arranged so that there were an equal number of students on each side of the aisle (25 students per side).

2.2 Problem Statements

The problem chosen for this study is the design of an automatic burrito folding machine. This problem statement was adapted from a previous project in a sophomore mechanical engineering course. Earlier research on morphological charts [13] had adapted the original class project problem statement (which encompassed several weeks of a semester) into a problem statement more suited for the time frame of this study (a single class period). Using this previous work as a guideline, an overall problem statement for the experiment was created.

This overall problem statement is composed of the general problem statement, the five functions that the burrito must satisfy, the instructions for means generation, and the instructions for means combination. The following sections outline each part of the overall problem statement.

General Problem Statement

This particular design problem is chosen because the scope is similar to what the participants would have experienced in other classes. As it was previously used as a sophomore design project, the difficulty and complexity of the problem was not considered too challenging for the time given and for the participants’ level of expertise. None of the students in the senior class had taken the sophomore class when this problem was used. Moreover, the problem was generally novel enough such that the participants would have had little preconceived thoughts about in advance of the study is presented. Therefore, it is important to note that the time elapsed between the assignment of the original problems statement and the conduction of this experiment is such that there is no overlap in participants. By choosing a problem which all participants have an equal understanding in the beginning, an additional variable of previous knowledge of the problem is not introduced. The general problem statement shown in Figure 4 establishes the need for the novel concept and lists some of the materials involved in the problem.

In general, the food service industry has a great need for speed, efficiency, and cleanliness as preparing large amounts of food quickly is their main goal. As a result, a local restaurant has identified the need for a machine to fold their burritos. Each burrito is made up of a ten inch tortilla shell and 2 ounces of filling.

The restaurant has identified the five main functions that the burrito folding machine must accomplish.

Figure 4: General Problem Statement Provided to Experiment Participants

Functions of the Burrito Folder

After this general problem statement, the five functions that the burrito folder must perform are presented in one of two ways, as a function list or as a functions structure. Each participant is given one of the two representations of the functions.

The function list is as follows:

- Store Filling
- Position Tortilla
- Fill Tortilla
- Fold Burrito
- Dispense Burrito

In the function list, each function of the burrito folder is represented as a verb-noun pair. While all of the verbs are unique, some of the nouns are repeated between functions.

Figure 3: Function Description of the Problem Statement Given to Groups #2 of Participants (Function Structure)
These functions have been arranged into a function structure as in Figure 4. In the function structure, the function blocks contain the same verb-noun pairs seen in the function list. The key differences between the function list and the function structure are the graphical representation of the functions and the flows between the functions.

**Instructions for Means Generation**

After the participants are given the functions that the burrito folder must perform, they are instructed to populate the morphological chart as shown in Figure 5. The participants are explicitly reminded that price, number of components, and ease of use will be used as criteria against which to judge the quality of the generated means and concepts.

You will generate ideas for performing each function through the use of a morphological chart. Each function will have its own row in the chart with space to provide up to six means to perform the task. The means will be evaluated in terms of

- price
- number of components
- and ease of use.

Please keep these criteria in mind when designing your product.

**Figure 5: Instructions for Generating Means in Morphological Chart Provided to Experiment Participants**

**Instructions for Means Combination**

After means generation, the individual means from the morphological chart are integrated into concepts. The participants are asked to generate three concepts from the means listed in their own populated morphological charts. As seen in Figure 6, the participants are instructed to develop three concepts from their means. Again, the participants are reminded that price, number of components, and ease of use are the criteria in the problem. Based on the results from the previous work with a similar scope and problem statement [13,14], this is considered to be an appropriate number of concepts for the students to generate.

Using the Morphological Chart that you created please develop three concepts for performing the task of folding a burrito. As shown in the example you do not need to rewrite each of your means. Please refer to the means using the FX.Y notation where X is the function number and Y is the mean number. The concepts will be evaluated in terms of

- price
- number of components
- and ease of use.

Please keep these criteria in mind when composing your concepts.

**Figure 6: Instructions for Generating Concepts Provided to Experiment Participants**

**2.3 Data Collection**

In addition to the problem statement previously outlined, documents to capture the data created by the participants were created. First, two blank morphological charts were formed for the participants to record the means to perform each function. Participants were given the morphological chart which coincided with their problem statement. In the first morphological chart, the functions list is provided in the first column with the remaining columns available for participants to record their means. For the second morphological chart, a function is listed in each cell of the left hand column along with a miniature figure which shows the participants where that function occurs in the function structure.

In addition to the instructions for combining means into concepts, each group of participants was given space to record their concepts. For the function list participants, a table with the function list in the first column, an example concept in the second column, and three additional columns for the participant’s concepts was created. For the participants with the function structures to record concepts, a series of four function structures was created. The first function structure served as an example concept while the three remaining function structures were blank for the participant to record their own concept.

**2.4 Procedure**

To begin the experiment, a brief introduction of morphological charts by the author and the use of the data from the experiment were explained to the participants. Next, two handouts were distributed: one handout detailing the problem statement and one handout with an empty, unfilled morphological chart. As previously stated, the participants were physically split in the classroom by an aisle so that 25 participants received problem statements with the function list and 25 participants received problem statements with the function structure. Although the participants were grouped, the problem statements were randomly assigned as there was no control over which side of the classroom the participants were assigned at the beginning of the class. The first handout was read aloud to the participants and time was given to allow the participants to ask clarification questions about the two problem statements.

Next, the participants were given twelve minutes to generate means in their respective morphological charts. As the participants were completing the morphological charts, the author walked around the room to keep the participants on task, confirm that twelve minutes was an appropriate amount of time for idea generation, and make observations about how the participants were generating means (by row, by column, or randomly). Once the twelve minutes were finished, a third handout containing the appropriate figures (function structures) or table (function lists) was distributed. The function structures and function lists were not mixed. For example, participants with function structures on their first handout received function structures on their second handout. The third handout defined a space for the participants to generate concepts to solve the design problem based on the means previously generated. The author read the instructions aloud and allowed the participants to ask questions. The participants were permitted five minutes to complete the concept integration. Once the participants were finished, all papers were collected.

**3 PROTOCOL FOR ANALYSIS**

The analysis of the data collected will focus on two metrics of effectiveness (variety and novelty [24]). In order to prove that function representation does not affect the means generated, all means generated by the participants will be evaluated for variety and novelty, then the results from each group of participants will be compared. The protocol for analysis is outlined in the following sections.

**3.1 Variety**

The means for each function will be organized in to classifications. In order to do this, the author begins by
compiling all of the means entered into the morphological chart by function. For each function, the means are divided into classifications which represent similar means (i.e. for the function "store filling", a classification of *individual portions* would encompass *individual bags, individual tubes, individual containers and pre-packaged portions*). Variety is determined by counting the number of classifications represented for each function. Measures of variety will be observed both by participant and by function. Classifications which appear most frequently in the morphological charts will be identified. A comparison between the most popular classifications of means for each function will be considered for the two groups of participants.

### 3.2 Novelty

Novelty of the means generated will be based on the classifications of functions previously mentioned. A means will be considered novel if it does not correspond to any of the classifications and only appears in one participant’s morphological chart. Measures of novelty will be observed by participant, by row of the morphological chart (function), and by column of the morphological chart. Functions which encourage novel means will be identified. In addition, a notion for whether novelty occurs more in early means generation (first three columns) or late means generation (last three columns) will be formed.

#### 3.3 Summary

Once the user study is complete, the data collected is organized and recorded for analysis. The analysis presented in this paper explores the ideation metrics of variety and novelty [8]. Quality and quantity from this experiment are found in [25]. Variety will show trends in the classifications of means entered in the morphological charts. Novelty will be used to determine which functions encourage innovative thinking and where (which column) this innovative thinking is most likely to occur. The research hypothesis is that the function representations will not impact the variety or novelty of means generated within the morphological charts. This is due to the fact that the individual means are generated for specific functions and the interconnectedness or topological information which distinguishes function lists and function structures should have no influence.

### 4 ANALYSIS AND RESULTS

The analysis of the results of the experiment will follow the protocol outlined in Section 2. Each of the two metrics (variety and novelty) will help to determine if introducing function structures into morphological charts had any secondary effects on the means and concept generation process.

#### 4.1 Variety

As previously mentioned in Section 2, the means of the morphological charts were divided into classifications for each function to ensure the reliability of the morphological chart scoring. With these classifications created, it is possible to gain insight to how the means which were entered into the morphological charts varied between the two groups of participants. To begin this classification process, means with the same wording which appeared in more than one morphological chart were distinguished as the original classifications. In general, this step of the classification process was capable of organizing about a quarter of the entries in the morphological charts. Next, means which were synonyms to the classifications were identified. For example, for the function *dispense burrito*, means such as "ramp" and "chute" were identified as part of the classification "slide". After this step of the classification process, approximately three quarters of the means generated had been classified. The final step in the process involved connecting means which were semantically equivalent to the classifications. For example, for the function *position tortilla*, a means such as "semi circle lip" is considered semantically the same as the classification of "mold". While the first two steps of the classification process are mostly objective, the final step of the classification process is somewhat subjective as the author must make determinations of the intentions of the participants as they recorded the means.

Once all of the means are classified, the variety of classifications represented by each participant is determined. By counting the number of classifications used by each participant, the impact that the function representation had on the variety of classifications used can be quantified. A summary of the number of classifications used by each group of participants for each function is shown in Table 3.

<table>
<thead>
<tr>
<th>Function List</th>
<th>Function Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Store Filling</td>
<td>3.20 1.35 2.56 1.00</td>
</tr>
<tr>
<td>Position Tortilla</td>
<td>2.88 1.20 2.52 0.96</td>
</tr>
<tr>
<td>Fill Tortilla</td>
<td>3.08 1.12 2.84 1.18</td>
</tr>
<tr>
<td>Fold Burrito</td>
<td>2.40 1.04 2.44 0.87</td>
</tr>
<tr>
<td>Dispense Burrito</td>
<td>3.12 1.33 2.92 1.32</td>
</tr>
</tbody>
</table>

Table 3: Number of Classifications of Means Appearing per Participant

From these results, there is little evidence of a significant change in the average number of classifications which participants used comparing the two function representations. For the first function, *store filling*, there seems to be a noticeable difference between the function list and function structure participants. Further review of the classifications (and their frequencies) is necessary.

The number of means generated for each function by the two groups of participants is shown in Table 4. Also, the percentage difference between the number of means generated by the two groups of participants is calculated and shown. Three of the functions (*store filling, fill tortilla, and fold burrito*) do not show a significant difference in the number of means generated. The two functions which show a significant difference in the number of means generated are *position tortilla* and *dispense burrito*.
In Figure 7, the classifications (and their frequencies) for the function store filling are shown. For both groups of participants, the frequent classifications include “individual portions”, “bucket/basket”, “bottle/tube”, “bag/bladder”, and “tank”. The function structure group of participants identified “bag/bladder” significantly more than any other means by either group for any function. This is the root cause for the difference between the two groups of participants in average classifications appearing per participant for this function (as seen Table 3). A deeper investigation into why this was the case might provide insights as to whether the representation has an influence on types of means generated, but is currently out of scope for this paper.

Table 4: Number of Means Generated for Each Function

<table>
<thead>
<tr>
<th>Function</th>
<th>Function List</th>
<th>Function Structure</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Store Filling</td>
<td>88</td>
<td>92</td>
<td>5%</td>
</tr>
<tr>
<td>Position Tortilla</td>
<td>96</td>
<td>83</td>
<td>14%</td>
</tr>
<tr>
<td>Fill Tortilla</td>
<td>90</td>
<td>91</td>
<td>1%</td>
</tr>
<tr>
<td>Fold Burrito</td>
<td>75</td>
<td>71</td>
<td>5%</td>
</tr>
<tr>
<td>Dispense Burrito</td>
<td>94</td>
<td>79</td>
<td>16%</td>
</tr>
</tbody>
</table>

In Figure 8, the classifications (and their frequencies) for the function fill tortilla are shown. For both groups of participants, “manual”, “scoop”, and “squeeze/caulk gun” are the most frequent classifications. “Vacuum/pneumatics” is much more frequent for the function list participants than the function structure participants.

In Figure 9, the classifications (and their frequencies) for the function fold burrito are shown. The two groups of participants generated a similar number of means and shared the most frequent classifications: “flaps/fold” and “linkage/lever”.

Figure 7: Frequency of Classifications – Store Filling

In Figure 8, the classifications (and their frequencies) for the function position tortilla are shown. For both groups of participants, the most common classifications are “conveyor”, “mold/fixture”, and “arm/linkage”. The difference in number of means (13 more means generated by the function list participants) is largely accounted for in the classifications “slide/ramp”, “laser/suction”, and “laser/optics”. These classifications were moderately frequent for the function list participants but nearly non-existent for the function structure participants.

Figure 8: Frequency of Classifications - Position Tortilla

In Figure 9, the classifications (and their frequencies) for the function fill tortilla are shown. For both groups of participants, “manual”, “scoop”, and “squeeze/caulk gun” are the most frequent classifications. “Vacuum/pneumatics” is much more frequent for the function list participants than the function structure participants.

Figure 9: Frequency of Classifications - Fill Tortilla

In Figure 10, the classifications (and their frequencies) for the function fold burrito are shown. The two groups of participants generated a similar number of means and shared the most frequent classifications: “flaps/fold” and “linkage/lever”.

Figure 10: Frequency of Classifications - Fold Burrito
In Figure 11, the classifications (and their frequencies) for the function dispense burrito are shown. Although the function list group of participants generated significantly more means than the function structure group of participants for this function, the two groups of participants share the most frequent classifications: “conveyor”, “slide/ramp”, “manual (by hand)”, “actuator”, and “mechanical arm”.

Figure 11: Frequency of Classifications - Dispense Burrito

Using the classifications created as described in this section, means which are novel ways of satisfying each function can be identified. The following section will detail the process of identifying these means and which, if any, functions were more likely to produce novel means.

4.2 Novelty

For this research, a novel means is any means which does not fit into a classification and does not appear in more than one participant’s morphological chart. While the extent to which these means would satisfy the function is uncertain, any means which fit this description is considered novel. In Table 5, each novel means which was identified is shown. The functions which produced the most novel means were fill tortilla, fold burrito, and dispense burrito while position tortilla produced the fewest novel means. Although “shaker” and “vibrating plate” are similar semantically, they appear as means to satisfy two different functions (and were listed by two separate participants), so they are still considered novel. In addition, “bucket system” and “gondola bucket system” are also considered novel for the same reasons. Of the fifty participants, only one participant generated more than one novel means (“shaker” and “ladder”).

For store filling, participants from the functions structure group identified “large heated griddle” and “just in time from supplier” as a means to satisfy the function, while there were no novel means produced by the function list group of participants. For position tortilla, a participant of the function list group identified “shaker” as a means to satisfy the function, while there were no novel means produced by the function structure group of participants. For fill tortilla, a participant from the function list group identified “bowl feeder” while participants from the function structure group identified “bucket system” and “toaster-style process”. For fold burrito, a participant from the function list group identified “tooth pick” and “edible happy face stickers” as means to satisfy the function, while a participant from the function structure group identified “slap bracelet”. For dispense burrito, a participant from the function list group identified “vibrating plate”, while participants from the function structure group identified “ladder” and “gondola bucket system”. Only twelve means out of the 859 means generated are considered novel. Although the function structure participants generated more novel means than the function list participants (7 to 5), there is not enough evidence to confirm a significant difference between the two groups.

5 CONCLUSIONS

The research presented details the design and execution of a user study to determine the relationship between function structures and morphological charts in the context of idea generation effectiveness with an emphasis on the novelty and variety of means generated. The experimental method, protocol of data analysis, and the results of the study are described. Based upon the results presented in this paper, there is no statistical evidence which shows a difference in the novelty and variety of means generated using function lists and function structures. Therefore, the research hypothesis presented in Section 3.3 is confirmed. The conclusion of this research is that the variety and novelty of means generated in a morphological chart is not dependent on the function representation (function list or function structure). This work also serves as an exemplar on how systematic studies in ideation and concept generation/integration tools might be conducted. In order to develop informed guidelines to aid novice designers, the underlying relationships between the elements of ideation tools and effectiveness must first be studied.

6 REFERENCES


A Synthesis Decision Framework for Early-Stage Innovative Design

S. C.-Y. Lu, A. Liu
Viterbi School of Engineering
University of Southern California
Los Angeles, California, 90089 USA
sclu@usc.edu; angliu@usc.edu

Abstract
Design synthesis is an abductive reasoning process from abstract intents to concrete instantiations under constraints. This process involves two stages: purposeful “alternative creation” and systematic “alternative selection”. This paper presents a synthesis framework to support decision making during these two phases for early-stage innovative design. For alternative creation, a decision process that uses three synthesis reasoning operations to generate innovative options is provided. For alternative selection, an aggregation method that combines multiple preferences to choose the most preferred option is developed. This decision framework can be seen as a theoretical generalization of the Axiomatic Design Theory to better support design synthesis.

Keywords: Synthesis, Axiomatic Design, Innovative Design

1 INTRODUCTION
Generally speaking, synthesis is a creative human activity that synthetically combines several things to form a thing that did not exist before. In formal logic, synthesis calls for a “cognitive leap” from an intangible subject to a more tangible predicate by making abductive propositions. Accordingly, design synthesis can be seen as an abductive reasoning process from an abstract (or conceptual) intent (e.g., what) to some concrete (or detailed) instantiations (e.g., how). While synthesis is inherently subjective, especially during the early design stages, it should be structured to better support design innovations. With a structured decision framework, designers can have a common basis to objectively communicate, compare, negotiate and select various propositions based on individual subjectivities to synthetically innovate good design options.

The design synthesis task consists of two stages, namely purposeful alternative creation and systematic alternative selection under constraints. This task is especially difficult during the early stages when both the design intent and constraints are less tangible and hard to quantify. For alternative creation, the abstract “thing-free” (or “solution-neutral”) thinking is preferable, because it opens more innovation possibilities to create artifacts that are new and previously unseen. Meanwhile, due to the high abstraction level, alternative selection at these same stages is more driven by subjective human preferences than objective domain criteria. Earlier studies have already shown the difficulty of making group decisions with diverse subjectivities in general [1-2]; finding a preference aggregation method for abstract “thing-free” thinking is even harder. In short, alternative creation in design synthesis calls for “abstract reasoning” that should be solution-free to enhance design innovation, whereas alternative selection needs “concrete information” that must be detailed enough for systematic comparison. This dilemma illustrates one of the many challenges of supporting design synthesis at early stages.

Despite its importance, little design research has been devoted to supporting synthesis reasoning for early-stage innovative design. Most design theories to date focus on “analysis-based” alternative selection, leaving alternative creation to ah-hoc human decisions. Axiomatic Design (AD) theory [3] is a notable exception in this regard. However, to date the AD theory is most often used to select, rather than create, alternatives in design practice [4]; hence still fails short to fully support synthesis task at early design stages.

The rest of this paper is organized as following. Section 2 explains some relevant theoretical background of this research. Section 3 presents a synthesis-based design framework to support alternative creation and selection. Section 4 includes the conclusions and future work.

2 SOME THEORETICAL BACKGROUNDS
2.1 Synthesis in Logic and Design
Because synthesis occurs ubiquitously in many creative human activities, it has come to mean different things to different people. According to one dictionary definition, in philosophy, synthesis means a purposeful reasoning from the general to the particular [5]. It typically begins with an intangible (abstract) general (e.g., thought, intent, goal, objective, etc.) and ends with a more tangible (concrete and detailed) particular (e.g., embodiment, solution, plan, artifact, etc.). In science and engineering, synthesis occurs in all kinds of designing and planning activities where detailed embodiments are thought to satisfy a general goal. For instance, in electronics, logic synthesis is a process by which an abstract form of desired circuit behavior is turned into a concrete design implementation using logic gates. In chemistry, chemical synthesis is a purposeful execution of chemical reactions to get a desirable product. In biology, biosynthesis is a catalyzed process in cells of living organisms by which substrates are converted to more complex (i.e., detailed and structured) products. In mechanical design, design synthesis of mechanisms is the
transformation from abstract specification of required behaviors into detail description of machine's structure [6]. Except for these few instances, most synthesis activity in engineering are neither fully understood in research nor well supported in practice. Today, synthesis is still often carried out intuitively based on the designer's subjective experience rather than systematic reasoning. Iterative analyses with optimization algorithms to indirectly "mimic" synthesis reasoning via expensive trial-and-error are still the common practice. But, synthesis is fundamentally different from analysis in design - it cannot be achieved or replaced by analysis. Firstly, design synthesis is to create new things that have not yet existed, while design analysis is to investigate some existing (or will exist) things. Whenever analysis is to be carried out, it implies that something, either physically or conceptually, has been (or will be) there to be analyzed. Next, design analysis is based on the notion of optimality, which is different from design synthesis which must rely on the concept of rationality. Only if the domain governing laws of the thing (i.e., how things work) are known, can designers analyze the system to improve its behaviors towards the optimal. Lastly, the reasoning direction of "ends-means" synthesis is directly opposite to that of the "means-ends" analysis.

In formal logic, synthesis is an intentional "abductive" reasoning from abstract subjective to concrete predicate. Compared with synthesis, analysis and evaluation are reasoning based on deduction and induction respectively. Analysis plays the role of examining the predicate created through synthesis; whereas evaluation functions to verify if the predicate can satisfy the subject successfully. The iterative loop (Figure 1) formed by synthesis, analysis, and evaluation matches perfectly with the logical inquiry cycle: abduction generates new hypotheses; deduction analyzes the hypotheses; and then induction justifies the hypothesis [7].

In design practice, synthesis can be viewed as repeatedly making purposeful and rational propositions (i.e., subject-predicate pairs) from ENDS (what) to MEANS (how) under constraints. During such a process, an abstract thought is refilled to become some concrete things by encoding distinguished properties. Note that, elements in the ENDS (what) realm are always relatively abstract, compared with elements in the MEANS (how) domain which are more concrete. For instance, "objective" as an intangible WHAT can be realized by "object" which is a tangible HOW. In other words, "object" as a MEANS is a specific realization of "objective" that is a concrete ENDS. In contrast, analysis and evaluation go through a reverse direction from MEANS (how) to ENDS (what) by iteratively examining and assessing specific performances. When applied in tandem, synthesis, analysis, and evaluation continuously drive designers' decisions to move forward in design practice.

Abductive reasoning process in synthesis involves both hypothesis construction and hypothesis selection [8]. The design synthesis task is no exception; it corresponds to a "purposeful alternative creation" stage and a "systematic alternative selection" stage, respectively. Relevant theoretical backgrounds for each stage is provided below.

### 2.2 Making Propositions in Alternative Creation

In design synthesis, "purposeful alternative creation" is achieved by making abductive propositions. Kant [9] used the terms "analytic" and "synthetic" to differentiate two different types of propositions. Analytic proposition is defined as a proposition whose predicate concept is "contained" in its subject concept. For instance, "all bachelors are married" and "all triangles have three sides" are both analytic propositions. Note that, in either case, the predicate concept is contained in the subject concept. The subject concept of "bachelor" consists of the meaning of "unmarried". In other words, the predicate concept of "unmarried" is part of the definition of "bachelor". Analytic proposition establishes the "part-of" (consist-of) relationships within a decision hierarchy. As will be explained next, it is the foundation of the "Specialization" operation in our synthesis framework.

In contrast, synthetic proposition is a proposition whose predicate concept is "not contained" in its subject concept. For example, "bachelors are unhappy", "power can be generated by internal combustion (IC) engine", and "travel long distance can be achieved by taking airplane" are all synthetic propositions. Take the "IC engine" case for instance, the predicate concept "IC engine" is not necessarily contained in the subject concept "power", but only one of the many "means" (e.g., solar generator, wind generator, etc.) to realize it. In other words, the relationship created by the synthetic proposition is totally different from the "part-of" relationship established by analytic proposition. This is a very important difference which must be distinguished and understood in design synthesis reasoning. As will be explained next, synthetic proposition is the foundation of the "Realization" operation in our synthesis framework.

### 2.3 Aggregating Preferences in Alternative Selection

Preference aggregation has been studied extensively by social choice research for decades [1-2, 10-11]. The goal is to find a universal social welfare function (SWF) [12] that can consistently convert multiple individual preferences...
over a set of candidate alternatives into a single collective preference for group decision making. One of the most well-known theorems in social choice research is called Arrow's Theorem [1-2]. It theoretically proves that, in general, there is no universal method (or SWF) that can convert the ranked preferences from individuals into a community-wide group ranking while also meeting a set of so-called "Arrow's rational conditions"; they are (1) unrestricted domain, (2) non-dictatorship, (3) Pareto efficiency (i.e., if every individual stakeholder prefers a certain alternative to another, then so must the resulting group societal preference order.), and (4) independence of irrelevant alternatives.

Recently, some efforts have been devoted to examining the applicability of this Arrow Theorem to engineering design [13-17]. From a theoretical point of view, to escape Arrow's Theorem, there is an unavoidable price to pay; that is at least one of his rational conditions must be weakened somehow. In light of the special characteristics of engineering problems, it may be feasible to relax some of the requirements from social choice research to make preference aggregation possible in engineering design practice [18]. However, for group decision making in design teams, each of these possibilities must be carefully evaluated to determine if the price is worth the gain in the context of design synthesis [19].

For example, nominating a dictator to synthesize design alternatives is obviously inappropriate, because during the early stages, it is important to incorporate, not to exclude, every designer's preference. Next, weakening the unrestricted domain condition seems to be a relatively reasonable option, as the "domains" in engineering design are mostly bounded by internal and external constraints, and "simple majority rule" as a SWF is mostly ordinal. However, its realization in a particular design team depends highly on the existence of strict ordering of alternatives and the preference pattern (e.g., single-peakedness) of every designer. Therefore, it is generally unsuited to all types of design problems, particularly those whose alternatives are still relatively intangible during the early stages. Then, the Pareto-extension rule can be regarded as a universal SWF during the early design stages as long as "quasi-transitivity" can be commonly accepted by the team. However, it may result in multiple "equally desirable" alternatives. Scoring methods can also lead to the possible result; however, they are not purely "ordinal", and the application is largely determined by the availability of cardinality of the particular design problem. Note that the measure of cardinality (e.g., based on estimation) is mostly arbitrary during the early stages when alternatives are relatively abstract.

Besides weakening some of Arrow's rational conditions, preference aggregation is also possible by measuring each designer's cardinal utility of preference directly during the early design stages. For this possibility, the recent development of "Multiple-Layer Surrogate Modeling" [20] is quite relevant and useful. These abstract qualitative surrogate models built from the detail quantitative domain knowledge can be used to support the estimation of cardinal utility during the early design stages [17]. Once the cardinal utility of preference becomes available, utilitarian rule can also be regarded as a qualified SWF in synthesis decisions.

Table 1 summarizes the three possible preference-aggregation methods during the early design stages: utilitarian rule, simple majority rule, and Pareto-extension rule. Each method can be chosen as the SWF to synthesize multiple preferences at early design stages, if, and only if, the required informational basis can be satisfied.

3 A SYNTHESIS DECISION FRAMEWORK

3.1 Different Reasoning Operations

Table 2 summarizes the notations used to define the three different reasoning operations in our proposed synthesis decision framework for early stage design.

<table>
<thead>
<tr>
<th>Utility Type</th>
<th>Utilitarian Rule</th>
<th>Simple Majority Rule</th>
<th>Pareto-extension Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Voters</td>
<td>No</td>
<td>Odd</td>
<td>No</td>
</tr>
<tr>
<td>Violation of Arrow's condition</td>
<td>No</td>
<td>Unrestricted Domain</td>
<td>Pareto Efficiency</td>
</tr>
<tr>
<td>Required Informational Basis</td>
<td>Interpersonal comparison of utility</td>
<td>-Strict ordering of alternative -Single peaked preference</td>
<td>No</td>
</tr>
<tr>
<td>Advantage</td>
<td>Result is more precise</td>
<td>Purely &quot;ordinal&quot;, easy to carry out</td>
<td>Universal</td>
</tr>
<tr>
<td>Disadvantage</td>
<td>Cardinal utility is difficult to measure</td>
<td>Less precise</td>
<td>Many equivalent alternative</td>
</tr>
</tbody>
</table>

Table 1: Different preference-aggregation methods.

Table 2: Notations of the synthesis framework.

First, design synthesis (\( \mathcal{A} \)) is represented as a function that acts on the abstract subject (\( \mathcal{A} \)) as input to produce the concrete predicate (\( \mathcal{C} \)) as output.

\[
\mathcal{A} \mapsto \mathcal{C}
\]

Where, \( \mathcal{A} \) = Design Synthesis Function, \( \mathcal{A} \) = the "Abstract" (e.g., the ends, what, intent, or the subject)

\( \mathcal{C} \) = the "Concrete" (e.g., the means, how, instantiation, or the predicate)

Next, the (\( \mathcal{A} \)) function is further defined as making a "synthetic proposition" via a Realization (\( \mathcal{R} \)) operation and
an "analytic proposition" via a Specialization ($\hat{S}$) operation under the constraints via Bounding ($\hat{I}$) operations. To clearly specify these three basic reasoning operations, subscripts $i,j$ are used to denote the synthetic and analytic propositions, respectively (see Figure 2 below). Accordingly, the X-axis and Y-axis in the framework shown in Figure 2 represent a horizontal "conceptual-concrete" spectrum and a vertical "abstract-detail" spectrum of design synthesis.

The Realization Operation ($\hat{R}$)

($\hat{R}$) is defined as a generative type of abductive reasoning that makes synthetic propositions to transform a conceptual subject, that is relatively intangible in the upstream, to a more concrete (i.e., touchable, executable, etc.) predicate, that is more tangible in the downstream. ($\hat{R}$) creates the "means-of" (or "realized-by") dependency between subject and predicate. In design synthesis, ($\hat{R}$) can only be performed horizontally across two adjunct domains from $i=1$ to $i=m$. The transformation performed by ($\hat{R}$) is represented by Equation (2). Note that ($\hat{R}$) can be seen as a generalization of the horizontal "mapping" operation prescribed in the AD theory.

$$R(P_{ij}) \Rightarrow P_{i+1,j} \quad (2)$$

The Specialization Operation ($\hat{S}$)

($\hat{S}$) is defined as a derivative type of abductive reasoning that makes analytic propositions to transform an abstract subject, that is comparatively intangible at upper layers, to a detailed predicate, that is more tangible at lower layers. Unlike ($\hat{R}$), ($\hat{S}$) establishes a "part-of" dependency between the subject and predicate. In design synthesis, ($\hat{S}$) can only take place vertically from $j=1$ to $j=n$. The transformation performed by ($\hat{S}$) is represented by Equation (3). Note that ($\hat{S}$) can be regarded as a generalization of the vertical "decomposition" operation prescribed in the AD theory.

$$S(P_{ij}) \Rightarrow P_{i,j+1} \quad (3)$$

Specialization (or decomposition) is a more common operation (than Realization) that has been studied extensively by many existing analytical approaches (i.e., AHP analytical hierarchy process [21]), which can provide detailed guidance of implementation.

The Bounding Operation ($\hat{I}$)

($\hat{I}$) is defined as using constraints to limit transformation from a conceptual and abstract subject to a concrete and detailed predicate. In design synthesis, ($\hat{I}$) takes place diagonally from the upper abstraction layer and the downstream domain $\sum_{k=2}^{m} P_{i,k,j+1}$ to limit the transformation from $P_{ij}$ to $P_{i+1,j+1}$, as illustrated by Equation (4).

$$\hat{I}(P_{i,j+1}) \Rightarrow \hat{B}(\sum_{k=2}^{m} P_{i,k,j+1}) \Rightarrow P_{i+1,j+1} \quad (4)$$

incorporates those already known constraints imposed by external parties or those resulted from previous propositions made to focus (or limit) the consideration of possible solution alternatives ($P_{i+1,j+1}$).

3.2 The Design Synthesis Reasoning ($\mathcal{D}$)

Having defined the three basic reasoning operations in Section 3.1, a theoretical framework that guides the execution of a design synthesis reasoning process from a conceptual (abstract) subject to a concrete (detailed) predicate can be presented in Equations 5 and 6:

For any proposition within the bounded region of $i \in \{1, m\}$ and $j \in \{1, n\}$, given a set of boundary conditions imposed at $i=1$, $m$ and $j=1$, $n$, design synthesis ($\mathcal{D}$) is a reasoning function that transforms the input to the output as follow:

$$\mathcal{D}(P_{ij}) = \sum_{k=2}^{m} \hat{I}(P_{i,j+1}) \Rightarrow \hat{S}(P_{i+1,j+1}) \Rightarrow \hat{R}(P_{i,j}) \Rightarrow P_{i+1,j+1} \quad (5)$$

where $P_{i+1,j+1}$ must satisfy:

$$P_{i+1,j+1} \subseteq \bigcup_{q=1}^{n} \sum_{k=2}^{m} \hat{S}(P_{i,j}) \cap \hat{R}(P_{i,j}) \cap \hat{B}(P_{i,j}) \Rightarrow P_{i+1,j+1} \quad (6)$$

Although the two equations defined above cannot be solved mathematically, they can be regarded as some useful mental guidelines to drive the designer's thinking direction in design synthesis.

3.3 A Design Synthesis Framework

So far, we have defined synthesis reasoning in formal logic as a cognitive "leap" from a relatively intangible subject (= $P_{ij}$) to a more tangible predicate (= $P_{i+1,j+1}$) by making abductive propositions, where $i$ and $j$ denote the synthetic and analytic propositions, respectively. Based on this logic foundation, the design synthesis task can be modeled as abduction from an abstract intent (i.e., "what") to a concrete instantiation (i.e., "how"). When carrying out a design synthesis (see Equation 5), the designer must go through two stages sequentially: the alternative creation stage and the alternative selection stage. The goal of alternative creation is to ideate some possible instantiations for further comparisons. The quality of obtained alternatives (for example, in terms of their level of innovativeness and diversity) is the main concern at this stage.

Figure 2: Three basic reasoning operations for design synthesis decision framework.
stage. The goal of alternative selection is to choose a unique instantiation as the final outcome (i.e., $P_{i+1,j+1}$) of design synthesis; efficiency, for example, in terms of the required effort of choosing the most appropriate instantiation, is the main focus at this stage. Certainly, the alternative creation quality and the alternative selection efficiency are mutually related to each other in design practice.

In the alternative creation stage, given $P_i$, the designer must first form a “nucleus” mentally in order to focus his/her creative attentions to satisfy $P_i$. In other words, starting from “all things are possible” initially (i.e., the solution-free thinking desired by innovative design should begin with all possible alternatives without any limitation from past solutions or creative hindrances), a bounded small “space for consideration” (to which $P_{i+1,j+1}$ must belong) must be carefully established first. Unlike deduction or induction for which some fixed procedures or algorithms can be used, this ideation (i.e., nucleation of ideas) process is mainly of the abduction type, and hence can only be systematically guided (or mentally driven) by abduction-based synthesis operations.

The three synthesis operations defined in Section 3.1 must be applied in tandem here. The first is the “Realization operation” ($R$) that uses synthetic proposition to create the “means-of” relationship between $P_i$ and $P_{i+1}$. In other words, the designer must think horizontally first along the “$i,j$” operation (“$i,j$” means-of” relationship between $P_i$ and $P_{i+1}$). In other words, the designer must then think vertically within the same decision domain ($i$) and ask “what are the possible $P_{i+1}$ that could be a part of $P_i$?” The second is the “Formation operation” ($F$) that assures that the resulting $P_{i+1,j+1}$ are limited by both domain-independent axioms and domain-dependent constraints. In other words, the designer must also think diagonally across the domain and one layer, and ask “what are the possible $P_{i+1,j+1}$ that would be within the boundary of limits imposed by these axioms and constraints?” In short, during the alternative creation stage, the resulting limited “space for consideration” ($P_{i+1,j+1}$) is formed by jointly and simultaneously considering the intersection between $P_{i,j}$ and $P_{i+1}$ that also meet some domain-independent and dependent constraints.

The domain-independent constraints that must be included in the Bounding operation ($F$) can include three axiom-based criteria from the AD: i.e., complete, minimal and independence. That is to say, that those ideal alternatives within the small limited space for consideration ($P_{i+1,j+1}$) must: (1) completely satisfy the design intent (i.e., the subject of synthesis) expressed by $P_i$, (2) without any redundancy (or duplication) among themselves, and (3) be functionally independent from each other. The domain-dependent constraints that must be included in the bounding operation can be one of two kinds. When $i=1$ and $j=1$, for $P_{i,j}$, and $P_{i+1,j}$, the constraints along the boundary are those restrictions imposed onto the designer by corporations, policies, regulations, and markets as well as other known resources (such as time, budget, etc.) limits. For other instances (i.e., $i>1, j>1$), the constraints in the interior are those propositions that have been made previously and agreed upon by the designers before at the upper abstraction layer and the downstream domain (i.e., $\sum_{k=2}^i P_{i+1,k} \cdot j$).

The combined considerations among the above “means-of,” “part-of,” and “constraint-by” (using both domain-independent axioms and domain-dependent constraints) operations lead to a small limited space for consideration that consists of “a few high quality alternatives” at the conclusion of the Alternative Creation stage of design synthesis. These few qualified alternatives will then become the candidates of comparison and choice among designers during the alternative selection stage next.

For alternative selection, three selection methods derived from social choice research (see Table 1) can be used, while the challenge now lies in which method to choose under the specific circumstances of the design problem at hand. Different informational bases determine the applicability of different methods, and different methods might lead to completely different aggregation outcomes [22]. It can be hypothesized that, during the early design stages, the required informational basis for each method is related to the alternative’s level of abstraction. In other words, the alternative’s level of abstraction can influence the choice of the most suitable selection method, which will in turn affect the final aggregation result.

In design synthesis, the candidate alternatives to be rank-ordered are a set of concrete instantiations, whose level of abstraction has been relatively lower compared with that of the initial abstract intent. Therefore, the designers should first check if cardinal utility of preference can be measured (e.g. by qualitative surrogate modeling) for the specific design problem at hand. If that is the case, the Utilitarian rule can be chosen as the SWF, because it yields the most precise aggregation result. Sometimes, due to the lack of domain-dependent models, it is very difficult to directly measure cardinal utility. Under such circumstances, it is more appropriate for the designers to utilize some purely “ordinal” procedures. Hence, they should check if the “single-peakedness” condition can be satisfied, if yes, then choose simple majority rule as the SWF. Otherwise the Pareto-extension rule should be used to combine team preferences, which could bring out some equally desirable alternatives that cannot be compared by subjectivity at the present abstraction level to the next design synthesis, and transform all of them to more tangible instantiations for further comparison.

3.4 The Design Synthesis Process

The Alternative Creation Stage

Figure 3 below illustrates a typical alternative creation process in design synthesis from subject $P_i$ to predicate $P_{i+1,j+1}$. The reasoning process at this stage consists of two consecutive steps, namely Formation, and Ideation.

![Figure 3: The design synthesis process.](image-url)

The Formation step in alternative creation stage involves three sub-steps: elicit design intent, identify boundary conditions, and establish a small limited “space for consideration”.

- Elicit the design intent:
  
  Purposeful alternative creation in design synthesis begins with eliciting the design intent. In the following discussion,
it is assumed that the designer has already arrived at a certain intent (i.e., \( P_{i,j} \)) from previous stages.

- Identify the boundary conditions:
  
  Boundary conditions during the synthesis reasoning process are identified as those domain-dependent constraints. In Figure 3, they are represented by the horizontal axis (from \( i=1 \) to \( i=m \)) and the vertical axis (from \( j=1 \) to \( j=n \)).

- Establish a small limited “space for consideration”:
  
  1) Create the means-of dependency

  Two (\( R \)) operations are first carried out to create means-of dependency along the horizontal direction across three separate hierarchies between the subject (i.e., \( P_i \)) and the predicate (i.e., \( P_{i+1} \)) and (i.e., \( P_{i+2} \)) via two synthetic propositions. As shown in Figure 3, the links (a) and (b) are established by applying Equation (2) previously defined in Section 3.1 twice.

  \[
  \hat{R}(P_{i,j}) \Leftarrow P_{i+1,j} \quad (i.e., \text{link a}) \tag{7}
  \]

  \[
  \hat{R}(P_{i+1,j}) \Leftarrow P_{i+2,j} \quad (i.e., \text{link b}) \tag{8}
  \]

  Note that, if the level of abstraction of \( P_i \) is relatively high, one (or more) additional realization operations may be further carried out to create more means-of dependency (i.e., \( \hat{k}(c) \)) along the horizontal direction (i.e., \( P_{i+3} \))

  2) Create the part-of dependency

  One (\( S \)) operation is now performed to create the part-of dependency along the vertical direction within a single hierarchy between the subject (i.e., \( P_i \)) and the predicate (i.e., \( P_{i,j+1} \)). As shown in Figure 3, the link (d) is established by applying Equation (3) once.

  \[
  \hat{S}(P_{i,j}) \Leftarrow P_{i,j+1} \quad (i.e., \text{link d}) \tag{9}
  \]

  3) Identify the domain-independent constraints

  As explained before, the domain-independent constraints that must be included via the (\( \hat{1} \)) operation include the three criteria of the first Axiom from the AD theory: i.e., complete, minimal and independence.

  4) Identify the domain-dependent constraints

  Domain-dependent constraints include previous propositions made at the upper abstraction layer and the downstream domain of \( P_{i+1,j+1} \), which can be represented as

  \[
  \sum_{k=2}^{m} P_{i+k,j} \quad .
  \]

  During the ideation step of alternative creation stage, within the small limited “space for consideration” formed above, the designer must now “ideate” a few concrete (or more detailed) instantiations for further selection. The ideation of possible instantiations can be seen as a deductive reasoning influenced by “forces” from several directions within the bounded space, as indicated by links (e), (f), (g), (h), and (i) in Figure 3. They include a horizontal force acting from the upstream \( P_{i+1} \) via a (\( R \)) operation (i.e., link (e)), a vertical force acting from the upper-layer \( P_{i+1} \) via a (\( S \)) operation (i.e., link (f)), a few diagonal forces acting from the upper-layer and downstream domain-dependent constraints.

  \[
  \sum_{k=2}^{m} P_{i+k,j} \quad \text{via a bounding operation (i.e., link (g) and link (h)), and one last force acting from the domain-independent axioms via a (\( \hat{1} \)) operation (i.e., link (i)). In other words, the ideated instantiations must at the same time be “means-of” \( P_{i+1} \), “part-of” \( P_{i+1} \), and constrained by
  \]

  \[
  \sum_{k=2}^{m} P_{i+k,j} \quad .
  \]

  These combined operations lead to Equation (6) in Section 3.2.

Equation (6) will yield a small option space from which the designer can ideate a few concrete instantiations, which completes the alternative creation process of synthesis during the early design stages. With these ideated instantiations as candidate alternatives, the designer can now move on to the alternative selection stage to choose a unique \( P_{r_{i+1}} \) based on multiple preferences.

The Alternative Selection Stage

As part of the proposed synthesis design framework, a specific preference-aggregation model is developed to support alternative selection. The model guides designers in a design team to collaboratively go through three sequential steps, namely preference formation, preference evaluation, and preference aggregation, to orderly and rationally combine multiple individual preferences into a single team preference for selection.

The preference formation step consists of two sub-steps: eliciting candidate alternatives and discourse individual preferences.

- Elicit candidate alternatives:

  The preference formation step begins with eliciting candidate alternatives that are yet to be rank-ordered. In case there are new stakeholders joining the design synthesis process from the selection stage, a review sub-step is performed if needed. The objective is to ensure that every stakeholder fully understands the intent to be satisfied and all previous operations that were carried out to ideate these instantiations. Some techniques for reviewing the sub-step can be found in Social Technical Co-Construction Process for collaborative engineering [23]. Once a common understanding is established among all stakeholders, these instantiations formally become candidate alternatives to be compared and selected.

- Discourse individual preferences:

  Having elicited a set of candidate alternatives from all stakeholders, designers can now discourse their individual preferences as multiple orderings over alternatives from the most to the least desirable.

The preference evaluation step involves two sub-steps: access the utility type of preference and measure cardinal utility of preference.

- Assess the utility type of preference:

  The preference evaluation step begins with assessing the reasonable utility type (i.e., ordinal utility or cardinal utility) of already expressed preferences. If only ordinal utility is available or possible, then proceed to choose the SWF to carry out preference aggregation (see item 1 in Aggregation step).

- Measure cardinal utility of preference:

  If possible, use relevant domain-dependent models (e.g., multi-layer surrogate modeling) to assess cardinal utility.

The preference aggregation step consists of two sub-steps: choose SWF for preference aggregation and combine multiple preferences.

- Choose SWF for preference aggregation:

  Based on available informational bases, various selection methods can be chosen as the SWF. It takes three sub-steps to identify the most suitable SWF for a particular design problem.

(1): If cardinal utility can be quantified, Utilitarian rule should be chosen as the SWF. If not, then go to sub-step 2 below.

(2): If a strict ordering exists among the set of alternatives and every individual preference is single-peaked, plus the number of stakeholders is odd, then simple majority rule can be chosen as the SWF.
(3): Choose the Pareto-extension rule as the SWF; the Pareto-extension rule is automatically promoted to be SWF.

- Combine multiple individual preferences into a single team preference:

Following the above guidance of choosing the SWF, a team preference of ordering candidate alternatives can be systematically produced. After completing the preference-aggregation process, if there remain a few “equivalently” desirable alternatives, all of them should be brought back to the alternative creation stage to transform them into more tangible instantiations to be further compared systematically.

3.5 An Example of Using the Synthesis Framework in Early-stage Innovative Design

This section provides an example of conceptual design of a computer keyboard using the proposed synthesis decision framework described in this paper.

One important customer need (CN) of keyboard design is to “alleviate Repetitive Stress Injury” (RSI). We assume that this is the abstract intent (or subject) to begin the design synthesis task. The boundary conditions imposed by external parties in this case are: competing products in the market (e.g., existing ergonomic keyboards), limited budget, available materials, etc. To form a limited “space for consideration”, two synthetic propositions, i.e., (1) “alleviation of RSI” can be realized by adjusting user’s posture gradually, and (2) “posture adjustment can be realized by adaptive motion control device”, are first made to create means-of dependencies horizontally. This is followed by making an analytic proposition, i.e., “alleviation of RSI includes elimination of static repetition” to establish “part-of” dependencies vertically.

During the ideation step, three instantiations are created within the “limited space for consideration”. They all represent possible qualified solutions that can “completely, minimally, and independently” satisfy the initial design intent to alleviate RSI.

- Instantiation 1: to adaptively adjust the key zones by software programming
- Instantiation 2: to continuously adjust the gable of keyboard by mechanical system
- Instantiation 3: to manually adjust the key zones by the user

These synthetically generated instantiations conclude the alternative creation stage of design synthesis. The entire creation process is illustrated in Figure 4 below.

Next, in the alternative selection stage, multiple subjective preferences are utilized to rank-order the above three instantiations. Instantiation 3 is finally chosen as the most “preferred” alternative following the specific preference aggregation procedure prescribed in Section 3.4. This is the final outcome of this particular synthesis reasoning.

4 SUMMARY AND FUTURE WORK

Synthesis aims, given an abstract P_0, to arrive at a concrete P_i,j+1 that “is-a” tangible “thing”. In design, the result of synthesis can be obtained by querying certain databases (e.g. of relevant knowledge, theory, experience, etc.) [24]. Even without the guidance of any design theory or methodology, designers can still use some heuristics to query the database to reach a final result (i.e., “is-a”). The goal of design theory and methodology research is to provide designers with some specific query patterns (e.g. zigzagging in the AD theory) to carry out the query process more systemically and effectively. In a way, this research develops a new query pattern to effectively support design synthesis (see Figure 5).

Figure 5: Synthesis framework as a query pattern.

Our new query pattern combines the “means-of”, “part-of”, and “constrained-by” reasoning operations in a specific manner to arrive at the final query result (i.e., “is-a”), as illustrated in Equation (10) below. It is expected that the query process can be enhanced by applying the new query pattern guided by our synthesis reasoning framework. That is to say that, if more high-quality alternatives can be generated, the quality of query becomes higher.

\[
(is-a) = (\text{means-of}) \cap (\text{part-of}) \cap (\text{constrained-by}) \tag{10}
\]

After a set of alternatives are created following certain query pattern, the designer must then make rational decisions to choose a unique one as the outcome of design synthesis. Without the assistance of any preference aggregation method, the designer can still arrive at a choice by using their heuristics. However, selection methods suggested by our synthesis reasoning framework can help designers to compare and rank-order alternatives more systemically and efficiently. Some selection methods rely on objective criteria as merit for selection, whereas others utilize subjective preference. For early-stage innovative design which lies much closer to the subjective extreme of the design decision spectrum, using subjective preference for alternative selection is more appropriate. In this research, motivated by the nature of design synthesis (i.e., from intangible intent to tangible instantiation), the choice of the most suitable selection method is associated with the alternative’s level of abstraction, and a preference-
aggregation model is prescribed to help choose the most appropriate selection method under diverse situations.

In conclusion, the proposed synthesis design framework supports early-stage in innovative design in such a way that: for alternative creation, the structured synthesis-based query pattern provides a common basis to compare multiple propositions created by different designers; for alternative selection, the preference-aggregation method picks up the right selection method which in turn leads to a unique solution.

This paper presents the initial development of a synthesis design framework to support early-stage innovative design. The framework is expected to help enhance both the quality of alternative creation and the efficiency of alternative selection during the process of design synthesis during the early stages. In our future work, a series of design experiments with direct participation of multiple designers will be conducted, analyzed and compared to validate the foundation and improve the performance of this new design synthesis framework.

5 ACKNOWLEDGEMENTS

Our research in developing a design synthesis framework can be seen as an effort to use theories from formal logics to establish a theoretical foundation for the 2-D mapping and decomposition operations as well as the unique zigzagging procedure originally suggested by the Axiomatic Design Theory developed by Prof. Nam P. Suh. We received much inspiration and useful guidance from Prof. Suh during the course of this research.

6 REFERENCES

An Approach to Improving the Cooperation between the Departments of Product Design and Technical Production Planning

L. Weyand, H. Bley
Institute of Production Engineering / CAM, Saarland University, Saarbruecken, Germany
weyand@cam.uni-saarland.de, bley@cam.uni-saarland.de

Abstract
In industry, the cooperation of the departments of product design and production planning is an important key factor for success. As a consequence, an approach which allows performing a product variant-oriented production planning and clearly feeding back the information regarding the resulting process/resource variance to the product design department has been developed. In this paper, the focus is mainly on the second part of the approach. It is explained how the results can be fed back to the product design department in a fast and non-ambiguous way.

Keywords:
Cooperation, Product Design, Production Planning

1 INTRODUCTION
Product and production complexity is increasing in car manufacturing [1-3]. One reason for this is an increase in customer orientated approaches especially by European car manufacturers who are trying to concentrate more and more on individual wishes. Customers can now choose between many different car bodies, engines and equipment packages. As to the product development process, the cars have to be designed in a way that they can be individually configured but also easily produced in one production line. But up to now, problems occur if a high number of products and product variants has to be considered within the development process.

During the early phases of classic production planning, transparency regarding separate product variants and regarding the existing process/resource variance in the vehicle project is often missing [4]. Moreover, the cooperation between the product design department and the production planning department is partly suboptimal. This situation is surprising due to the fact that it is well-known that the cooperation of these two departments is an important key factor for success [5-7]. The old “throw it over the wall mentality” (cp. with Figure 1) is not the same problem as it was years ago, but there are still stones of the old wall left and those stones still cause trouble [8]. Misunderstandings between the departments can occur due to a lack of communication.

Consequently, there is a need for action in this area: the transparency within production planning has to be improved and the results have to be fed back to the product design department in a fast and non-ambiguous way. Product designers should be clearly informed if specific parts of product variants lead to a process/resource variance and whether they are rated as critical by the production planning department.

In this paper, an approach is presented that helps to improve the transparency of production planning and the cooperation between the departments of product design and production planning. Based on the presented approach, it becomes immediately visible for product designers which specific parts of product variants lead to a process/resource variance and whether they are rated as critical by the production planning department.

2 APPROACH TO A TRANSPARENT, VARIANT-ORIENTED PRODUCTION PLANNING
Especially during the early phases of traditional production planning, a part-based planning is often preferred. Assembly planners consider the elements of the product structure they are responsible for and they define the processes and resources which are needed for joining the parts to their destination point in the vehicle. The planning results are stored in a digital planning environment and so-called product-process-resource models (PPR models) [10] have to be created. Figure 2 shows a simple example of a process definition including resource information.
<table>
<thead>
<tr>
<th>Door handle (outside)</th>
<th>Assembly processes</th>
<th>Required resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. classic / normal</td>
<td>1. pick up</td>
<td>- one pair of pliers</td>
</tr>
<tr>
<td></td>
<td>2. fix (7 s)</td>
<td>- one screwdriver</td>
</tr>
<tr>
<td></td>
<td>3. screw (2x)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. check</td>
<td></td>
</tr>
<tr>
<td>(component of the product variants x)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. elegance / radio</td>
<td>1. pick up</td>
<td>- one pair of pliers</td>
</tr>
<tr>
<td></td>
<td>2. connect</td>
<td>- one screwdriver</td>
</tr>
<tr>
<td></td>
<td>3. fix (10 s)</td>
<td>- one power supply and</td>
</tr>
<tr>
<td></td>
<td>4. screw (2x)</td>
<td>transmitter unit</td>
</tr>
<tr>
<td></td>
<td>5. check by transmitter</td>
<td></td>
</tr>
<tr>
<td>(component of the product variants y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. sport / radio</td>
<td>1. pick up</td>
<td>- one pair of pliers</td>
</tr>
<tr>
<td></td>
<td>2. connect</td>
<td>- one screwdriver</td>
</tr>
<tr>
<td></td>
<td>3. fix (10 s)</td>
<td>- one power supply and</td>
</tr>
<tr>
<td></td>
<td>4. screw (3x)</td>
<td>transmitter unit</td>
</tr>
<tr>
<td></td>
<td>5. check by transmitter</td>
<td></td>
</tr>
<tr>
<td>(component of the product variants z)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product</th>
<th>Process</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Door handle classic / normal</td>
<td>A1: pick up</td>
<td>one pair of pliers</td>
</tr>
<tr>
<td></td>
<td>A2: connect</td>
<td>one screwdriver</td>
</tr>
<tr>
<td></td>
<td>A3: fix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A4: fix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A5: screw</td>
<td>one power supply and</td>
</tr>
<tr>
<td></td>
<td>A6: screw</td>
<td>transmitter unit</td>
</tr>
<tr>
<td></td>
<td>A7: check</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A8: check by transmitter</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Example of a process/resource definition regarding different door handles.

It is obvious that different assembly processes and resources are needed for assembling the different door handles to a side door (cp. with Figure 2). The relations between parts and processes/resources are clear, whereas it is difficult to see all of the differences regarding assembly aspects at a glance. There is a low transparency in respect of production differences.

While comparing pairs is easy, it can be a time-consuming and error-prone task in the case of a high number of alternative parts differing in lots of details. However, the transparency regarding the assembly differences is still lower in a classic PPR model due to the fact that processes and resources are not stored in a redundant way. Figure 3 shows a PPR model which contains the same information as Figure 2 but without redundant information.

<table>
<thead>
<tr>
<th>Product</th>
<th>Process</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Door handle classic / normal</td>
<td>A1: pick up</td>
<td>one pair of pliers</td>
</tr>
<tr>
<td></td>
<td>A2: connect</td>
<td>one screwdriver</td>
</tr>
<tr>
<td></td>
<td>A3: fix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A4: fix</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A5: screw</td>
<td>one power supply and</td>
</tr>
<tr>
<td></td>
<td>A6: screw</td>
<td>transmitter unit</td>
</tr>
<tr>
<td></td>
<td>A7: check</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A8: check by transmitter</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Example of a classic PPR model.

In order to enhance the transparency, it is proposed to create process/resource graphs which immediately visualise the production differences of product variants. In this context, it is rated as necessary to use additional elements in the graphs, not only boxes which represent processes or resources. It makes sense to use so-called variant elements that explain the differences between processes in a clear way. In many cases, it is sufficient to use the following four variant elements:

- The variant element “Technology variance (T)” which has to be used in the case of processes that are different regarding their joining technology (for instance in case of gluing and screwing processes).
- The variant element “Time variance (t)” which has to be used in the case of processes that are equal regarding their joining technology but that have different execution times (e.g., a different time is needed for joining three or for joining two screws).
- The variant element “Plus (+)” which has to be used in case not all product variants need the process and if the process is not relevant to the reference variant.
- The variant element “Minus (-)” which has to be used in case not all product variants need the process, but if the reference variant needs this specific process.

The variant-oriented planning workflow based on graphs can be described as follows: firstly, the variant with the highest expected sales figures should be declared as the reference variant and it should be chosen as a starting point. A process sequence has to be created for this variant – process variance is not possible at this planning stage. Afterwards, the product variant with the second highest expected sales figures should be considered. In this case, it is possible that further or different processes are necessary for producing the second product variant. The original process sequence has to be modified where necessary – in case of a modification, a process graph is created which contains a process variance. In a third step, the product variant with the next highest expected sales figures is dealt with, and so on. Consequently, instead of starting a new planning procedure for every new product variant, the already created process graph has to be checked in respect of its suitability for the next product variant. Figure 4 shows the resulting process graph regarding the example presented in Figure 2.
The process graph shown in Figure 4 offers high transparency regarding the assembly differences of the product variants which contain different door handles. With the exception of process step 1, process variance is given in all process steps. The different paths that the product variants take in the graph are clearly presented and can also be easily identified that the differences between product variants y, which contain the door handle classic/radio, and product variants z, which contain the door handle sport/radio, are low in this section of the graph. There is only one process step, process step 4, in which a different execution time is needed due to a different number of joining elements (door handle classic / radio: two screws, door handle sport / radio: three screws). Execution times are usually estimated and stored in the digital planning environment by experienced production planners who partly use MTM analyses in order to reach reliable results. The different shape and weight of the two door handles that are radio controlled do not affect the assembly process. As to the product variants x which contain the door handle classic / normal, it can be easily identified that these product variants often take a different path than the other variants. The process A2 is not relevant to the product variants x and in the process steps 3 and 5, special processes are needed for the variants x. In process step 5, technology variance occurs. In order to perform a function check of the radio controlled door handles, a power supply and transmitter unit is necessary, whereas the door handle to be opened by key does not need any special equipment. This function check can be performed manually.

All the points described above are visible in the graph and consequently will be understandable for production planners or decision makers if they receive training in how to create and read the graphs. As already mentioned in [4, 11-13], digital graphs consisting of special variant elements offer advantages. Moreover, it has to be considered that line balancing information can be stored in process graphs by using different types of edges. Furthermore, process graphs can be used to identify the standardisation degree on the process level by counting and assessing the number of variant elements existing in the graph as well as by analysing the number of processes enclosed by variant nodes. Additionally, it has to be mentioned that resource graphs can be derived from process graphs, whereas in this publication, it is not considered to be necessary to work with resource graphs. However, the main topic of this paper is to introduce an approach which determines how the relevant results of production planning can be clearly fed back to the product design department. Section 3 deals with this topic.

### 3 APPROACH TO FEEDING BACK THE RESULTS OF PRODUCTION PLANNING

In order to improve the cooperation of product design and production planning, some endeavours have been performed in several companies in the recent past. For instance, Concurrent Engineering [14-16] Design for X (DFX) [17-19] and computer-aided coordination and decision support tools [20-21] which represent a part of the Digital Factory [22-23] are introduced in practice. In general, it has been recognised that it makes sense to bring the responsible persons of the cooperating departments together. Cross-functional meetings can help to improve the development process and the outcome [24-26]. However, even in meetings, misunderstandings are possible and it cannot be expected that there is nothing but harmony in the meeting group – inter-disciplinary problems can occur [27]. Moreover, it has to be considered that there are time lags between meetings. Therefore, further possibilities for exchanging information in a clear and objective way are necessary. An automatic feedback of the results to the product department is considered to be necessary for the future, especially if the product range and, in this context, the production complexity further increases. Of course, only results which are interesting for designers have to be fed back. In the example dealing with different door handles (cp. with Section 2), it was identified that the shape and weight of the door handles do not affect the assembly process. This information is not interesting for the design engineers; instead, they need feedback that says that the working principle of the normal door handle with a key in comparison to the radio-controlled versions of the door handles causes production differences. The two radio-controlled door handles can be more or less equally produced – there is only one difference in process step 4 caused by the different number of screws to be joined. This information has to be fed back to the working environment of the product designers too.
In connection with previous research activities dealing with final assembly planning, seven different part properties which can cause production variance have been identified:

- Differences regarding the working principle of the part variants, also called functioning (F).
- The dimension (D) of the part variants is different.
- Differences regarding the materials (M) of the part variants.
- The weight (W) of the part variants is different.
- Differences regarding the number of joining elements/connectors (N) of the part variants.
- The colour (C) of the part variants is different.
- Differences regarding the shape (S) of the part variants.

As mentioned above, two aspects – the functioning (F) and number of joining elements (N) – have to be considered regarding the door handle example discussed in Section 2. This information has to be sent back to the product designers in an intelligible way. In order to reach this goal, it is proposed to create a so-called product graph which again contains special variant elements contrasting with the normal product structure. The variant elements have to visualise the differences between the parts that lead to production variance. Consequently, the variant elements have to take the differences (F), (D), (M), (W), (N), (C) and (S) listed above into account, whereas the use of the variant elements is case-specific. Figure 5 shows the resulting product graph that belongs to the door handle example discussed in Figure 4.

The product graph visualises the differences between the door handles which are relevant to production. The representation of the product variance causing production variance is non-ambiguous, even if a special case has to be considered in Figure 5. Based on the variant element (F), it is obvious that one important difference from the point of view of production is the different functioning of the door handles. Within the created product graph, it is obvious that the door handle with key has to be dealt with in a different way than the radio-controlled door handles as far as their production is concerned. Additionally, it is visible in the graph that the number of joining elements (N) varies regarding the two radio-controlled door handles. Moreover, further information can be generated in an indirect way. Due to the fact that there is no further variant element (N) placed on the level of the variant element (F), the product designer can conclude that the number of joining elements of the door handle classic, which needs a key, is equal to one of the radio-controlled versions. This is another important aspect. The pair of door handles which have the same number of screws can be easily identified in a second step by analysing their attributes in the digital environment. Consequently, it is shown that the representation of the differences which are relevant to production is non-ambiguous in the product graph. A complete scenario of a side door assembly (cp. with Figure 6) has been successfully discussed.

However, the product designers still have no feedback concerning whether the identified differences are rated as critical or not by the production planners at this point. In order to make a statement about critical conditions, process and resource analyses have to be performed. In this paper, especially process analyses are discussed in more detail.

![Product Graph](image1)

![Product Graph](image2)
Based on the processes included in the process graph, time analyses can be performed in order to make a statement about whether line balancing problems have to be expected or not. It is proposed to perform the following time analysis:

- **Step one**, the product variants to be compared must be selected (the selection depends on the relevant elements of the product graph to be analysed).
- **Step two**, the maximum time deviation of the relevant processes has to be determined for each process step.
- **Step three**, all of the maximum time deviations of the process steps considered have to be summed up.
- **Step four**, the result of step three has to be divided by the intended cycle time of the production line.
- **Step five**, the result of step four has to be evaluated.

Based on the examples discussed, it can be said that the result of step four is not critical if it is lower than 10%. Even a result of up to 20% is usually still acceptable, whereas results higher than 20% could cause larger problems within the subsequent phase of line balancing. In this paper, it is proposed to use three different colours to illustrate the three different intervals mentioned – green for uncritical situations, yellow for still acceptable situations and red for possibly critical conditions. In general, there is no doubt that the presented time analysis is only a first step. If the results of the time investigation are good (i.e. green or yellow), problems during the step of line balancing do not have to be expected – at least if the assembly time of the most time-consuming product variant is still lower than the intended cycle time in the analysed section. If the results of the investigation are unacceptable (i.e. red), the situation could but does not have to lead to problems within later steps of production planning. However, it makes sense to identify theoretically critical conditions as early as possible so that countermeasures can be defined on the production level or product level. Figure 7 illustrates the discussed time assessment method based on the example of a door handle assembly.

Additionally, a resource investigation has to be performed. It has to be identified whether product variance leads to resource variance and consequently to additional investment costs. The result of the investigation has to be classified again. The following three different intervals are proposed: green for uncritical situations due to low investments in additional resources, yellow for still acceptable investments in additional resources and red for unacceptably high investments. The limits have to be defined in advance. For the side door example, it was decided to define intervals from 0€ to 1000€ (green section) and from 1000€ to 2000€ (yellow section). Values greater than or equal to 2000€ (red section) are rated as possibly critical instead. The performed resource investigation is also shown in Figure 7. In this case, it does not have any influence on the final assessment. In order to achieve a final rating, the results of the process and resource analysis have to be superimposed. In case of the door handle example, the differences on the process level are more critical than those on the resource level. This fact is also considered in the final feedback sent to the product designers (cp. with Figure 7). It always dominates the most critical result coming either from the process level or from the resource level. In connection with the discussed side door example, it has also been identified that it is sometimes reasonable to reorganise the initial product structure. Figure 8 shows one example which led to a restructuring.
Figure 8: Example leading to a restructuring.
As shown in Figure 8, the initial product structure and the final product structure created after feeding back the results of production planning are different. From the point of view of a production-oriented product development, the final product structure offers advantages. This structure classifies the part variants based on their properties which lead to production variance and it describes the impact of the differences. In case of the different door panels, the different number of connectors (here: cables to be connected) which is needed if a car with sound system 1 or 2 has to be produced or if a car without any sound system is ordered by a customer has an impact on the assembly operation of the door panels. This information is available in the product graph and it is also obvious that the production differences are rated as unacceptable from the production planning department. Thus, a meeting with product designers and product planners to discuss and optimise the door panel design with the objective of deleting unacceptable assembly differences caused by design differences is not necessary.

As to the door handle example shown in Figure 7, the situation is different. The product graph illustrates in a clear and objective way that there is a need for action regarding the door handles. In this case, the product designers have to study the results of production planning in detail and a meeting of responsible product designers and production planners could make sense. It can be summed up that the presented product graphs offer at least two advantages: transparency regarding part properties which lead to production variance and objectiveness regarding the impact of relevant differences. Both aspects are rated as absolutely necessary for a production development process of the future which considers individual customer wishes on a larger scale.

4 SUMMARY AND OUTLOOK

In this paper, a special method to improve the cooperation of the product design and production planning department is presented. It is proposed to enhance production planning by using special process/resource graphs which allow identifying the existing production variance at a glance and which deliver the basis for feeding back the results to the product designers in a fast and non-ambiguous way. The design department needs to know which product variants and part properties are causing production variance and whether the existing differences are critical from the point of view of production planning. A so-called product graph which consists of special variant elements and contains all relevant information in a clear and objective way is proposed. In the case of a critical feedback, cross-functional meetings with the aim of deleting suboptimal conditions have to be arranged with participants from both departments.

This special approach has been checked based on the scenario of a side door assembly. Referring back to the example, it was possible to create a non-ambiguous product graph. However, there is no doubt that some further examples have to be discussed in the future and that it could become necessary to define some further variant elements and rules in order to be able to feed back the results of production planning in a clear and non-ambiguous way.

Moreover, it has been discussed that it would be interesting to look at further production areas instead of focusing on final assembly planning only. For instance, the requirements in the area of manufacture of parts starting with unprocessed blanks are different and will lead to different product graphs. This topic could be interesting for automotive suppliers which partly have to deal with challenges that are different from those of the OEMs. The question of whether it finally makes sense to superpose product graphs created in a different context (e.g. superposing a product graph created in the context of final assembly planning) cannot be answered at this point. Further research activity will be necessary in this area.

Lastly, adequate software support has to be taken into account. Tests have to be conducted to clarify how the digital working environments of product designers and the digital working environments of production planners can be coupled in a bidirectional way and how the relevant information can be easily transferred. Overall, it should be possible to create product graphs completely automatically by software.

5 ACKNOWLEDGMENTS

This research activity was partly supported by the research project “Flexible Assembly Processes for the Car of the Third Millennium (MyCar)” funded by the Commission of the European Union.

6 REFERENCES

Risk Reduction via Prototyping in Customized Product Development

K. Sun, S. L. Chen
Division of Systems and Engineering Management, School of Mechanical & Aerospace Engineering,
Nanyang Technological University, Singapore 639798, Singapore
su0001ke@e.ntu.edu.sg, songlin@ntu.edu.sg

Abstract
Risks are inherent in customized product development for both customers and manufacturers due to their inability to accurately articulate requirements and estimate costs, respectively. The presence of risks creates transactional barriers when decision makers are risk averse. Prototyping, commonly used for customer requirements elicitation and manufacturing cost estimation, is interpreted in this paper as a means of risk reduction and modelled via a Bayesian estimation process. A quantitative risk model is subsequently developed to investigate the investment decision upon prototyping, taking into consideration the fidelity and cost of the prototype. This paper provides a decision framework for practitioners to understand and manage transaction risks in customized product development.

Keywords:
Risk reduction, customization, prototype, product development

1 INTRODUCTION
Customized products are designed and manufactured to fulfill the particular needs of individual customers [1]. There is an increasing output of customized products, spanning from capital goods like machinery, network servers, and information systems to consumer goods like personal computers, cars, golf clubs, and sneakers among many others [2]. Product customization has been recognized as a frontier for manufacturers in many different industries to gain a competitive edge in an increasingly diversified and dynamic marketplace [3]. However, in customized product development, both customers and manufacturers are faced with risks. Given the large solution space implied in customization, it is often difficult for customers to clearly articulate their requirements; and it is also difficult for manufacturers to clearly communicate their capabilities in sufficient details without confusing customers.

Prototypes are commonly used for risk reduction in product development [4]. For instance, in concept development, experimental prototypes can be built and tested to elicit customer needs; in detail design, prototypes are often created for customers’ reviews and comments; in testing and refinement, prototypes are tested to determine whether the product works as designed and whether the product meets customer needs [5]. Prototypes can be categorized into two types. The first type can be seen as a part of a manufacturer’s internal performance testing, assessing functionality or verifying fitness. The second type of prototype is primarily used for enriching communication between manufacturers and customers. For instance, architects often construct models of buildings to get design feedback. Through such prototypes, customers can convey their reviews back to manufacturers as well as to refine their requirements and update their estimated value of the final product. In this sense, prototypes can be taken as a tool for information collection and communication, which reduce uncertainties and risk exposure for both customers and manufacturers.

Despite the fact that prototypes are quite useful in risk reduction, they are not free and can be very costly. For instance, the prototype of an aircraft could cost up to hundreds of millions of dollars. There are a number of significant but difficult questions regarding the investment decision upon prototyping, especially in capital intensive industries where there is normally a high degree of customization. Typical questions include: is it cost-effective to build a prototype? Who should pay for it? And, how much should the final product be priced?

To answer these questions, this paper develops a quantitative risk model from a manufacturer’s perspective. The prototyping process is interpreted as a sampling of the final product with different degrees of fidelity. A prototype with a higher fidelity rate means that it can better represent the final product. The potential of risk reduction through prototyping is then modelled via a Bayesian estimation process. Decision models are subsequently developed to analyze the manufacturer’s decision in product customization, with or without prototyping. The decision models take into consideration both the fidelity and cost of the prototype. Numerical analysis-based simulation is conducted to investigate the prototyping decision with respect to a number of factors, including the manufacturer’s risk attitude, estimated cost and uncertainties of the final product, price quoted by manufacturer, the fidelity and cost of prototype, and the proportion of prototyping cost to be shared by manufacturer. This model thus provides a framework for investment decisions about prototyping in customized product development.

2 RELEVANT LITERATURE
This research relates to prototyping decision-making in product development. The past relevant literature can be...
generally categorized into two streams, which models prototyping either as a “trial and error” process or as a “learning process” to examine the utilities of a series of prototypes.

2.1 Trial and Error

Trial and error is a general method of problem solving. It is a process of reaching the final solution by experimenting with various methods until the error is sufficiently reduced. Prototypes can be seen as experience goods, which are defined as products or services whose quality is difficult to observe before consumption, where quality refers to any valued attribute such as taste, efficiency, or durability [6]. The fidelity of the prototype, i.e., how well the prototype resembles the final product, serves as an indicator of the quality of the prototype. By observing the outcome of the prototyping, designers can update their estimate of the final product. In this sense, prototypes are a means of trial and error to search for a good design solution.

A main drawback of random “trial and error” in prototyping is that it is usually cost and time consuming. The cost and time to build a test prototype depends highly on the available technology and the required degree of fidelity [7-8]. Prototyping costs vary from a few dollars to hundreds of thousands of dollars. For example, manufacturing a physical prototype used in automobile crash tests can cost hundreds of thousands of dollars and may take months to build. In such cases, it is not cost-effective to use a trial and error mechanism.

2.2 Learning Mechanism

Terwiesch and Loch has proposed a learning mechanism to search for product design in a series of prototypes [1]. In general, the customer chooses a design quality threshold as a stopping point and continues prototyping until this threshold is reached. This mechanism is investigated in both unstructured and structured design space. Unstructured design space prevents learning between prototypes. In this case, it is optimal for the manufacturer to offer a linear pricing scheme, and sell prototypes at cost. In structured design space, successive prototypes create learning about the optimal design solution. This method provides a model for the manufacturer to offer prototypes at a profit, at cost, or even for free based on the design problem and market characteristics. This learning mechanism assumes that prototyping is required in the development of custom-design products without regard to the costs and uncertainty outcomes of prototypes.

Although both the “trial and error” and the “learning mechanism” provide a framework to model prototyping decisions in product development, they tend to focus on the engineering aspects. This paper studies prototyping decision, especially its effect on risk reduction, in the context of customized product development. This paper also introduces Bayesian estimation as a novel new method to model prototyping decisions in the product development process.

3 RISK MODELING IN CUSTOMIZED PRODUCT DEVELOPMENT

3.1 Risk sources

In customized product development, both customers and manufacturers are faced with certain level of risk. For customers, a major source of risk stems from their inability to accurately articulate needs in terms of concrete and clear requirements, particularly when the product is complex and customers do not have sufficient technical knowledge [9]. Distorted need information will mislead manufacturers in design problem solving and result in solutions that are not what customers have expected. Costly design changes or disputes may ensue. Another source of risk for customers is their inability to accurately evaluate a customized solution. Customers are often not technically savvy and could get ‘confused’ by the large variety of solutions that are embedded in customization [10].

For manufacturers, a major source of risk stems from uncertainty concerning resources that may be required in product customization. Coinciding with customers’ inability to accurately articulate needs, manufacturers are often unable to accurately communicate their capabilities. It is often hard, if not impossible, to represent or describe a customized solution in sufficient details without confusing customers. Furthermore, manufacturers are often exposed to the risks of requirement changes from customers. Even though customers are contractually responsible for customer-initiated design changes, it is often the case in practice that manufacturers need to modify their solutions to cope with customers’ updated requirements. Such design changes are often costly, especially in the later stage of product development.

3.2 Risk attitude

In general, decision makers can be categorized into three kinds, risk averse, risk neutral and risk seeking, depending on their risk attitudes. In this paper, both customers and manufacturers are assumed to be risk averse, and an exponential utility function is assumed without loss of generality.

\[
u(x) = 1 - e^{-Rx},
\]

where \(u(x)\) represents the utility function, \(x\) is the evaluation measure, and \(R\) indicates the degree of risk aversion. \(R\) is a positive real number and higher value implies less risk aversion.

As the decision faced with the customer and a manufacturer in the contracting stage can be taken as symmetric, this paper focuses on the manufacturer’s decision without loss of generality. The true cost of the final product \(c_N\) remains unknown until observing it after manufacturing. We assume that \(c_0\) is the prior estimation of the product cost before prototyping, which is a random number that is assumed to follow a normal distribution:

\[
c_0 \sim N(\mu_0, Q_0).
\]

As Figure 1 shows, the manufacturer can weigh his decision based on his estimation of the product cost, \(c_0\).

![Figure 1: Manufacturer's Initial Decision Making](attachment:image1.png)

**Deal or No Deal?**

**Deal**

Surplus \(\pi_{\text{Deal}} = P_0 - c_0\)

**No Deal**

Surplus \(\pi_{\text{No Deal}} = 0\)

The manufacturer quotes price \(P_0\) after analyzing his costs, risks and profits. When the manufacturer’s estimated cost is lower than the selling price \(P_0\), making the deal would generate the buyer a surplus of \(\pi_{\text{Deal}}\). If \(c_0\) is larger than \(P_0\), the manufacturer is unwilling to make the deal, which generating 0 surplus.
3.3 Certainty equivalent

\( \pi_{\theta_0} \) is a random number depending on \( c_0 \). To help the manufacturer make the decision in the face of uncertainty, this paper introduces the concept of "certainty equivalent", which transforms a set of random outputs into a certain value taking into account the decision maker's risk attitude. For instance, if \( \hat{x} \) is the certainty equivalent of a lottery \( L \), a decision maker would be indifferent between lottery \( L \) and \( \hat{x} \). The mathematical relationship between lottery \( L \) and \( \hat{x} \) is given as [12]:

\[
\hat{x} = E[u(x)],
\]

where \( x \) indicates the uncertain outcome of lottery \( L \).

Given that the probability density function of \( x \) is \( f(x) \), the certainty equivalent can be solved based on the following relationship:

\[
u(\hat{x}) - E[u(x)] = \int u(x)f(x)dx.
\]

As \( f(c_0) \) is the probability density function of the estimated cost \( c_0 \), the certainty equivalent of \( \pi_{\theta_0} \) is calculated as,

\[
\hat{\pi}_{\theta_0} = \int u(\pi_{\theta_0})f(c_0)dc_0.
\]

\[
\hat{\pi}_{\theta_0} = \int \left(1 - e^{-\frac{\pi_{\theta_0}-c_0}{R}}\right) \left(1 - e^{-\frac{\pi_{\theta_0}-c_0}{2\sigma^2}}\right)dc_0.
\]

\[
u(\hat{\pi}_{\theta_0}) \text{ can be represented in a form of } g(P_0,R,\mu^c,Q^c), \text{ such that the inverse function for } u(\hat{\pi}_{\theta_0}) \text{ is } \hat{\pi}_{\theta_0} = u^{-1}[g(P_0,R,\mu^c,Q^c)]. \text{ By calculations, } \hat{\pi}_{\theta_0} \text{ can be mathematically expressed as:}
\]

\[
\hat{\pi}_{\theta_0} = P_0 - \mu^c - \frac{Q^c}{2R}.
\]

Given that \( \pi_{\theta_0} = 0 \), the certainty equivalent of which is fixed at 0 as well, \( \hat{\pi}_{\theta_0} = 0 \). The manufacturer can thus weigh the decision between Deal and No Deal by comparing \( \hat{\pi}_{\theta_0} \) with \( \hat{\pi}_{\theta_0} \).

Equation (6) indicates that the manufacturer's economic surplus decreases as \( R \) increases, which implies that the value of customization would be less if the manufacturer is more risk averse. Similar conclusions can be drawn for the customer. Thus, the presence of risk creates a barrier in customized product development, as it reduces the perceived value of customization.

4 RISK REDUCTION THROUGH PROTOTYPING

4.1 Prototyping as sampling

Figure 2 shows the manufacturer's decision-making process when he is given a third option of prototyping other than deal or no deal. \( \hat{\pi}_{\theta_0} \geq 0 \) means that without prototyping, making deal is expected to be profitable.

![Figure 2: Manufacturer's third option of prototyping](image)

In product development, the cost of the final product could be reflected from its prototypes in certain degrees of fidelity. A prototype with a higher fidelity rate can be used to better estimate the cost of the final product for the manufacturer. The product cost reflected from prototyping \( c_0 \) can be taken as a sample of actual cost \( C \) distorted by a noise factor \( \epsilon \), which is assumed to be a non-biased normal random variables with variance \( \Sigma \), \( \epsilon \sim N(0,\Sigma) \).

\[
c_0 = c + \epsilon.
\]

\( \Sigma \) indicates the fidelity of the prototype, with lower variance implying higher fidelity. Prototypes that are closer to the final production of the product generally have higher fidelity than those closer to the early conceptual design stage. Before prototyping, the manufacturer's best estimate of the value of \( c \) is \( c_0 \):

\[
c_0 = c_0 + \epsilon,
\]

\[
c_0 \sim N(\mu^c,Q^c+\Sigma).
\]

4.2 Bayesian updating

Conditional on the outcome of prototyping, the manufacturer can update his estimated cost of the final product. The updating process can be generally modelled via Bayesian estimation. The posterior estimated cost after prototyping is represented by \( c_0 \), which is assumed to follow normal distribution with mean \( \mu^c \) and variance \( Q^c_0 \):

\[
c_0 \sim N(\mu^c,Q^c_0 + \Sigma).
\]

The mean (\( \mu^c \)) and variance (\( Q^c_0 \)) can be calculated via Bayesian updating as:

\[
\mu^c = \mu^c_0 + \frac{(c_0 - \mu^c_0)Q^c_0}{Q^c_0 + \Sigma},
\]

\[
Q^c_0 = \frac{Q^c_0 \Sigma}{Q^c_0 + \Sigma}.
\]

From equation (11) and (12), it can be observed that the mean value of the new estimates will shift from its initial value \( \mu^c_0 \) towards \( \mu^c \) after prototyping. The variance of the new estimates, \( Q^c_0 \), is smaller than that of initial estimates, \( Q^c_0 \), which indicates risk reduction. Thus, if the mean cost does not increase dramatically while risk reduction is significant, the transaction could become possible.
4.3 Prototyping decisions

The Bayesian estimation process described above provides a qualitative interpretation of the use of prototypes for risk reduction. However, prototypes can be very expensive in some cases. This section develops a quantitative model to assist decision-making, regarding whether a prototype is justified under different situations. The decision-making procedures are summarized in Figure 3.

The manufacturer could decide to have a deal or no deal after prototyping. However, the question remains in terms of favouring prototyping or not. The manufacturer may only want to build a prototype when his expected surplus after prototyping, \( \hat{\pi}_{11} \), is higher than the expected surplus without a prototype, which can be represented as:

\[
\max(\hat{\pi}_{01}, \hat{\pi}_{02})
\]

Suppose the cost of prototyping is \( d \), and manufacturer shares \( \omega \) (0 ≤ \( \omega \) ≤ 100%) proportion of the cost, which is \( \omega d \). The manufacturer asks for a new price, \( P_1 \), which is adjusted after observing the prototype. \( \pi_{11} \) and \( \pi_{12} \) are the surpluses of deal and no deal after prototyping, respectively, with certainty equivalents \( \hat{\pi}_{11} \) and \( \hat{\pi}_{12} \).

By adopting the same method of calculating \( u(\hat{\pi}_{01}) \) and \( u(\hat{\pi}_{11}) \) and \( u(\hat{\pi}_{12}) \), they are represented as below:

\[
u(\hat{\pi}_{11}) = \int u(\pi_{11}) f(c_{p}) dc_{p}
\]

\[
= \int \left(1- e^{-\frac{P_1 - \omega c_{p} - \omega d - Q_{c}^{0}}{\delta}} \right) \left(1 - \frac{\omega d}{2\pi Q_{c}^{0}} \right) dc_{p}, \quad (13)
\]

\[
\hat{\pi}_{11} = P_1 - \mu_{c}^{e} - \omega d - \frac{Q_{c}^{0}}{2R}, \quad (14)
\]

where \( \mu_{c}^{e} = \mu_{c}^{e} + \frac{(c_{p}^{e} - \mu_{c}^{e})Q_{c}^{0}}{Q_{c}^{0} + \Sigma} \) and \( Q_{c}^{0} = \frac{Q_{0}^{0} + \Sigma}{Q_{c}^{0} + \Sigma} \).

Given that \( \pi_{12} = -\omega d \), the certainty equivalent of which is \( \hat{\pi}_{12} = -\omega d \) as well. The choice between \( \hat{\pi}_{11} \) and \( \hat{\pi}_{12} \) depends on several parameters, that include \( \omega, d, P_1, R, \mu_{c}^{e}, Q_{c}^{0}, \Sigma, \) and \( c_{p}^{e} \). All other variables could be assumed as given, except \( c_{p}^{e} \), which is a random number with mean \( \mu_{c}^{e} \) and variance \( (Q_{c}^{0} + \Sigma) \). This decision-making process could be seen as a lottery with possible outputs \( \hat{\pi}_{11} \) or \( \hat{\pi}_{12} \) depending on the outcome of prototype, \( c_{p}^{e} \).

\( c_{p}^{e} \) is defined as a threshold of \( c_{p}^{e} \), where the manufacturer will choose to make a deal if \( c_{p}^{e} > c_{p}^{e} \), otherwise no deal. \( c_{p}^{e} \) if the value that satisfies the condition:

\[
\hat{\pi}_{11} = \hat{\pi}_{12}.
\]

\[
u(\hat{\pi}_{1}) = \int u(\hat{\pi}_{11}) f(c_{p}) dc_{p} + \int u(\hat{\pi}_{12}) f(c_{p}) dc_{p}. \quad (15)
\]

\( \hat{\pi}_{1} \) can be calculated from the inverse function of equation (15). Thus, the manufacturer could make the prototyping decision based on the comparisons between \( \hat{\pi}_{01} \) and \( \hat{\pi}_{1} \). In the case of no deal, where \( \hat{\pi}_{01} < 0 \) and \( \hat{\pi}_{1} < 0 \), the decision maker could make the deal possible through adjusting some parameters, such as \( d \) or \( \omega \).

Four simulated studies are investigated with different settings in Section 5 to provide a quantitative guide in applying the model.

5 SIMULATIONS

The risk model is implemented in Mathematica® with parameter settings as in Table 1. The input parameters include the manufacturer's initial estimates of the product cost, \( c_{0} \sim N(\mu_{c}^{e}, Q_{c}^{0}) \), risk attitudes, and the fidelity and cost of the prototypes. Four scenarios are investigated in the simulation study. In setting 1, the relationship between \( \mu_{c}^{e} \) and expected surpluses (\( \hat{\pi}_{01} \) and \( \hat{\pi}_{1} \)) is studied.

Setting 2 studies the effects of cost variance \( Q_{c}^{0} \), which represents the risk level in product cost estimation.

Setting 3 focuses on the changes of \( \pi_{12} \) with

Figure 3: Decision making process
respect to the increase of $d$. The output is concerned with choices among expected economic surpluses at different conditions. Last but not least, setting 4 investigates the prototype cost sharing $\omega$, which is an important factor in contracting between customer and manufacturer.

| Setting | $\mu^c_0$ | $Q^c_0$ | $R$ | $P_0$ | $P_1$ | $\Sigma$ | $d$ | $\omega$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0–10</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>0.75</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0.1–10</td>
<td>1</td>
<td>10</td>
<td>0.75</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>0–1</td>
</tr>
</tbody>
</table>

Table 1: Variables and descriptions

5.1 The effect of mean product cost, $\mu_0$

The manufacturer's risk attitude can be measured in a quantitative way. In this scenario, $R$ is assumed to be 1, which indicates that the manufacturer is risk averse. Given $Q^c_0 = 5$, the manufacturer is uncertain about cost. The variance of fidelity is given as $\Sigma = 1$, which indicates a fairly good fidelity rate. The prototype costs $d = 0.5$, which is 5% of the initial price, $P_0 = 10$. The manufacturer shares 50% of prototype cost, $\omega = 0.5$. $P_1$ decreases to 9.75 in order to make up for customer's prototype cost. The decision outcome is illustrated in Figure 4, which is plotted with the horizontal axis representing estimated mean value of product cost, $\hat{\mu}_0$, and the vertical axis representing $\hat{\pi}_0$ and $\hat{\pi}_1$.

The results show that expected economic surpluses ($\hat{\pi}_0$ and $\hat{\pi}_1$) are decreasing with respect to the increase of $\mu_0$. $\hat{\pi}_0$ and $\hat{\pi}_1$ cross at (5.8, 1.9). This indicates that beyond this threshold point, prototyping is expected to generate more surpluses for the manufacturer. If the expected cost is not significant, say $\mu_0 = 4$, there's no need to prototype in this situation.

5.2 The effect of product cost variance, $Q_0$

It is assumed that $c_0 \sim N(7, 5)$. The manufacturer has the same risk attitude as the previous cases, $R = 1$. In this scenario, it is given that $P_1 = P_2 = 10$, which means that the manufacturer doesn't compensate the customer for prototyping cost, 50% of which is borne by manufacturer. The results are shown in Figure 5.

$\hat{\pi}_0$ is a horizontal line, as it is the certainty equivalent before prototyping, not a function of $d$. $\hat{\pi}_1$ decreases with the increase of $d$. The two lines cross at (2.1, 0.5), which means that it is worth to prototype when $d < 2.1$ with the given parameter settings.

5.3 The effect of prototype cost, $d$

In setting 4 the prototyping cost is fixed to 5. $\hat{\pi}_0$ and $\hat{\pi}_1$ cross at (0.2, 0.5), which means that in this given situation, the manufacturer can only bear less than 20% of the prototyping cost in order to profit from this deal.

5.4 The effect of prototyping cost sharing, $\omega$

In setting 1 the prototyping cost is fixed to 5. $\hat{\pi}_0$ and $\hat{\pi}_1$ cross at (0.2, 0.5), which means that in this given situation, the manufacturer can only bear less than 20% of the prototyping cost in order to profit from this deal.
6 AN ILLUSTRATIVE CASE

This section provides the simulation of an illustrative case which helps to explain how the risk model can be applied to a practical design scenario.

In the case of house construction, an architect (manufacturer) cannot identify what the customer wants. As a result he is uncertain of his estimated cost due to the information barrier. Through the initial brief discussion with the customer, the architect may roughly estimate the cost $c_2$ based on the house size, style, material, decoration, etc. Here, two cases are designed to illustrate the effects of prototyping in risk reduction.

In case 1, $c_2$ is assumed to follow a normal distribution with mean $\mu_0 = 7$ ($\text{million}$) and variance $Q_0 = 5$ ($\text{million}$); and in case 2, $c_2$ follows a normal distribution with mean $\mu_2 = 7$ and variance $Q_2 = 10$. Assume that through a specifically designed survey, the customer's risk tolerance, $R$, is measured to be 1 ($\text{million}$). The price that architect asks for is $P_0 = 10$ ($\text{million}$). The settings of the scenario are summarized in Table 2.

Table 2: Architect's initial setting

<table>
<thead>
<tr>
<th>Case</th>
<th>$\mu_0$</th>
<th>$Q_0$</th>
<th>$R$</th>
<th>$P_0$</th>
<th>$d$</th>
<th>$\omega$</th>
<th>$\hat{x}_0$</th>
<th>$\hat{x}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.78</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>3</td>
<td>0.5</td>
<td>-2</td>
<td>0.52</td>
</tr>
</tbody>
</table>

The certainty equivalent of the architect's surplus without prototyping, $\hat{x}_0$, is calculated to be 0.5 in case 1 and -2 in case 2. This means that in case 1, the architect is willing to make the deal; however, he declines the deal in case 2 because of the high risk.

These two cases are illustrated in Figure 8, which is presented in 3-dimensional figure with horizontal axes representing $R$ and $Q_0$, vertical axis representing $\hat{x}_0$, and the horizontal surface indicating the frontier of $\hat{x}_0$. 

Figure 8: Architect's surplus without prototyping

The prototype of a construction project could be in different forms, such as detailed drawings, 3D rendered graphs, or small scale physical model. Such prototypes help the architect collect more information which results in a more accurate estimation of the final product cost.

In addition to the initial settings without prototyping (Table 2), the prototyping cost is assumed to be, $d - 1$ ($\text{million}$). This cost consists of consulting fees, model construction fees, etc. The architect and the customer share half of the prototyping cost each, $\omega = 0.5$. The prototype is assumed to well represent the final product, $\Sigma = 3$. The architect charges the same amount even after prototyping, $P_t = 10$.

The same prototype settings (Table 3) are applied to both cases. Then, the certainty equivalents of the architect's surplus after prototyping, $\hat{x}_1$, could be found for each case.

Table 3: Prototype setting

<table>
<thead>
<tr>
<th>Case</th>
<th>$\mu_0$</th>
<th>$Q_0$</th>
<th>$R$</th>
<th>$P_0$</th>
<th>$\hat{x}_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>-2</td>
</tr>
</tbody>
</table>

In case 1, $\hat{x}_1 = 0.78$ which is higher than $\hat{x}_0 = 0.5$. This means that the architect is willing to prototype even though he could make the deal without prototyping. This is because of the risk reduction and relative low prototyping cost. For case 2, $\hat{x}_1 = -2$ which is significantly higher than $\hat{x}_0 = -2$, this positive result shows that the architect is willing to return to deal although he refused it before prototyping. This change from no deal to deal explains the risk reduction effect of prototyping. The results are shown in Figure 9.
7 CONCLUSION

The risk model developed in this study shows that the inherent uncertainties and the consequent risks concerning the value and cost of product create barriers for customers and manufacturers to engage in making deals in product customization. Unlike the prevailing research considering series of prototypes in product development, this study provides a quantitative guide for manufacturers in making the prototype-initiating decision. The use of prototypes in product development is interpreted as a means for risk reduction. Information updating via prototyping is modeled as a Bayesian estimation process, which succinctly captures the dynamics of risk evolution. This paper also develops a prototyping decision model based on risk analysis, taking into consideration a number of factors including manufacturers' risk attitudes, their initial estimation accuracies, and the fidelity and cost of the prototype, as well as the proportion of prototype cost shared by the manufacturer. A numerical study based on simulation reveals that an informed decision upon prototyping requires an intricate balance among the fidelity and cost of the prototype, sharing cost and price of the final product.

To fully understand the risks associated with customized product development, this study can be extended and enriched in a number of directions. First, this model can be developed to investigate multiple prototypes in product development. A second direction to extend this research is to consider multiple customers and manufactures in a more general sales or procurement scenario concerning product development. Although the competition among multiple parties makes decisions more complex, it provides incentives for truthful information sharing as well as a pricing mechanism. This paper makes a contribution towards this end by providing a general risk model.

8 ACKNOWLEDGMENTS

This research is supported by the Academic Research Fund Tier-1 (RG 27/08) of Singapore for research in “Product Line Design and Strategic Platform Planning for Mass Customization”.

9 REFERENCES


Abstract

The research of the authors within product development is addressed at developing a tool for innovation in industry, namely Integrated Product and Process Re-engineering (IPPR), whose overall objective is enhancing and harmonizing ideation, design and manufacturing with a product lifecycle approach. The module of IPPR indicated as Process Value Analysis (PVA) is aimed at ranking the phases of a business process according to their contribution to customer satisfaction with respect to the employed resources. The original contribution of the present paper is complementing the method with information concerning drops in customer satisfaction as a result of poorly performed process phases. By accepting the non-linear relationship between satisfaction and attributes' quality level and the different roles played by customer requirements according to Kano categories, the authors propose a preliminary method to provide quantitative evaluations of the effects of process phases that do not thoroughly fulfill the intended objectives. An exemplary application here presented refers to the cosmetic industry, by investigating the production process of lipsticks, to which PVA was previously applied with encouraging outcomes.

Keywords

Process phases performance, customer satisfaction, Kano categories, Process Value Analysis, lipstick production

1 INTRODUCTION

Several works in the literature demonstrate that Business Process Reengineering (BPR) tasks have failed to pursue the objective of generating value for companies' customers. Among others, Holland and Kumar [1] assessed that 60–80% of BPR initiatives haven't fulfilled the declared objectives. More specifically, the analysis performed by Hall et al. [2] remarks that, in order to carry out successful BPR initiatives, redesign efforts should be focused not only on cost and time reduction, but mainly on the areas of the business process which have the most direct impact on customer value. These results show that managers must reengineer their core processes starting from the customer perspective.

The focus on customers and on their perception of products and services has consequently played a relevant role within companies by affecting industrial practices, policies [3] and product development cycles, leading moreover to a great impact on the required managerial skills [4].

In this context the authors have developed a methodological approach namely Process Value Analysis (PVA), aimed at analyzing business and manufacturing processes and identifying the contribution of each phase in determining customer satisfaction through quantitative metrics [5]. The overall objective of this research is the enhancement of BPR practices by providing information about the phases which generate consistent value for the customers and into which improvement efforts should be channeled. In this way, re-engineering actions can focus on the industrial activities which may be substantially improved through the adoption of novel or emerging manufacturing activities in terms of expected performance and employed resources by aligning them with the whole production process [6].

The first version of the methodology has been applied to several case studies, obtaining positive feedback that have highlighted its potential. However, in order to enhance the robustness of the approach and widen the applicability of the methodology, some development guidelines have been identified. Among them are the need to make the methodology more systematic with a particular emphasis on the partial subjectivity of certain steps and to improve the models that are used to evaluate the extent of customer satisfaction by taking into account their performance and reliability. This is particularly important because PVA [5] assumes that each phase works at its maximum potential, thus neglecting the effects on customer satisfaction played by limited robustness and/or malfunctions of the process.

The link between process performance and expected customer appreciation (and consequently the market success) of products and services is relevant not only for anticipating the drawbacks of mistakes and negligence, but also for supervising ongoing activities. In the industrial world, increasing interest is addressed by predictive metrics for anticipating the outcomes of a business activity and by monitoring its related processes [7].

With the aim of widening the set of indications provided by the PVA method, this paper presents a preliminary investigation of models for assessing the impact of process performance on customer satisfaction.

Section 2 summarizes the logic of PVA, then overviews the state of the art about the models providing quantitative evaluations of customer satisfaction. In section 3 the integration of the most acknowledged models within the PVA module for customer satisfaction...
analysis is presented. Section 4 reports the application of the proposed models to a case study in the cosmetics industry, with the aim of showing their functionality. The outcomes of the case study are discussed in Section 5, while the considerations about the preliminary model are drawn in Section 6. Section 6 also outlines expected future activities.

2 STATE OF THE ART

2.1 Brief outline of the PVA methodology

The PVA methodology follows a step by step procedure, aimed at supporting business process re-engineering activities based on the evaluation of the impact of each process phase on the perceived value of products and/or services and taking into account the employed resources. The results of this assessment are used to identify suitable guidelines for process evolution strategies which will preserve and improve the market competitiveness of products and services.

The first stage concerns the gathering of the information related to the business process under investigation. IDEF0 models are suitable tools to represent the flows of information and materials along the process phases, as well as the employed technology, machinery, human skills. Complementary data are collected to map the expenditures and the timing of each process phase.

The next step of the methodology regards the investigation of the customer requirements that are intended to be delivered during the business process. Each of these product/service attributes are characterized in terms of Kano categories (One-dimensional, Attractive, Must-be) [8] and relevance $R$ [9], meant as the relative importance of the customer requirements within the bundle of benefits provided by the business process (on a scale ranging from 1 to 5). Both the Kano category and the attribute relevance are established through customers interviews or stated by business experts.

Then, on the basis of business experts' evaluation, the coefficients $k_j (0 < k_j < 1)$ that relate each phase $i$ to each attribute $j$ are identified, in terms of the accounted contribution of the process stages to fulfill the customer requirements when delivered at their maximum achievable offering level. Such parameters allow a quantitative link to be established among the process steps and the arisen benefits. The subsequent result is the evaluation of the role played by each phase contribution in both providing unexpected benefits perceived in the marketplace (Customer Satisfaction, CS) and avoiding user discontent (Customer Dissatisfaction, CD), according to mathematical expressions dealing with the implications of Kano model fundamentals. Thus, at the current step, the PVA evaluates the contribution of each process phase in determining customer satisfaction under the hypothesis of achieving the greatest potential, without taking into account the effects of poor repeatability of the outputs and underperformances.

According to Value Engineering, employed resources, costs and time necessary to carry out the process activities are measured and compared with the terms expressing the extents of customer satisfaction. Thus the phases are estimated in terms of their capability to provide both basic quality and unexpected features.

However, such a step is not relevant for the scope of the present paper, since the purpose of this contribution concerns a deeper investigation of the value offered to the customer by each process phase. More in detail, in order to evaluate the effects of process underperformances on the delivered value, the objective is to establish a relationship among the phases' performance and customer satisfaction. Therefore, while $k_j$ coefficients represent the connection between the process stages and the customer requirements, it is necessary to quantitatively link the attributes quality with customer satisfaction parameters.

2.2 Models for quantitative evaluation of customer satisfaction

The task of a quantitative evaluation of customer satisfaction according to the level at which the attributes are fulfilled has been already investigated in the literature. However, an acknowledged well validated model has not been identified yet. By recognizing the non-linear relationship between satisfaction and quality, as well as the different relevance shares of the attributes, the Kano curves remain the reference model for this research.

Thus, a sort of mathematical extrapolation of the curves standing out for must-be, one-dimensional and attractive features, has represented the starting point for building such models. By taking into account the shape of the curves, Tan and Shen [10] relate the ratio between two degrees of satisfaction to the ratio between the corresponding performance levels, according to attributes Kano categories. With respect to their formulation:

$$S_j = \left( \frac{P_j}{P_0} \right)^{k_j}$$  \hspace{1cm} (1)

where $S_j$ and $S_0$ are the extents of satisfaction matching the performance levels $P_j$ and $P_0$, respectively, and the coefficient $k_j$ depends on the attribute category according to the Kano classification: for attractive attributes, $k_j > 1$; for one-dimensional attributes, $k_j = 1$; and for must-be attributes, $0 < k_j < 1$. However $k_j$ commonly assumes the values $k=2$ and $k=1/2$ for attractive and must-be customer requirements, respectively. The ratio between the satisfaction levels is commonly indicated as adjusted Improving Ratio ($IR_{adj}$), while the ratio between the performances is generally indicated with Improving Ratio ($IR$), as depicted in Figure 1(a). It is worth noting that prior to the publication of models for the quantitative evaluation of customer satisfaction investigating the non-linear relationships between achieved performances and originated benefits, only the $IR$ was commonly used.

The expression (1), developed to support Quality Function Deployment (QFD) tasks, was first applied to an exemplary case in the Internet field and has been further employed in different domains, such as catering [11], mobile communication [12], education [13].

A modification of the recalled formula has been recently proposed by Chadua et al. [14], who introduced a corrective coefficient $m$ in the relationship between the improving ratios:

$$IR_{adj} = IR \times (1 + m)^k$$  \hspace{1cm} (2)

The $m$ coefficient, with $0 < m < 1$, depends on the distribution of the attributes within the Kano categories, as well as on their relevance. The authors of the model describe an application of the proposed model by surveying the field of webpage design.

The models proposed in [10] and [14] aim to perform a comparison among different product profiles or to evaluate the effects of improvement scenarios by using a QFD approach, but they have never been used for other tasks.

In the literature there are also other models to evaluate customer satisfaction that are supposed to overcome the
rigid distinction among Kano categories. However such approaches haven't yet found an appreciable consensus, nor a wide application.

"A-Kano" [15] considers two different importance scores for each attribute, by separately taking into account their supposed relevance in being triggers for customer satisfaction and means for avoiding discontentment. Subsequently, the product profile is summarized in a bi-dimensional graph, where the two different relevance scores represent the coordinates. By placing the attributes in the graph, depicted in Figure 1(b), their categorization depends on the inclination of the segment connecting the representative point with the origin, while their overall importance is evaluated through the length of the same segment. Falls in the attribute performances determine proportional drops in the capability to generate satisfaction and avoid dissatisfaction. The new representative point thus stands on the same segment, but at a minor distance from the origin. A demonstrative application regards the automotive field.

With respect to [10], [14] and [15], Xi et al. [16] propose a different model to represent the satisfaction curves on the basis of attribute quality. Their characterization first considers the supposed maximum dissatisfaction (minimum attribute quality) and the greatest satisfaction (maximum attribute quality). Subsequently, the two representative points are connected by a 3-segment broken line, whose inclination in the three portions depends on the local rates of satisfaction growth, as in Figure 1(c). The curves shown in the manuscript refer to attribute investigations within the financial field. The model depicts then very different shapes of the curves, whose end segments of the broken lines are however frequently flat, recovering therefore the concept of satisfaction thresholds expressed in [17].

![Figure 1: Models for quantitative evaluation of customer satisfaction as a function of attributes performance [13, 15, 16].](image)

2.3 Critical review of the existing models

On the basis of the previously performed analysis of existing approaches for the quantitative estimation of customer satisfaction, the purpose is to choose a reference model to be embedded in the PVA. In order to fulfill this task it is necessary to make further considerations about the reliability and the applicability of the recalled models.

None of the models foresee markedly different patterns for customer satisfaction and dissatisfaction, neglecting the dissimilar effect they give rise to. Just as in "A-Kano", the two indexes follow different functions; however the simultaneous modification of customer satisfaction and dissatisfaction doesn't comply with the trajectories of Kano curves.

Furthermore, in [10] and conversely in [14], the employed approach is tailored for incremental levels of attributes' quality modifications and, in such a context it can be advantageously applied just by considering positive values of satisfaction.

On the contrary the model depicted in [16] envisages both negative and positive values, allowing a graphical representation for the whole range of potential attributes' performance.

However, beyond their more limited scientific evidence, the main limitation of the models represented in [15] and [16] relates to the need of introducing further coefficients, mostly consisting of personal estimations of the attributes' roles. Such issues can potentially increase the degree of subjectivity impacting the outcomes of the PVA, hindering the pursued goal of enhancing the systematic level of the whole methodology.

3 EVALUATING THE IMPACT OF PRODUCT DEVELOPMENT PHASES ON CUSTOMER SATISFACTION

This Section describes the criteria leading to the choice of the reference model, according to the literature review and the PVA requirements. Then the integration of such a model within the evaluation module of PVA is described. The readers who don't know the details of PVA proposed in [5-6] can better comprehend the contribution of the current paper through the exemplary application reported in Section 4.

3.1 Adapting the reference model for its implementation within PVA

A suitable model for satisfaction estimation within the PVA should:

- express mathematical formulas for both customer satisfaction (CS) and dissatisfaction (CD) indexes, according to the various categories of Kano model;
- consider the minimum and the maximum extents of both coefficients with respect of the criteria established within Kano framework;
- relate both coefficients with attributes quality ranging from 0 to 100%.

According to the performed survey and the limitations of the reviewed models, the authors have chosen a reference framework and adapted it according to the intended purposes. The choice of employing the model...
described in [10] as a starting point for developing a novel framework to measure customer satisfaction indexes lies on the following observations:

- it can be considered the only contribution facing significant scientific evidence;
- it is based on a mathematical interpretation of the trajectories of the original Kano curves;
- under certain circumstances and objectives, resumed further on, no additional parameters must be defined in order to make it suitable for the implementation within PVA.

Indeed, although it is not possible to adopt the model “as is”, the criteria involving linear, quadratic and square root like progress of the improvement ratios can be maintained. With reference to concepts and equations belonging to the original Kano formulation or to acknowledged scientific contributions aimed at developing its model [18-22], the authors have taken into consideration the following hypotheses in order to build the novel framework:

- Customer Dissatisfaction \( CD \) that is neglected in the reference model [10], has to be considered as the extent of avoided discontentment when delivering a certain attribute at a proper level (with the minus sign); hence, it makes reference to the maximum dissatisfaction that a poor attribute quality can deliver.
- Customer Satisfaction \( CS \) will be considered as the extent of generated unexpected benefits, thus considering the satisfaction curves above the zero-level.
- Attractive attributes play no role within \( CD \). Their maximum amount of generated \( CS \) equals to their relevance \( R \). In case of minimum quality the \( CS \) turns to 0.
- For must-be attributes, the progress of the satisfaction related to (1), stands for the avoidance of potential dissatisfaction when the performance of the customer requirement grows. The same concept can be employed even for one-dimensional features when considering the negative part of the Kano curve.
- Must-be attributes play no role within \( CS \). Their maximum amount of \( CD \) is equal to the absolute value of their relevance \( R \). In the event of their lowest quality, the \( CD \) turns to 0.
- One-dimensional attributes influence both \( CD \) and \( CS \). The relevance \( R \) characterizes the extent of both the generated benefits and the avoided dissatisfaction. Hence, due to such symmetry:
  - They generate \( CS \) when their performance overshoots 50% of the range; the maximum \( CS \) equals to \( R \).
  - The related \( CD \) equal to \(-R\) when the performance is greater than 50%; \( CD \) turns to 0 in event of the minimal performance.
- Both \( CD \) and \( CS \) are proportional with the attribute relevance \( R \).

According to the above hypotheses about Kano categories, Table 1 summarizes the equations for calculating the extents \( CS \) and \( CD \) with reference to the \( j \)th customer requirement when varying its degree of fulfilment \( q_j \) (like the performance \( p \) in the model represented in [10]).

<table>
<thead>
<tr>
<th>Kano category</th>
<th>Quality range</th>
<th>( CS )</th>
<th>( CD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Must-be</td>
<td>( 0 \leq q_j \leq 1 )</td>
<td>0</td>
<td>(-R q_j)</td>
</tr>
<tr>
<td>One-dimensional</td>
<td>( 0 &lt; q_j &lt; 0.5 )</td>
<td>0</td>
<td>(-2R q_j)</td>
</tr>
<tr>
<td>One-dimensional</td>
<td>( 0.5 &lt; q_j &lt; 1 )</td>
<td>( R(2q_j-1) )</td>
<td>(-R)</td>
</tr>
<tr>
<td>Attractive</td>
<td>( 0 &lt; q_j \leq 1 )</td>
<td>(-R q_j)</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Equations of the satisfaction indexes.

### 3.2 Implementation of the presented metrics within the PVA methodology

The PVA methodology evaluates the process phases according to their contribution to customer satisfaction through the determination of proper indexes. The novel evaluation module requires the input data related to the business process that are introduced throughout the PVA previously to the calculation of satisfaction indexes. This includes:

- the process phases;
- the customer requirements, their relevance and classification according to Kano model;
- the extent of each phase in fulfilling each product and service attribute \( k_i \) coefficients;
- the performance of each phase in determining product or service attributes \( p_i \) coefficients. This is assessed through a score ranging from 0 to 1, in accordance with the specific productivity of the phase, or other metrics related to the compliance of its outcomes to the expected characteristics. If these data are not available, the score is assigned by the experts according to their knowledge of the process performances through a discrete evaluation, e.g. with 5 discrete performance levels.

Given this starting point, the evaluation module consists of a procedure that includes the following calculation stages:

1) the determination of the attributes’ quality \( q_j \) through the expression (3);
2) the estimation of modified \( CD \) and \( CS \) coefficients through the formulas summarized in Table 1;
3) the calculation of the sums of current \( CS \) and \( CD \) values, indicated as modified Global Customer Satisfaction/Dissatisfaction \( GCS_{mod} \) and \( GCD_{mod} \), representing therefore the overall amount of generated satisfaction;
4) the calculation of the sums of maximum \( CS \) and \( CD \) values, indicated as hypothetical Global Customer Satisfaction/Dissatisfaction \( GCS_{Hyp} \) and \( GCD_{Hyp} \), representing therefore the achievable amount of generated satisfaction in event of best phases performances \( p_i=1 \forall i \);
5) the evaluation through the expressions (4) and (5) of not-exploited margins of satisfaction, due to limited performances, both in terms of supplied benefits \( \Delta CS \) and avoided dissatisfaction \( \Delta CD \):

\[
q_j = \sum_p p_j \times K_{ij} \quad (3);
\]

\[
\Delta CS = GCS_{mod} - GCS_{mod} \quad (4);
\]

\[
\Delta CD = GCD_{mod} - GCD_{mod} \quad (5).
\]

Equation (3), determined as a modification of the model employed within PVA, stands for a preliminary evaluation of the effects of poorly performed phases on the attributes' degree of fulfilment. At this stage of the research, the authors aim to take into account just the
linear impacts of phases performances, without considering second order interrelations among the phases in the effective determination of the attributes quality values. The performed module is viable to evaluate the effects of a single phase showing non-optimal performance, as well as the impact of a set of $p_i$ scores attributed to various process tasks and operations.

4 EXEMPLARY APPLICATION
This Section illustrates the application of the Process Value Analysis for an industrial case with a particular reference to the module for investigating the effects of poorly performed phases.

4.1 Outline of the case study and of the application of Process Value Analysis
The cosmetics industry in Italy has undergone relevant transformations in the last years from the viewpoint of the firms operating in such sector. Although activities which were once predominantly craftsmanship tasks have been turned into massive industrial productions regulated by policies which are common for big companies, the manufacturing technology hasn't been subjected to radical modifications. With reference to lipstick design and manufacturing, the case study under investigation deals with the business process of a medium-sized firm, whose main customers are global players in the field of women's make-up products and fashion in general.

<table>
<thead>
<tr>
<th>Phases</th>
<th>A11</th>
<th>A121</th>
<th>A122</th>
<th>A123</th>
<th>A13</th>
<th>A21</th>
<th>A22</th>
<th>A23</th>
<th>A31</th>
<th>A32</th>
<th>A33</th>
<th>A34</th>
<th>A35</th>
<th>A36</th>
<th>A41</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td>Marketing Briefing</td>
<td>Stick technical design</td>
<td>Packaging technical design</td>
<td>Prototyping</td>
<td>Industrial engineering operations</td>
<td>Laboratory research</td>
<td>Laboratory analysis</td>
<td>Development of the formula</td>
<td>Quality control and warehousing of raw materials</td>
<td>Additional manufacturing operations</td>
<td>Bulk manufacturing</td>
<td>Liquid bulk casting</td>
<td>Assembly</td>
<td>Quality control of final product</td>
<td>Warehousing of final product</td>
</tr>
<tr>
<td>CR1 Stick colour</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0.45</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0.05</td>
<td>0</td>
<td>0.05</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>CR2 Stick colour precision</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.45</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR3 Stick taste</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.35</td>
<td>0.4</td>
<td>0</td>
<td>0.05</td>
<td>0.1</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR4 Stick scent</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.35</td>
<td>0.4</td>
<td>0</td>
<td>0.05</td>
<td>0.1</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR5 Absence of foreign bodies in the stick</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.65</td>
<td>0.3</td>
<td>0</td>
<td>0.05</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR6 Stick surface porosity</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.45</td>
<td>0.5</td>
<td>0</td>
<td>0.05</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR7 Lipstick applicability</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR8 Presence of active principles in the lipstick</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.45</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.05</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR9 Lipstick resistance on the lips</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR10 Avoiding irritation phenomena</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0.05</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>CR11 Quantity of product in the lipstick</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.65</td>
<td>0.3</td>
<td>0.05</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR12 Duration of lipstick properties</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td>0.1</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>CR13 Customizable stick shape</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>CR14 Special effects providing</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>CR15 Customizable chemical formulation</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>CR16 Compatibility of the primary packaging with the stick</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>CR17 Colour of the primary packaging</td>
<td>0.2</td>
<td>0</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>CR18 Customizable primary packaging</td>
<td>0.2</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>CR19 Shapes variety</td>
<td>0.2</td>
<td>0</td>
<td>0.4</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>CR20 Resistance of primary packaging</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>CR21 Functionalities of primary packaging</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.2</td>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>CR22 Technical dossier</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>CR23 Product labelling</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.4</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Table 2: Calculation of attributes quality according to phases performances.
Table 3: Calculation of satisfaction coefficients according to attributes quality.

<table>
<thead>
<tr>
<th>Attribute quality (q_j)</th>
<th>Kano category</th>
<th>Relevance (R)</th>
<th>CS_i</th>
<th>CD_i</th>
<th>Maximum CS_i</th>
<th>Maximum CD_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR1</td>
<td>0.50</td>
<td>MB</td>
<td>5</td>
<td>0.0</td>
<td>-3.5</td>
<td>0</td>
</tr>
<tr>
<td>CR2</td>
<td>0.91</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-3.8</td>
<td>0</td>
</tr>
<tr>
<td>CR3</td>
<td>0.73</td>
<td>MB</td>
<td>3</td>
<td>0.0</td>
<td>-2.6</td>
<td>0</td>
</tr>
<tr>
<td>CR4</td>
<td>0.73</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-3.4</td>
<td>0</td>
</tr>
<tr>
<td>CR5</td>
<td>0.72</td>
<td>MB</td>
<td>5</td>
<td>0.0</td>
<td>-4.2</td>
<td>0</td>
</tr>
<tr>
<td>CR6</td>
<td>0.66</td>
<td>MB</td>
<td>3</td>
<td>0.0</td>
<td>-2.4</td>
<td>0</td>
</tr>
<tr>
<td>CR7</td>
<td>0.55</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-3.0</td>
<td>0</td>
</tr>
<tr>
<td>CR8</td>
<td>0.45</td>
<td>MB</td>
<td>3</td>
<td>0.0</td>
<td>-2.0</td>
<td>0</td>
</tr>
<tr>
<td>CR9</td>
<td>0.20</td>
<td>OD</td>
<td>3</td>
<td>0.0</td>
<td>-1.2</td>
<td>3</td>
</tr>
<tr>
<td>CR10</td>
<td>0.53</td>
<td>MB</td>
<td>5</td>
<td>0.0</td>
<td>-3.6</td>
<td>0</td>
</tr>
<tr>
<td>CR11</td>
<td>0.68</td>
<td>MB</td>
<td>3</td>
<td>0.0</td>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td>CR12</td>
<td>0.43</td>
<td>MB</td>
<td>3</td>
<td>0.0</td>
<td>-2.0</td>
<td>0</td>
</tr>
<tr>
<td>CR13</td>
<td>0.74</td>
<td>AT</td>
<td>4</td>
<td>2.2</td>
<td>0.0</td>
<td>4</td>
</tr>
<tr>
<td>CR14</td>
<td>0.80</td>
<td>AT</td>
<td>4</td>
<td>2.6</td>
<td>0.0</td>
<td>4</td>
</tr>
<tr>
<td>CR15</td>
<td>0.20</td>
<td>AT</td>
<td>3</td>
<td>0.1</td>
<td>0.0</td>
<td>3</td>
</tr>
<tr>
<td>CR16</td>
<td>0.64</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-3.2</td>
<td>0</td>
</tr>
<tr>
<td>CR17</td>
<td>0.85</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-3.7</td>
<td>0</td>
</tr>
<tr>
<td>CR18</td>
<td>0.84</td>
<td>AT</td>
<td>5</td>
<td>3.5</td>
<td>0.0</td>
<td>5</td>
</tr>
<tr>
<td>CR19</td>
<td>0.76</td>
<td>OD</td>
<td>4</td>
<td>2.1</td>
<td>-4.0</td>
<td>4</td>
</tr>
<tr>
<td>CR20</td>
<td>0.88</td>
<td>MB</td>
<td>1</td>
<td>0.0</td>
<td>-0.9</td>
<td>0</td>
</tr>
<tr>
<td>CR21</td>
<td>0.88</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-3.8</td>
<td>0</td>
</tr>
<tr>
<td>CR22</td>
<td>0.46</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-2.7</td>
<td>0</td>
</tr>
<tr>
<td>CR23</td>
<td>0.62</td>
<td>MB</td>
<td>4</td>
<td>0.0</td>
<td>-3.1</td>
<td>0</td>
</tr>
</tbody>
</table>

The goal of the firm was to evaluate value bottlenecks and phases reliability within the process (schematized with an IDEF0 framework in compliance with PVA and here omitted due to space reasons) for producing lipstick batches (as summarized Table 2) and to consequently support the ideas of technological enhancement that had been previously undertaken.

The business process under investigation ranges from marketing studies intended to determine the characteristics of new trendy lipsticks to warehousing and distribution activities, throughout the phases concerning laboratory testing, prototyping, raw materials supplying, manufacturing and quality control. A set of 15 phases is deemed to fulfil 23 customer requirements that include functional and emotional features regarding the bulk and the primary packaging, the make-up quality and characteristics, the compliance with the regulations concerning the presence of foreign bodies and the skin irritation, up to the technical documentation for the clients. Table 2 summarizes the phases (listed in the first row) and the recalled attributes (reported in the first column), as well as their mutual influence along the process through the $k_i$ coefficients. Attributes' Kano category and relevance are indicated in Table 3.

For space reasons, this paper omits indications about the employed resources in terms of consumption materials, machinery, structures, energy, human skills, labour and time, that have led to the end results about value estimations. Such final outcomes are briefly summarized in the following:

- certain routine tasks, such as bulk manufacturing, laboratory analysis and raw materials warehousing provide a low amount of value for the business process;
- the technical design of the primary packaging has resulted in the most valuable phase, by consistently contributing to the fulfilment of both basic and unexpected requirements with a limited resources consumption. Analogous conclusions can be drawn also for the outputs of laboratory research;
- several phases concerning product detailed design and manufacturing belong to the basic performance area, whereas the activities ranked in the industrial engineering phase (such as the definition of product specifications, manufacturing planning, determination of materials to be purchased, etc.) show the best value in avoiding dissatisfaction;
- creative tasks give rise to exciting performance, with a particular emphasis on prototyping operations.

The indications issued from the analysis have provided general support for the reengineering directions that the firm was intending to undertake. Indeed, while a technological development was envisaged for the creative phases and the fundamental manufacturing activities, the
outsourcing of routine tasks with perceived low value for the customers was debated within the enterprise management.

4.2 Application of the integrative module

According to the business process described in Section 3 and with reference to Table 2, the bottom row, labelled as “Phase performance”, allows the introduction of $p_i$ coefficients into the worksheet. This set of values represents the input according to which the impact in terms of reduced attribute quality, and consequently unexploited benefits, are evaluated. As a result of phases that partially fulfill the intended purpose ($p_i^{\neq 1}$, highlighted with grey cells background), $q$ values are calculated according to (3) and summarized in the right column, labelled as “Attribute quality”.

In Table 3, on the basis of the quality (as calculated in Table 2), the Kano category and the relevance $R$ of each attribute, the columns labelled as $CS$, and $CD$, indicate the modified contributions to customer satisfaction according to the equations summarized in Table 1. Their sums reported in the bottom row stand for the $GCS_{mod}$ and $GCD_{mod}$ indexes. Conversely, according to the greatest extent of customer satisfaction and dissatisfaction that each attribute can give rise to, the $GCS$ and $GCD$ indexes are calculated; such coefficients are reported at the bottom right of the Table. It is therefore possible to subsequently calculate $\Delta CS$ and $\Delta CD$ through (4) and (5). The exemplary distribution of phases performances reported in Table 2 would thus lead to not exploited margins of about 54% and 20% in terms of customer satisfaction and avoided dissatisfaction respectively.

With the integrative module it is possible to investigate the effects of reduced outcomes regarding each phase singularly and thus the amount of not exploited margins due uniquely to its underperformances. Figure 2 illustrates the effects of the limited performance of various phases, with $p_i$ indexes ranging from 0 to 1, under the condition that all the other process activities are undertaken at an optimal level. The depicted curves refer to process phases with meaningful value indexes according to the outputs of PVA analysis summarized in the Subsection 4.1. The abscessas refers to the phase performance quality, while the ordinates indicated the $\Delta CS$ and $\Delta CD$ coefficients, which are represented with dotted grey and continuous black lines respectively.

5 DISCUSSION OF THE RESULTS

Through the diagrams of Figure 2 it is possible to appreciate the different paths for the fall of customer satisfaction and dissatisfaction, as well as dissimilar maximum extents of unexploited margins. The curves’ trajectories depend consistently on the kind of attributes that are primarily fulfilled along the process phases. On the other side, the peaks of unexploited satisfaction are related with the relevance of the customer requirements and the degree of influence played ($k_j$ coefficients). It is worth to notice that, while some diagrams follow gradual trajectories along the whole abscessas range, in some cases the slope of the curves is significantly different according to the performance level. Thus, with reference to Figure 1b an intermediate level of Laboratory Research performance would result in a consistent drop of unexpected benefits, while it doesn’t meaningfully affect the avoidance of customer dissatisfaction.

Thanks to the introduced model, the firm dealing with the process underpinning lipsticks production can better evaluate the consequences of underperformance and effects of further limited reliability of the phases. A more detailed survey of the business process can also highlight the risks arising from outsourcing, in the event that the accounted activities aren’t undertaken at the current performance, thus providing additional insights for the firm within the strategic decisions regarding the business process. As a whole, the integrative module allows a more careful investigation of the as-is state of the process and of the results of hypothetical to-be situations.

![Figure 2: Not exploited margins of satisfaction according to phases performance.](image-url)
notwithstanding the acknowledged scientific findings it is based on.

Indeed, the planned future activities include an all-
comprising evaluation by the cosmetic firm of the
briefly summarized outputs, assessing whether the
outcomes have been judged reliable and if the
consequent decisions will be supported by positive
confirmations. Furthermore, the experimentation of
the proposed approach, aimed at assessing not only its
effectiveness but also robustness and usability,
represents an ongoing activity.

Another thread of research will involve the implementation
of different models for the evaluation of customer
satisfaction extents according to attributes quality. The
survey will include the models mentioned in Section 2 that
weren’t firstly chosen for the recalled task and the results
will be critically compared. Also the link (3) between
phases’ performance and customer requirements quality
has to be further investigated. More specifically the
interplay of the phases in determining the provided level
of the attributes requires a more detailed examination,
taking into account diversified patterns to affect the
quality, depending on the nature of the product features.

Once the above mentioned open issues will be surveyed
and the employed tools regarding the integrative module
will be validated, analogous evaluations upon the phases’
value can be carried out with respect to those currently
concerning their contribution to customer satisfaction.

Further extensions will regard also the capabilities of the
PVA methodology to interpret further negative outcomes
of the ongoing business process (such as limited market
appeal, non-conformities, unpredicted drawbacks) by
revealing the inadequate phases and pointing out the
measures to be attained and the proper resources to be
channelled.

ACKNOWLEDGMENTS

The authors would like to thank Luca Lazzarini for the
efforts spent in scrupulously analyzing the business
process object of the exemplary application.

REFERENCES

[1] Holland, D., Kumar, S., 1995, Getting past the
obstacles to successful reengineering, Business
Horizons, 38: 79-85.

make reengineering really work, Harvard Business

Relations among TQM practices, plant performance,
and customer satisfaction, Journal of Operations
Management, 17: 59-75.

Knowledge Development: Antecedents and Impact
on New Product Performance, The Journal of
Marketing, 68: 47-59.

Value Analysis for Business Process Re-
eengineering, Proceedings of IMechE Part B: Journal
of Engineering Manufacture, 224: 305-327.

driven Process Value Analysis, in 20th CIRP Design

measurement system design: A
literature review and research agenda, International

Attractive quality and must-be quality, The Journal of


model in the planning matrix of quality function
deployment, Total Quality Management, 11: 1141-
1151.

satisfaction attributes using competitive analysis of
the improvement gap, International Journal of

2007, Integrating Kano’s Model into Quality Function
Deployment to Facilitate Decision Analysis for
Service Quality, in 8th WSEAS Int. Conference on

Evaluation of a Digital Library by Means of Quality
Function Deployment (QFD) and the Kano Model,
The Journal of Academic Librarianship, 36: 128-
132.

2010, Integration of Kano’s Model into quality
function deployment (QFD), International Journal of
Advanced Manufacturing Technology, in press.

Piecewise Regression: A Tool for Understanding
The Relationship between the Attribute-Level
Performance and Customer Satisfaction, in
Conference on Information Management, Cheng
Yun, Taiwan, January 26, 2010.

Discontinuities: A Trigger for Radical Technological
Change, Managerial and Decision Economics, 2-3:
79-97.

customer-defined quality, Center for Quality
Management Journal, 2: 3-36.

[18] Sauerwein, E., Baimi, F., Matzker, K., Hinterhuber,
H.H., 1996, The Kano model: how to delight your
customers, in IX International Working Seminar on
Production Economics, Innsbruck, Austria, February
19-23, 1996.

the QFD using a modified Kano Model, Journal of
the Academy of Business and Economics, 2: 103-
113.

of Kano’s Model Into QFD for Multiple Product
Design, IEEE Transactions on Engineering

Study on the Evaluation of Customers’ Satisfaction -
The perspective of Quality, in 4th International
Quality Conference, Kragujevac, Serbia, May 19,
2010.
State of the Art - Lean Development

U. Dombrowski, T. Zahn, S. Schulze
Institute for Production Management and Enterprise Research, Langer Kamp 19, Braunschweig, Germany
u.dombrowski@tu-bs.de, thimo.zahn@tu-bs.de, sven.schulze@tu-bs.de

Abstract
The role of product development is becoming increasingly important to overall business success. Therefore it is necessary to establish fast, efficient and reliable product development processes. One approach to reach this goal is the implementation of lean thinking in the development processes, e.g. the establishment of a Lean Development System (LDS). LDS are based on the Toyota Product Development System, but also include additional aspects. This paper describes and evaluates the basic approaches of lean development in a literature review.

Keywords:
Lean Production, Lean Development, Lean Innovation, Toyota Product Development System, Global Product Development

1 INTRODUCTION
Manufacturing enterprises must continually improve their processes in order to stay competitive. In the past enterprises have implemented Lean Production Systems (LPS) based on the Toyota Production System (TPS) in order to improve productivity and flexibility [1-3]. An LPS is created using lean principles. Although the application of LPS is not restricted to specific processes or departments, the LPS-principles have mainly been applied in production, assembly, logistics, maintenance, and quality management [2]. In most cases lean principles are not applied to other important processes in the enterprise.

There is a lack of sophisticated concepts available for implementing lean thinking in product development, although product development is rapidly becoming a more important factor in strategic business success than production [4]. In order to reach a sustainable and long-term competitive advantage, it is important for manufacturing enterprises to develop innovative and - from a customer point of view - reasonably priced high quality products in the shortest time possible [5-6]. Enterprises like Toyota which work according to lean thinking in product development develop higher quality products with significantly shorter time-to-markets and for lower costs. [4]. In comparison to its competitors, Toyota is twice as fast, twice as efficient and twice as profitable [7]. This paper starts with a literature review focusing on the application of the lean thinking in product development and the creation of Lean Development Systems (LDS). Afterwards, the LDS concepts presented are evaluated according to criteria like whether a systematic implementation concept is given.

The paper is divided into four parts. First the development process with the most important phases is presented and the goals of product development are identified. Second, the criteria enabling an evaluation of LDS-concepts are formulated. Third, existing LDS-concepts are described from a literature review and evaluated according to the chosen criteria. Finally a summary identifies further areas of research.

2 PHASES AND GOALS OF PRODUCT DEVELOPMENT
Product development is defined as “the process, until a product can be used: Starting with the product planning and the search for ideas, the definition of the product, respectively with the single part production: starting with the order up to the delivery of the product at the customer” [8]. The product development process can either start with an indefinite customer or a determined customer [6]. During the development process, the lifecycle-related characteristics of a product, with increasing degrees of concretion and decreasing insecurity, are determined concerning structure, design and materials [8].

2.1 Phases of product development
Underlying every product is a lifecycle extending from the first idea to the final disposal of the product [9]. The product lifecycle can roughly be divided into three phases: development, market presence, and disposal. Product development itself can be divided into five phases. In practice these phases will overlap instead of taking place in sequence. The specification of the product development process depends on many factors, such as the product, the organization of the enterprise and the available resources [10].

Not every manufacturing enterprise has its own research department. Nevertheless, there is an essential interest in using new discoveries for one’s own enterprise. The goal of research is to systematically gain knowledge in terms of scientific and engineering discovery [11].

Product planning represents the interface between customer and the manufacturing enterprise. Based on market expectations and customer requirements, product planning systematically develops ideas for new innovative products or services while taking the business strategy into account. By determining unique product characteristics, these ideas can be made more specific [12]. Product planning ends with a decision about whether a product concept will be realized and the corresponding development project for that product [12].

The development of an appropriate realization concept, which includes all product functions, is carried out during the subsequent design and development [12]. Here the structure, design, material and lifecycle of the product are determined based on the functional requirements of product planning [8]. Concurrently, a prototype is manufactured and tested.
After the design and development phase, the industrial engineering team plans and controls all technical and organizational tasks necessary for an economic production. The link between industrial engineering and manufacturing and assembly is the start of production. This step is divided into the pilot series, comprising of a preproduction and zero series, and the serial production ramp-up. The purpose of the pilot series is to identify and eliminate any problems which have not been detected in the previous phases of development in order to ensure that all quality objectives are met at the start of production [13].

The final phase of the product development process is manufacturing and assembly. Lean thinking in this phase is accounted for in LPS and therefore is not considered in the LDS concepts. LDS is only applicable from product planning through industrial engineering.

2.2 Goals of product development

Basically, the product development process is part of a manufacturing enterprise, whose main goal is making profit. Successful performance in the market is a result of fulfilling customer requirements while differentiating from competitors [14]. Current literature derives the goals of product development from the general goals of an enterprise. According to [15] and [16] the goals are:

- High quality of products respective services
- Low costs during product development
- Low product lifecycle costs
- Short time-to-market
- High degree of sustainability (social, ecological, economical)
- High degree of innovation
- High product and service acceptance of the customers

3 CRITERIA FOR THE EVALUATION OF LDS-CONCEPTS

The implementation of LPS principles has led to great successes in manufacturing and assembly departments [2]. However, the promotion of the lean thinking in adjacent business processes like product development has not been done, or only fragmentarily pursued. There are many reasons for the absence of lean thinking in adjacent business processes. For one thing, the processes of product development differ significantly from manufacturing, assembly, logistics, maintenance and quality processes. For instance, many enterprises are not aware of the definition of value creation and waste in product development as here material flow comes second to information flow. Furthermore, during product development employee creativity is of particular importance. Employees are given scopes of development which cannot be automatically classified as waste during the creative process. An application of LPS principles to product development without adjustment is consequently not possible [4]. LDS concepts must be derived systematically and tailored to the LPS concept in order to achieve an optimally integration with downstream processes such as manufacturing and assembly. During the development and implementation of a lean development concept it must be ensured that there are no conflicting goals between the LDS-principles and the LPS-principles.

In order to evaluate the approaches described in literature in a systematic and structured way, objective criteria are necessary. Five criteria could be identified which, in part, match the requirements for the implementation of a lean production system as described in [2,4].

3.1 Criterion 1: Development of specific LDS-principles

The continuous and integrated orientation of all development activities according to specific principles is a central criterion of lean product development. These principles have to be used as fundamental guidelines to convey the LDS philosophy and to ensure integrity. Meanwhile, the specific requirements of product development must be taken into consideration (see above). In the following discussion of concepts either existing LPS principles are directly transferred, slightly adjusted, or specific development principles are derived.

3.2 Criterion 2: Integrated perspective and field of application

A systematical application of LDS principles to all phases of product development will ensure an integrated elimination of waste. Focusing on individual activities of product development and implementing individual methods without any linkage to other methods during the optimization efforts should be avoided. Thus, LDS concepts have to take all phases of product development into account, starting with research, product planning, design and development through industrial engineering, including the start of production. Moreover, the interfaces between different departments, especially towards manufacturing and assembly, should be taken into consideration to achieve synchroniziation with LPS rules, standards, methods and tools. Furthermore, the scope of LDS concepts should be extended to all activities and relevant support processes. Relevant support processes include project management [17], knowledge management [18], supplier management [19], quality management [20] and variant management [21].

3.3 Criterion 3: Enterprise-specific selection and configuration of methods and tools

Because there are differences in culture and philosophy among enterprises, and also product and branch specific peculiarities within enterprises it is not possible to define a universally valid LDS. Imitating a successful LDS from another enterprise will not necessarily lead to success because each company may have different requirements. Instead, it is necessary to configure an enterprise specific LDS for each application. This fact is congruent with the configuration of an LPS, which also must obey the specific requirements of the enterprise [1,22]. Using established LDS concepts each enterprise must select its own methods that will meet its specific requirements. LDS methods should be catalogued in a way such that an enterprise can select a method and customize it according to its requirements.

3.4 Criterion 4: Specification of an implementation process

Since the implementation of an LDS involves a fundamental change in the organization and culture of the development department, a well structured implementation process is required. Furthermore, the implementation will take several years since reorganization and employee education is required. To a large extent, obstacles arising during the implementation of LDS include planning mistakes, an inadequate business culture, leadership mistakes, a lack of methodological skills, and an ineffective organizational structure [23]. In order to overcome these obstacles, the LDS concepts must be implemented in a structured manner.
3.5 Criterion 5: Methods and tools to evaluate the success of implementation
During the implementation of a lean development system, it is important to detect problems immediately. Even in realized pilot projects, it is necessary to determine the success of the measures applied before the start of the rollout. Therefore, key performance indicators, an audit system, or stage models are necessary to evaluate the implementation. The evaluation methods should be included in the LDS concept.
By formulating the five criteria necessary for LDS-conception it is possible to evaluate the LDS concepts of the literature review and their practical application in enterprises.

4 LITERATURE REVIEW OF LDS-CONCEPTS
In this section, several LDS concepts are presented and evaluated based on the previously described criteria. The selection of concepts was based on a literature review. Due to the fast-paced dynamic of this field and the multitude of concepts available, this paper has selected only the most common and distinctive LDS concepts to review.

4.1 Lean Product Development System (LPDS) according to Morgan and Liker
The Lean Product Development System (LPDS), according to Morgan and Liker [4], describes product development at the Toyota Motor Corporation. The investigative results have been summarized in thirteen principles and are categorized according to four subsystems: Process, Skilled People, Tools and Technology.

The Process subsystem comprises all activities and operations accumulating in the product development from product planning to the start of production. The value stream ranges from customer requirements and product ideas to the final draft of the product which is handed over to the production. The subsystem comprises the following principles:

- Establish customer-defined value to separate value-added activity from waste
- Front-load the product development process while there is maximum design space to explore alternative solutions thoroughly
- Create a leveled product development process flow
- Utilize rigorous standardization to reduce variation, and create flexibility and predictable outcomes

The Skilled People subsystem is comprised of the employees involved in product development. There are six principles in this subsystem:

- Develop a chief engineer system to integrate development from start to finish
- Organize to balance functional expertise and cross-functional integration
- Develop towering technical competence in all engineers
- Fully integrate suppliers into the product development system
- Build in learning and continuous improvement
- Build a culture to support excellence and relentless improvement

The third subsystem includes all Tools and Technologies used during product development. This subsystem supports employees with the completion of their tasks. The principles are:

- Adapt technology to fit your people and processes
- Align your organization through simple, visual communication
- Use powerful tools for standardization and organizational learning

Apart from the principles, the LPDS provides hints about the integration of subsystems and explains approaches to culture change.

The LPDS according to Morgan and Liker, represents the currently most widely recognized approach to the systematization of LDS. The concept comprises essential contents by [24-30].

The principles within the LPDS are conclusive, structured and explained in subsystems like the Toyota system that LPDS is based on. The LPDS principles share similarities with common principles of Lean Production Systems. For example, the principle of a continuous improvement process and standardization can also be found in the LPDS. Therefore, criteria 1 and 2 are met by the concept because of the integrated perspective which covers the entire product development process and the supporting processes. Furthermore, it offers structured references about the implementation (criterion 4). However, criteria 3 and 5 are only partly fulfilled. The concept does not provide any standardized catalogues of methods to facilitate selection and customization. There is also a lack of methods and tools for evaluating the implementation and success of the measures.

4.2 Lean Product and Process Development according to Ward
In his book, “Lean Product and Process Development,” Ward presents a further concept for a Lean Development System [7]. First, he defines how value and performance should be interpreted and measured in the framework of product development and which types of waste occur in product development. Then the LDS “as practiced by Toyota and its suppliers” is summarized to five principles. These principles are:

- Value focus: Focus on creating knowledge and hardware for consistently profitable
- Embody this focus in entrepreneur system designers (ESDs)
- Support ESD with set-based concurrent engineering (SBCE)
- Support SBCE with cadence flow and pull project management
- Support flow and pull management with teams of responsible experts

Additionally, an overview of the success factors of a sustainable implementation of the LDS is given.

The approach focuses on the contents applied by Toyota. Some of the contents described in [4] are taken into consideration and explained in detail. It should be mentioned that the work offers approaches for the evaluation of waste, value and performance. Ward has defined principles, fulfilling criterion 1. The field of application is not described in depth; therefore criterion 2 cannot be evaluated. Criterion 3 is not fulfilled. There are several hints for the implementation (criterion 4) and the evaluation for the success (criterion 5), but not an integrated concept.
4.3 Lean Innovation according to Schuh, Lenders et al.

The term Lean Innovation according to [31-32] stands for a systematic application of lean thinking in innovation management and product development. Lean Innovation should help to avoid waste in the processes of product development and direct all processes and contents towards the customer. In order to reach these goals several principles have been conceived and developed. The Lean Innovation principles can be categorized as: “position unambiguously”, “structure early”, “synchronize easily” and “adapt securely”. The principles according to [32] are described below.

To ensure competitiveness and to support the selected business strategy, continuous control and adaptation of the planned products and product range is necessary. With the help of the “position unambiguously” principle these goals shall be reached. The “position unambiguously” principle includes the following sub principles:
- Strategic positioning
- Evident hierarchization
- Roadmapping

An important aspect of product development is the control of the complexity of different projects and activities by structuring early. There are three sub principles integrated in the “structure early” principle.
- Control of the solution space
- Design of the product architecture
- Assortment optimization

A continuous and consistent synchronization of all activities during product development is necessary to achieve stable processes with a maximum use of the projects’ internal and external synergies. The principle “synchronize easily” includes three sub principles.
- Pulsing
- Consistency of information
- Optimization of the value stream

The “adapt securely” principle calls for the ability to adapt products to changing requirements during the entire product life cycle and includes:
- Continuous improvement
- Release-engineering
- Innovation controlling

The fundamentals and principles are described in many different publications with different focuses and degrees of detail. These variations make the application of lean development principles in product development and their support processes difficult. Other aspects, mentioned in [4] or [7], e.g. supplier integration, are only marginally considered. Therefore it is not clear which principles should be applied for supplier integration. Furthermore, the presented principles are often not true principles but rather tasks or methods. For example, the principles “innovation controlling” and “assortment optimization” represent tasks. The fact that the principles and methods are constantly evolving makes their application problematic. Due to these facts, the criterion 1 is fulfilled only partially. In the course of lean product development a basic and comprehensive modification of structure and content of the process has taken place. The focus of this approach is on the product planning, and later steps in the process are neglected (criterion 2). A methodical selection and configuration of the methods is not published yet (criterion 3). The presented approach offers a rough concept for the implementation and measures the degree of implementation through a maturity level scale (criteria 4 and 5).

4.4 Lean Software Development according to Poppendieck and Poppendieck

Software engineering also features approaches based on the lean thinking. One of the most popular approaches is the Lean Software Development concept according to [33]. This concept is based on seven principles and 22 corresponding tools:
- Eliminate waste
- Amplify learning
- Decide as late as possible
- Deliver as fast as possible
- Empower the team
- Incorporate integrity
- See the whole

In [33], the individual principles and the tools related to software development are presented (criterion 1). However, the origin of the Lean Software Development principles and the determination of the compilation of the 22 tools are not explained. Topics like standardization, continuous improvement, and quality related elements are almost not taken into consideration. Due to the focus on software development, the field of application is not as it is described in criterion 2. Additionally, there is a lack of solutions supporting criteria 3 and 5. The concept does define success factors, which give some indication of implementation methods specific to the size of an enterprise (criterion 4).

4.5 Lean Development according to Balle and Balle

In “Lean Development: A Knowledge System” Balle and Balle describe a system divided into different categories [34]. There are four key factors concerning product development, which are meant to describe the goals of every development process. The factors are:
- Listen to the voice of the customer
- Limit late engineering changes
- Master the flow of drawings and tool elaboration
- Focus on quality and cost in production

In order to reach these goals, and consequently a high customer satisfaction, short time-to-market and low costs in several so-called layers of the process must be realized. Within the practice layer the authors identified the elements of technical careers, pull communication, continuous improvement, and supplier integration from lean manufacturing practices. Furthermore, the organizational layer with the component platform center, and the culture layer comprising the knowledge-based paradigm, is identified. In the process layer the elements of frontloading with “concept with chief engineer”, delaying key decisions with “system design with set-based engineering”, reducing variability by “detailed design with standards”, and using lean principles from production with “prototypes and tools with lean manufacturing” have to be considered.

The concept by Balle and Balle includes aspects of [4] and [7]. However, the contents are structured differently. The “aims of every development process”, which can be seen as principles (criterion 1), and the categorization in different layers, allow a structured detailing of the components. In the same way a rough draft of all phases is carried out. However, the concept does not offer either structured method descriptions (criterion 3), an implementation systematic (criterion 4) or evaluation
methods (criterion 5). The field of application is not described in depth; therefore criterion 2 cannot be evaluated.

4.6 Innovative Lean Development according to Schipper and Swets

Schipper and Swets assume that it is not only the principles of lean manufacturing - like flow, pace, pitch, and elimination of waste - that contribute to a successful product development process [35]. Instead, the principles of LPS are combined with the contents of "structured innovation" to an LDS. Within the integrated concept of Innovative Lean Development the following six principles have been identified:

- Identify and fill user gaps
- Use multiple learning cycles
- Stabilize the development process
- Capture knowledge
- Use rapid prototyping
- Apply LPS-principles, including learning cycles and visual boards

The authors particularly focus on using fast learning cycles to create, implement, and maintain a learning culture. The concept of Innovative Lean Development has an integrated field of application since it includes the whole process, from the product planning as the starting point of product development up to the ramp-up (criterion 2). Furthermore it discusses aspects of cultural change. In contrast, there are only a few contents linked to methods and there is a lack of concrete instructions for action (criterion 3). Furthermore, some elements of lean development, and other concepts known as central LDS elements (e.g. supplier integration, frontloading, set-based-engineering), are not described in detail. It is not clear whether the six principles serve to support the cultural change, or if the product development process (support processes, activities, and methods) should be designed in accordance with the principles in order to make it more innovative and lean (criterion 1). Neither the implementation, nor the evaluation of the implementation success is described in depth (criterion 4 and 5).

4.7 Lean Development based on Lean Production

Apart from the highlighted concepts, there are approaches for the systematization of lean development in [36-38] based on the principles for lean production defined by Womack and Jones [22]. The principles are:

- Value
- Value Stream
- Flow
- Pull
- Perfection

These approaches emphasize different points (e.g. the implementation) and offer a systematic but superficial transfer of the principles. The approaches focus on the transfer of the principles without providing detail about the methods. Therefore in most cases, only criterion 1 is fulfilled. The additional criteria are neglected in the concepts.

4.8 Lean Development based on PTC

Another approach to the design of product development processes is shown in [39]. This concept is composed of six initiatives:

- "Frontloading": Shifting important decisions into the earliest possible phase of product development.
- Visual planning and completion
- Standardization of working-processes
- Systematic gathering and reuse of knowledge
- Partnership with manufacturing service providers and suppliers
- Efficient coordination of development processes and results

These six initiatives can be used as principles for lean development systems, fulfilling criterion 1. Frontloading, which also has been published in other papers [40], is mentioned explicitly. The field of application is not described (criterion 2). In addition, the approach does not offer any structured derivation of methods (criterion 3), proceedings for implementation (criterion 4), or evaluation methods (criterion 5).

4.9 Lean Product and Process Development (LeanPPD)

The research project Lean Product and Process Development (LeanPPD) considers four main blocks [41]:

- Lean self-assessment tool
- Product development value mapping tool
- Knowledge-based engineering
- Set-based Lean design tool

The lean self-assessment tool provides key performance indicators, helps to identify the starting point of the company, and helps to measure the success of the transformation to lean development. The value mapping tool helps to display the processes according to their value creation. The knowledge provision was identified in the project as main feature in the lean development. All decisions in a development project should be based on proven knowledge and experience. The knowledge based engineering tool will support knowledge acquisition and re-use of previous projects to support the application of lean development. The set-based design tool helps to identify the lean product design which can be manufactured in a LPS. [41]

The project is not completed as of the present, therefore some of the tools were only drafted and the exact function has not been published yet. Up to now there have been no principles defined (criterion 1). The field of application is widespread, but not described in detail (criterion 2); and the selection of tools and methods was not described (criterion 3). The implementation process (criterion 4) has not yet been published. The lean self-assessment tool can be used to measure the success of the implementation (criterion 5).

4.10 Lean Development approaches in enterprises

The LDS concepts described and evaluated in sections 4.1 - 4.9 are based on theoretical approaches. At the same time some enterprises have developed their own approaches for increased efficiency in product development by implementing rules and standards based on lean principles. Two representative innovative enterprise approaches are presented here. According to [42], Robert Bosch GmbH (rank 5) and Siemens AG (rank 3) are innovative enterprises with a high number of patent applications in Europe in 2 009.

Robert Bosch GmbH operates with the Bosch Product Engineering System (BES) [43], which focuses on product development and its environment. Product development is optimized through the use of best practice-processes and qualified employees. Principles of the BES according to [44] are:

- Market and customer orientation
In contrast, according to [45] the Siemens AG has defined different fields of action within particular categories of lean development. These categories are:

- Leadership-principles and strategy
- Product lifecycle management
- Supply chain management
- Customer relationship management
- Project and performance management
- Human resources and continuous improvement

For example, the category of product lifecycle management consists of frontloading, reliable operational processes and synchronous operations. There are different methods specifically designed for every field of action.

Both of these approaches are enterprise-specific (criterion 3) and not a universally valid concept. There are lean principles defined as mentioned in criterion 1. The field of application is the whole product development process and parts of the support processes (criterion 2). The implementation (criterion 4), and the evaluation of the success (criterion 5) are not described.

4.11 Summary of evaluations

Criterion 1 “Development of specific LDS-principles”: The variety of principles within the LDS concepts shows that several different influencing factors have been taken into consideration. Some approaches, especially those of [4] and [7] are based on principles of Toyota. Other authors suggest that lean development represents the application of lean production principles. In addition to these two directions, mixed concepts have been developed taking “implied principles” into consideration, which have proven themselves valuable in product development practices in the past. In general, nearly every concept has LDS principles, despite the variety in content and detail.

Criterion 2 “Integrated perspective and field of application”: In general, the principles defined are described in a way that allows a standardized application and hence suggest coverage of all product development phases as well as their support processes. However, most concepts lack specific descriptions of the field of application for the support processes, making them difficult to evaluate.

Criterion 3 “Enterprise-specific selection and configuration of methods and tools”: LDS concepts partly offer a specification of the particular LDS principles and sporadically describe methods. A methodology for a systematic selection of methods, an integrated description of activities, and corresponding methods and tools do not exist.

Criterion 4 “Specification of an implementation process”: Only in some concepts, e.g. [4], does a comprehensive process exist for the implementation of a Lean Development System that takes into account planning, corporate culture, leadership, knowledge, and organizational structures.

Criterion 5 “Methods and tools to measure the success of implementation”: Key performance indicators, an audit system, or stage models for the evaluation of the implementation of LDS are mentioned in some approaches. But even there, a detailed description is missing.

As shown in table 1 no approach fulfills all requirements identified as necessary for a lean development concept. Therefore further research has to be carried out.

![Diagram showing evaluation of LDS concepts]

Table 1: Summary of the evaluation of the LDS concepts

5 SUMMARY

In order to stay competitive, enterprises are forced to minimize waste in their processes and focus on value creation with integrated thinking. The implementation of lean thinking in all business processes is a promising approach. However, there are a multitude of concepts for the implementation of lean principles in product development. This paper provides a literature review and an evaluation of some of the common approaches to lean development.

There is a strong disagreement regarding the contents and proceedings within the scope of Lean Development Systems. Although some concepts provide approaches for the configuration of a LDS, there are very few scientific studies published, which can be used to draw conclusions concerning the individual principles and method as well as their conduciveness to success. Especially the criteria concerning the implementation of a LPS and the measurement of the success lack suitable solutions.

6 REFERENCES

Robustness of Inventive Design Solutions - Transferring the Robust Design Focus from Production Process to Development Process

F. Wuttke, M. Bohn, N. A. Suyam-Welakwe, A. Bohn

1 Institute for Information Management in Engineering, Karlsruhe Institute of Technology, Germany
2 Daimler AG Research & Development, Benzstrasse, 71059 Sindelfingen, Germany
3 Group Research & Advanced Engineering, Daimler AG, Wilhelm-Runge-Str. 11, 89081 Ulm, Germany
4 Chair of Product Development and Machine Elements, Technische Universität Darmstadt, Germany
{Fabian.Wuttke, Martin.Bohn, Nick-Ange.Suyam_Welakwe}@daimler.com; bohn@pmd.tu-darmstadt.de

Abstract
Conventional Robust Design methodologies consider deviations of product parameters during production and use. While it is acknowledged to integrate Robust Design methodologies as early as possible into the product development, most approaches neglect the systematic consideration of uncertainty due to the development process itself. This leads to a major drawback for inventive design solutions, especially within highly standardized and rapid development processes like car development. The proposed solution overcomes the major obstacles for the application of Robust Design in early design stages by identifying all relevant types of uncertainty. Furthermore, methods for the evaluation of robustness against uncertainty are shown. Finally, an optimization process is proposed and verified using an automotive example.

Keywords:
Early Design Stage, Robust Design, Inventive Design

1 INTRODUCTION
Current car development is focused on building modularized cars assembled by a certain amount of modules. Ideally, companies define one technical standard solution for each module, the standard module. This standard module has to be capable of being integrated into the highly standardized and rapid development of every new car project of the whole company. The development of the standard module solution needs to take place in the first design stages. The influence of uncertainties of the development process itself dominates decisions in early design stages. Therefore, one major issue of module development is the question which possible solution is best capable of handling deviations from the car development process. Currently, no method is established in module based development to systematically handle uncertainty. Therefore, the decision about the standard module is based on estimations instead of calculations. Consequently, conventional and approved solutions are usually favored in industries with highly standardized development processes. This leads to a major disadvantage for unconventional solutions possibly found by inventive design methodologies.

The aim of this paper is closing the gap between inventive design solutions and their integration into development processes based on the module paradigm.

1.1 Inventive Design Solutions
At the beginning of a module development, developers strive to obtain the maximum number of solutions possible. Therefore, both conventional and inventive development methods like TRIZ [1] are usually taken into account. Recent research activities in the field of Computer-Aided Innovation (CAI) intensively investigated the question of how to systematically come to innovative or inventive products by methods or tools [2]. The increasing applicability of methods like TRIZ enables developers of automotive modules to reach a higher level of inventiveness during the very first stages of car development. However, the demanding functional requirements of the automotive industry and tough timing combined with the high sensitivity for uncertainties during the development process lead to problems of acceptance of such tools. As a consequence, designers tend to fall back to proven and tested conventional modules. It can be concluded that innovative or inventive solutions based on CAI methods only will be considered in early stages of module development, if they sufficiently account for uncertainties that arise during later stages of the car development process.

1.2 Challenges of early design stage
The automotive development process is structured into the phases ‘strategic development’, ‘preliminary development’ and ‘mass-production development’. Due to the increasing variety of future product portfolios, the need to use more standardized modules is inevitable. This means that a vehicle is assembled by a certain amount of modules. The module paradigm aims at the comprehensive standardization of every module within an OEM. Thus, the standard modules are adapted specifically for each car project. This eventually leads to reduced costs because of scaling effects.

The development of standard modules requires very early evidence about the fulfillment of functional requirements in very different car concepts. Standards for modules are developed within the strategic development phase. Later development stages focus on the integration of those modules into the assembled product.

Uncertainties are present during every stage of product development including production and use of the final product, see Figure 1. Typically, large uncertainties are present at the beginning of the car development process and decrease over the timetable reaching their minimum after the ramp up of the mass production. Once a product...
reaches the area of unreliability in use, uncertainty in terms of probability of failure increases again.

As shown in Figure 1, the decision about the standard module has to be done very early and influenced by the highest level of uncertainty possible. The neglect of those uncertainties within the development of a standard module typically leads to high changing costs during the car integration phases. Furthermore, the consolidation of different solutions for standard modules likely results in proved solutions because designers have no method to handle the uncertainties influencing innovative solutions during car development.

Therefore, a new approach is needed. This approach focuses on identifying the most robust concept for standard modules taking into account the high amount of uncertainty during the development process.

2 ROBUSTNESS IN EARLY DESIGN STAGES

The targets mentioned above can be translated into an extended application of robust design methodologies. A brief summary of the progress of robust design approach from design paradigm to probabilistic based simulation method builds the fundamentals for introducing the new approach of early optimization of robustness.

2.1 Design to Robustness

Taguchi's idea of minimizing processes and use deviations of a product during its development caused a fundamental change in quality improvement methods [3]. Based on the international acknowledgment of Taguchi's paradigm, further works introduced Design for Six Sigma (DFSS) [4-5]. DFSS methods aim at systematically reducing the level of failure probability to a level of 4.5 $\sigma$ or higher during the development of a product. Hence, the ideal of a zero defect production is sought starting in the early design stages.

Within this framework, description methods like the P-Diagram were developed. A P-Diagram interprets a product as an input/output system with signal, noise and control factors influencing the product's output. In early design stages, P-Diagrams are appreciated as a demonstrative way to gather noise and its impact on the product [6].

Nee [7] interpreted a main potential of the Taguchi approach in identifying significant design factors prior to optimization. Gu [8] later combined the robust design paradigm with Axiomatic Design. This extension of engineering design transparently enables the assessment of an analyzed product design in comparison to its robust and ideal design. An ideal design is characterized by independence between uncertainties and functional features of the product. A robust design represents minimal influence of uncertainties on functional features.

Product designs can be changed easiest during the first stages of development. Because decisions about conceptual designs typically have to be done under a high level of uncertainty, Gheorghe [9] developed methods to aid concept decisions with the problem of high fuzziness during early design stages against the performance and profit of the final product.

Uncertainties are often equated with the complexity of early concept decisions. Because interactions and interdependencies between product features and noise factors are initially unknown, works like [10] seek to analyze concept interdependencies including all areas of product development (e.g. product functions, production concepts).

All in all, it can be concluded that Robust Design methodologies aim at characterizing and identifying the dependencies of different factors determining product behavior during early design stages. Based upon statistical analyzes, developers are able to identify deficits in order to design robust products.

2.2 Simulation based optimization to Robustness

Due to significantly decreasing costs for computer calculation power over the past years, robust design approaches extended from design paradigms to simulation methods. Latest developments enhance the robustness analysis by adding optimization algorithms. In mathematics, Robust Optimization (RO) covers the optimization of systems under uncertainty [11].

The probabilistic variant of the RO approach typically focuses on production issues by optimizing designs (e.g. topology optimization) while considering product scattering during production and use in parallel [12]. The goal is to investigate and optimize designs in order to maximize reliability of products to the area of six sigma.

Unlike analytic RO in mathematics, the optimization process of probabilistic RO is dominated by stochastic artificial life science algorithms. Based on examples like evolutionary or genetic optimization, research works like Roy [13] demonstrated powerful extension possibilities by integrating uncertainty not only into design variables but also into decision criteria. Applications like Zhang [14] showed that even the tolerance specification of design parameters can be optimized simultaneously by using the probabilistic RO approach.

2.3 Summary

All presented forms and enhancements of the robust design approach have one aspect in common. A product is designed and optimized to be robust against uncertainties of the production and use processes. In the future car development aims to have cars assembled by standard modules exclusively. Standard modules need to be developed in very early design stages. Later in the car development process, these standard modules must be capable to handle the uncertainties caused by their integration into distinct vehicles. Currently, no method is available to handle uncertainties during car integration. On the one hand, this results in suboptimal performance of conventional standard modules. On the other hand, developers tend to avoid inventive solutions for standard modules because uncertainties for novel solutions are even larger. As future car development will be heavily affected by the module paradigm, an approach to handle these uncertainties is inevitable.
3 ROBUSTNESS OPTIMIZATION OF MODULES

The new approach evolves from the probabilistic RO paradigm. The system is described by a set of $m$ design parameters

$$d = [d_{1,1}, d_{1,2}, d_{2,1}, d_{2,2}, \ldots, d_{m,1}, d_{m,n}] .$$

The design space $d$ is described with lower bounds $d_{l,i}$ and upper bounds $d_{u,i}$ of each design parameter $d_i$. These parameters represent the design bounds for each respective standard module. In addition, $n$ uncertainty parameters

$$r = [r_1, r_2, \ldots, r_n] \quad (2)$$

are defined. $r$ contains all possible deviations during the development that influence the behavior of the simulated system. The approach works in a dual looped process, see Figure 2.

First, the robustness value $\delta_r(d,r)$ of a starting set of design parameters is determined within the robustness loop. Therefore, deviations $r$ have to be added to the parameters of an initial design parameter combination. Subsequently, system simulation with different combinations of $r$ is done based on appropriate stochastic sampling methods like Latin Hypercube Sampling. The simulation results in terms of functional requirements are then analyzed statistically. This finally results in an overall robustness value $\delta_r(d,r)$, based upon appropriate weighting of functional requirements.

Second, optimization seeks to find the best combination of design parameters for the investigated solution. In contrast to conventional probabilistic RO, it is proposed to reduce the optimization criteria to $\delta_r(d,r)$. Hence, the overall target $T$ of the process can simply be identified as

$$T(d) = \min(\delta_r(d,r)) . \quad (3)$$

Consequently, increasing robustness of the system results in decreasing values for $\delta_r$. This must be considered for the statistical analysis of functional requirements within the robustness loop (e.g. noise to signal ratio instead of signal to noise ratio).

Related works mainly aim at the application of probabilistic RO focusing on production and use of a product. They do not address uncertainties that modules face during their integration into the development process of different assemblies, e.g. cars. Furthermore, no evaluation methods for the robustness of those modules can be found that deal with those uncertainties. Hence, the major research effort of this paper is put on the investigation of the robustness loop.

### Specification of uncertainty

The special needs of early design stages are addressed by defining uncertainty or deviation values $r$. Those uncertainties can be relatively large as one module concept contains every future car targeted by the module strategy. These uncertainties can be broken down into four types illustrated in Figure 3 and explained below.

- **Uncertainties due to the car model validation process**
  One of the most dominant factors for the development of a car is styling. The design of the car model is revaluated repeatedly during the car development process. This builds a major uncertainty for automotive modules because styling decisions generally don’t consider potential drawbacks for the modules. Since the automotive industry established standard processes for new car projects, successor car projects and model upgrading projects, many potential uncertainties can be derived from experiences made. Those uncertainties at the interface between module and the rest of the assembled car are quantified and applied to the investigated solutions for the standard module.

- **Uncertainties due to early stage data availability**
  Frequently, data concerning geometry and function of parts affecting the system behavior of the module are not available at early stages. Therefore, this data has to be estimated based on similar or predecessor cars. Due to the fact that estimations quantify uncertainties, all possible variations have to be transferred into the formulation of deviation $r$.

- **Uncertainties due to used CAD methods**
  Assembled automotive parts are typically mapped to common computer aided design (CAD) environments in the course of car development. However, the complexity and heterogeneity of most automotive parts disallows exact evidence about data concerning geometry and function even at the end of mass production development. Hence, every single data evolving from CAD processes is uncertain. The magnitude of deviations is dependent on the complexity and granularity of the considered data. This increases the uncertainties due to early stage data availability because every assumption based on predecessor cars cannot be exact perse.

- **Uncertainties due to newly developed modules**
  Simulation models of proven concepts for modules standards are typically verified holistically for different car types. Thus, those simulation models can be adapted to all targeted cars in course of the standard module development. On the other hand, the simulation models of
solutions identified by inventive design approaches first have to be built. Novel solutions for modules often require more or other available space than predecessor solutions and do not fit into current cars. Therefore, most of those simulation models can only be validated and optimized decoupled from the context of the entire vehicle and interface, respectively. The transformation of validation conclusions from isolated module testing to the context of the entire vehicle results in uncertainties of modeling parameters of the standard module. In conclusion, this type of uncertainty mainly applies to inventive or innovative solutions.

**Evaluation of robustness**

Basically, all determination methods for robustness values refer to values of the output probability density function. Besides standardized formulations of the four statistic moments - mean value \( \mu \), variance \( \sigma^2 \), skewness \( \nu \) and kurtosis \( \gamma \) - robustness can also be determined by the coefficient of variation \( c_v \) or failure probability \( p \). Figure 4 illustrates how to gather the robustness value of a very simple system. The system is characterized by one functional requirement, the output \( o_1 \) with a related specification border. Furthermore, the uncertainty of early system evaluation is described by the scattering input parameter \( x_1 \).

For this system, robustness is built as a weighted sum of failure probability \( p \) and the coefficient of variation \( c_v \):

\[
\delta_R = a \cdot p + b \cdot c_v \tag{3}
\]

Both \( a \) and \( b \) represent weighting of the considered statistical values \( p, c_v \). Thus, robustness evaluation can focus on different paradigms. According to Figure 5, emphasis on \( c_v \) on the left hand side leads to independency between output and input. Dominant consideration of \( p \) on the right hand side results in increasing reliability of functional requirements fulfillment without taking actual robustness into account.

![Figure 5: Possible focus of robustness evaluation](image)

It has been shown that early design stage evaluation experiences large uncertainties. Therefore, striving for an axiomatic design seems to be the main goal. On the other hand, axiomatic designs represent theoretical ideal designs. They won’t appear in reality. Furthermore, designs with a very low \( c_v \) don’t regard the distance to the specification borderline of functional requirements. As input deviations are still very large due to uncertainty, the focus on the minimization of \( c_v \) easily leads to a high failure probability.

Simulations based on probabilistic RO approaches tend to focus on minimizing the probability of failure in order to reach a design for six sigma. Since the large uncertainties of module integration usually preclude reaching a level of six sigma design, the new approach proposes a paradigm change.

Reaching a certain level of failure probability has to be substituted by minimizing the failure probability. Thus, failure probability is allowed in the percent range during early design stages.

All in all, overall robustness \( \delta_R \) has to be evaluated on a balanced compromise of all statistical values considered. In practice, this compromise leads to an extension of functional requirements.

**4 CASE STUDY**

One fundamental automotive demand is to bring the latest styling trends and technologies into each product. This results in a high probability of change for early car designs. Hence, modules, which are directly dependent on the styling of the car, experience even larger uncertainties during the car integration phase. Therefore, the following case study focuses on the automatic tailgate module.

The environment of this module is characterized by a high degree of customer visibility, see Figure 6. As a result, interface modifications regarding car styling become very likely in early development stages. Those modifications have a direct effect on the performance of the module systems. Automatic tailgate concepts contain elements that apply forces from the Body-in-White (BiW) to the tailgate. This system has to work in every part of the world with high ambitions concerning comfort and safety. Therefore, high efforts have to be put into a maximum robust behavior.

![Figure 6: Simplified model of an automatic tailgate](image)
external design parameters like joints see Figure 6. The location of the joints plays a big role regarding kinematic behavior. Usually, the possible locations for joints at the Body-in-White are illustrated as half-opened cylinders. The geometrical characteristics of the cylinders strongly correlate with the car type. Hence, these design parameters have to be defined relatively because the standard modules have to fit into all targeted car types of a company.

**Specification of uncertainty**

According to chapter 2.2, some of the most relevant uncertainties for automatic tailgates are shown in table 1.

<table>
<thead>
<tr>
<th>Car model validation process</th>
<th>Early stage data availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>- location of hinge axis</td>
<td>- poor / no data availability of targeted cars</td>
</tr>
<tr>
<td>- design of rear lights</td>
<td>- Estimations based on predecessor cars</td>
</tr>
<tr>
<td>- design of tailgate</td>
<td>- novel future car concepts</td>
</tr>
<tr>
<td>(mass and center of gravity of tailgate)</td>
<td>- materials of tailgate</td>
</tr>
<tr>
<td>- design of BiW (joints, package restrictions)</td>
<td>- features included in tailgate</td>
</tr>
<tr>
<td>- design tailgate (joints, package restrictions)</td>
<td>(e.g. rear camera)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CAD method</th>
<th>Newly developed module solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- every system parameter</td>
<td>- unexpected lifetime behavior</td>
</tr>
<tr>
<td></td>
<td>- transformation from laboratory to reality</td>
</tr>
</tbody>
</table>

**Evaluation of robustness**

Unlike most other modules of a car, the development process of automatic tailgates is not only based on the fulfillment of strict functional requirements regarding performance (e.g. opening times) or applied loads. Rather, the behavior of the entire system during use is the fundamental development principle. Thus, the dimensions of robustness evaluation are extended in order to determine robustness considering kinematic parameters like opening angles, see Figure 7. Consequently, not only statistical values of one output probability density function are examined but also the behavior of an output plotted against the kinematic parameters [15].

![Figure 7: Dimension extension of robustness evaluation](image)

Possible solutions for automatic tailgates have to fulfill about 80 functional requirements that can be checked as outputs in 50 use cases. In order to handle this complexity, a new method has been established to evaluate multi use-case robustness. Figure 8 shows the stepwise determination of the overall robustness value \( \delta_R \) starting with the identification of statistical values of every output of each use-case. This step represents the main challenge of this method. First, the different statistical values (e.g. signal-to-noise-ratio, failure probability) have to be transformed into a comparable magnitude. Second, the specific values for behavior robustness have to be weighted. This means that all functional requirements have to be extended by weighted statistical values. Subsequently, the aggregated output robustness values result in a use-case robustness. Finally, the robustness values of differently weighted use-cases determine the overall robustness \( \delta_R \).

![Figure 8: Multi-use-case robustness evaluation](image)
Optimization of robustness

In contrast to probabilistic RO, the only objective of the presented approach is to minimize $\delta_R$ as all included statistical values in $\delta_R$ decrease with growing robustness. Exemplarily, two concepts have been optimized taking into account 37 statistical values of 24 considered functional requirements. The optimization has been done by an evolutionary algorithm. The robustness values of each iteration step have been determined by 30 Latin Hypercube Samples.

Figure 9 shows the results of the robustness optimization of two competing concepts for automatic tailgates in early design stage.

The solution of concept B is a one-sided electromechanical drive operating directly at the tailgate’s hinge. Concept D describes a hydraulic system applying its forces by a two-sided actuator.

Concept D started with a 2.4-times better robustness value in comparison to concept B. The optimization progress of concept D shows, that its robustness value has already almost reached its best value. Concept B in opposite showed a major capability to decrease the robustness value. Finally, concept B is able to perform slightly better than concept D.

5 CONCLUSION

The presented approach focuses on the handling of early design stage uncertainties of automotive modules. It is described that the uncertainties of production and use of modules are not the focus of early design stage decisions. Rather, the uncertainties due to the integration of modules into the development process of cars are dominant. Hence, 4 types of early design stage uncertainty are identified. Those uncertainties can only be handled if their effect on the fulfillment of functional requirements is made transparent. Therefore, basic investigations concerning robustness evaluation in early design stage are demonstrated. In contrast to probabilistic RO, designers have to accept failures in percent range. Furthermore, the coefficient of variation has to be taken into account to identify actual robustness. Consequently, this leads to an extension of functional requirements. The considered statistical values of each functional requirement and their weighting have to be identified. This eventually results in an overall robustness value. Finally a conceptual optimization algorithm derived from probabilistic RO is shown. It is proposed to exclusively minimize the overall robustness value.

Following, the implementation of this paper’s approach demonstrates the high unique effort which is necessary to extend the functional requirements of a showcase module automatic tailgate. Uncertainties are identified supported by the introduced structure of 4 types. Hence, robustness can be evaluated systematically. The execution of an evolutionary optimization algorithm illustrated that the minimization of the overall robustness value of two competing solutions for automatic tailgates is feasible for real problems.

In practice, the new approach requires a high unique effort for the implementation. E.g., the compromise about the extension of functional requirements has to be found on a high commitment level. Therefore, all responsible participants of module development are forced to analyze the module very early and to identify the really important functional requirements. This leads to frontloading and a higher level of transparency of the development process. Whereas robustness can be given on a technical basis, humans still have to use and execute the approach. This leads to questions concerning process robustness. Therefore, further research focuses on the human role and influence of the approach.

Additionally, the approach enables developers to take inventive and innovative solutions for automotive modules into account. However, this can result in a large solution space. In order to handle this complexity, methods and tools for the automation of the process have to be found. Combined with the premise of full embedding capability into car development, the process automation seeks to simplify and standardize further module development.

6 REFERENCES


Approaches to the Design of Micro Mechanical Systems

H. N. Hansen
Department of Mechanical Engineering, Technical University of Denmark,
Produktionstorvet, DK-2800 Kgs. Lyngby, Denmark
hnha@mek.dtu.dk

Abstract
The design and manufacture of micro mechanical systems is characterised by a multi-disciplinary environment. The use of established design approaches such as axiomatic design, functional analysis and biomimetic design can be used in the conceptual design phases. The concept of micro engineering is introduced to identify important characteristics of the design and secure a coupling to manufacturing possibilities at an early stage. The paper discusses the different design approaches applied to micro mechanical systems and illustrates some issues based on specific cases.

Keywords:
Design, Micro manufacturing, Micro mechanical system

1 INTRODUCTION
Innovation within the field of micro and nano technology is to a great extent characterized by cross-disciplinary factors. The traditional disciplines like physics, biology, medicine and engineering are united in a common development process that can only take place in the presence of multi-disciplinary competences [1]. This requires a high degree of scientific specialization both from the point of view of the product designers and the production specialists. This fact makes the principle of concurrent engineering quite hard to implement. This paper describes some of the challenges related to the design of micro mechanical systems. Selected case examples will be used to illustrate the complexity of this area.

2 MICRO MECHANICAL SYSTEMS
A micro mechanical system is characterized by small dimensions, either of the system/component itself (one or more critical dimensions) or of functional features or structures on the system/component [2]. Generally, two categories of micro components are identified:
• Components with at least two critical dimensions in the sub-mm range, thus implying that the parts themselves are small and usually with a very low mass, e.g. parts for hearing aids.
• Relatively large components with functional features in the µm range, e.g. DVDs.

This viewpoint itself makes a clear-cut definition hard, since features by nature can be one or more orders of magnitude smaller than the dimensions of the products. From a geometrical point of view micro products can be organised into three groups [2]:
• Two-dimensional structures (2D), such as optical gratings.
• 2D-structures with a third dimension (2½D), for example fluid sensors (the structure of the channel system itself is two-dimensional, but since the channels have a finite depth they can be characterised as 2½D).
• Real three-dimensional structures (3D), for example components for hearing aids.

The geometry affects the possible manufacturing methods and the associated production support in terms of handling, assembly and metrology.

Another important characteristic of micro products is the integration of functions, integration of different functioning principles (physical, chemical, biological etc.) and integration of intelligence into products in terms of information processing and control (sensors and actuators).

The integration of different length scales into the same component/product is discussed in [3]. Here length scale integration is defined as the integration into a single product of functional features of different characteristic scales, e.g. nano structured surfaces on a micro fluidic device. Two different types of "solutions" to this problem are identified in [3]: assembly related solutions and multi-scale machining solutions. Solutions are linked to the absolute dimensions in question. If the critical dimensions are in the sub-mm range (as defined previously in this section) typically assembly solutions are identified to deal with length scale integration. If the absolute dimensions become so small that physical manipulation is difficult (due to force interaction, loss of visual coordination etc.) standard solutions are not available and other principles need to be employed (e.g. self assembly etc.). Assembly can to some extent be avoided if all manufacturing processes can be performed on the same specimen. Solutions employing a single process on multiple scales are seen (e.g. micro electrical discharge machining) as well as multiple processes on multiple scales (e.g. application of two processes in sequence). In both cases process resolution, alignment and referencing errors as well as process realization become key challenges.
3 DEVELOPMENT OF MICRO MECHANICAL SYSTEMS

3.1 General considerations

The process of coming from the first idea to an industrially manufactured product is long and must eventually include engineering skills. The ideas for functionalities to be obtained via micro products stem from many scientific areas. This will challenge the ability of engineers to create functional products and to choose between the many possible solutions. Some common mistakes during concept development encompass the following points:

- Consideration of few alternatives and failure to consider other concepts employed.
- Ineffective integration of promising partial solutions.
- Failure to consider entire categories of solutions.

These points are valid for macro products but become particularly relevant in micro product development since the designer often is limited by a specific scientific background. The integration of semiconductor technologies with conventional manufacturing technologies and material science represents the biggest challenge to micro product development but also the most promising trend in terms of innovation and value creation.

The generic product development process (figure 1) can be adapted to micro products with some modifications [4-5]. Early in the process special considerations have to be given to material choice and the subsequent manufacturing technology. If the principle structure is based on a fixed combination of materials and related production technologies (e.g. silicon and etching technologies) the subsequent development becomes an optimization of this combination [6]. In this case no real possibilities exist for changing materials or processes, thereby influencing product performance and cost. A premature choice of materials and processes also limits the possible geometries to be used for the single parts/components (2D-2½D-3D). Therefore the designer and product developer have to possess knowledge about alternative materials and production technologies to be able to develop the most optimal product for a given situation. A way to reduce development time and cost is through a systematic design approach to reach a decoupled design solution.

![Figure 1: Generic product development process [5].](image)

3.2 Systematic design approaches applied to micro systems

In any micro manufacturing technology and particularly in silicon micromachining, the manufacturing sequence influences the technical performances and quality of the product. Therefore, several constraints due to incompatibilities of materials, processes and geometries have to be considered while defining a manufacturing sequence [7]. Since each process step influences, in principle, the results of both the previous and the following process steps the process sequence has to be checked for consistency and incompatibilities must be identified. Eventually this will influence the design parameters. According to [8], the possible constraints due to incompatibilities can be grouped into three classes.

First the properties of material or functional elements may be affected by succeeding processes as for instance twisting or destruction of delicate mechanical structures due to thermally induced mechanical stresses or the attack of thin film materials by subsequent etching processes. The second class comprises the possible negative influence of the properties of materials and device geometry processed so far on the quality of succeeding technology steps. Again, examples are insufficient adhesion of adjacent thin film layers or inadequate planarity of layers deposited on top of 3D micro structures. The third kind of constraints concerns the feasibility of generating the intended device geometry using the specified fabrication processes. Even though the references refer to silicon based technologies, in principle the same types of constraints and incompatibilities are seen in micro manufacturing of components in polymers and metals as well.

A way to reduce development time and cost is through a systematic design approach to reach a decoupled design solution using the axiomatic design approach [9]. In fact, depending on the sequence of processes and process steps, a MEMS design can be coupled or decoupled. In many cases designers are not conscious of the coupled nature of their design and thus it becomes difficult to identify the correct changes in the process variables to improve (or even to obtain) the product performances. Often the process variables to fabricate MEMS devices are “randomly optimized”. In [10] the axiomatic design approach is used to design and manufacture a MEMS based device.

A systematic method based on functional analysis as described in [11-12] can also be used. A method based on the analysis of a macro scale device, followed by an analysis of which functions are influenced by the downscaling can be used. This approach has proven to be beneficial in pinpointing problem areas induced by downscaling.

Finally, the use of biomimetic approaches has been reported in micro product design [13]. Biomimetic design uses biological phenomena as analogies to help solve engineering problems. One well-known example of biomimetic design is the development of Velcro after observing that cockleburs attach to clothing and fur. The use of this methodology requires access to biological databases and the competence to interpret and translate analogies into engineering solutions.

4 MICRO ENGINEERING APPROACH

Micro engineering is introduced as a concept and it should be seen as the entire set of actions related to product development and manufacturing of micro products. In this context it becomes clear that a categorization as proposed above is not sufficient to comprise a full definition of the product. Important aspects...
such as geometrical complexity, integration of various materials, functionalities and components as well as requirements concerning mechanical, chemical and electrical performance all should be considered during the process of product development. This phase is strongly influenced by lack of design guidelines and tolerancing rules, and it is complicated by the fact that the tradition of concurrent engineering is far from dominating the development phase in micro engineering. The use of standard construction elements as in traditional mechanical engineering has until now not been adapted intensively although some commercial MEMS CAD tools contain standard elements such as beams and cantilevers. The fundamental issues of size effects when trying to apply macro rules to micro products and components are big e.g. [14]. In consequence, no uniform approach exists in this field, the consequence being that product developers run the risk of being limited to ‘known traditional solutions’ only.

The current research in the micro manufacturing area is focused very much on single manufacturing processes and their interaction with the material being processed. Focus is given to size effects e.g. [2,14]. The establishment of coherent process sequences, i.e. covering all necessary process steps from tooling over replication to assembly processes, is a very important research area. Often it is a quite challenging step and for the industrial realization of micro manufacturing a necessary step. When integrating single processes into coherent process chains and subsequently into production systems issues as material compatibility, relative accuracy, alignment precision, etc. must be considered. The necessary actions related to quality control comprise process validation and verification of tolerances as specified in the design.

Figure 2 illustrates the necessary components in a micro engineering approach identifying the most challenging parts. By experience some of the most restricting elements are the coupling of manufacturing possibilities (and constraints) to a conceptual design. This is where the specificities of micro scale processing are implied onto the design. Possibilities and restrictions of processing are described in many papers e.g. [15-18], but a major challenge is to establish an overview in order to make a qualified decision. So far the systematic design approaches described earlier only lead you a part of that way.

Figure 2: Elements in the concept of micro engineering.

Another challenge for micro product design is the detailed specification in terms of dimensioning and tolerancing. In macro scale engineering this discipline is well established and a long tradition has enabled distributed manufacturing based on a common technical terminology. In micro manufacturing, this is still an emerging area. The support from technical standards is virtually nonexistent at these scales. Furthermore, in a standard manufacturing environment, dimensional metrology is used to ensure the quality of the produced components. If the micro mechanical system is based on assemblies, extremely high demands are set to positioning and alignment accuracies in-between process steps as well as precise parts for subsequent assembly steps. This concept requires detailed knowledge of not only absolute dimensions and geometrical quantities, but also about the uncertainty of measurement, because this is a decisive parameter when dealing with machining capability in general.

In this context the verification of tolerances by means of dimensional and geometrical metrology becomes a key point. The specifications are usually given in terms of maximum deviations from an ideal, nominal dimension/form. The compliance with specifications are described in [20]. Figure 3 illustrates the principle. In order to be able to decide about a specific part, two points need to be fulfilled: a suitable measurement method must be identified to perform the measurement and the corresponding measurement uncertainty needs to be sufficiently small to be able to verify the tolerance. Upon downsizing of the absolute dimensions, usually the ratio of measurement uncertainty to tolerance becomes large, in this way leaving a smaller conformance zone for process variations.

Figure 3: Tolerance verification at micro scale. “U” indicates measurement uncertainty.

5 RE-DESIGN OF SWITCH FOR HEARING AIDS

In hearing aids many micro electro mechanical systems are found. This case is based upon a so-called push button found in many so-called “in-the-ear” solutions (figure 4). The push button can be used to turn on/off the system or change programs in the IC. The maximum diameter is 1.9 mm. As illustrated in figure 4, the current system is based on the scaling down of a traditional mechanical solution: screws, springs and structural elements. The single elements are manufactured using standard down-scaled manufacturing technologies, and the assembly is performed in a semi-automatic way. The chosen design is robust from a performance point-of-view, and experience shows relatively little sensitivity of the single manufacturing tolerances on the subsequent assembly. The turning process (for making the 0.5 mm screw) is close to the limit of state-of-the-art. The main challenge in the entire process chain lies in the assembly operation. Manual labour is required because a fully automated assembly line is too inflexible (typical production volume 100.000 pieces/year).
Applying the proposed methodology to the push button has resulted in a new proposed design and process chain [21]. Based on the list of specifications, an analysis based on functional analysis was performed. In parallel, possible processing scenarios were screened in order to be able to take advantage of new technological possibilities in the design phase. It was decided to opt for a solution based on two-component micro injection moulding. Figure 5 illustrates the principle. A core part consisting of two polymers, of which one has been metallised, secures the electrical conductivity from top to bottom. A flexible dome attached to an outer housing creates the electrical contact. The dimensions mentioned are about 2 times larger than the existing solution. This was chosen because of the challenges related to physical realisation of the tools. Figure 6 illustrates the plastic part after injection moulding of the two polymer materials. The material combination was chosen in such a way that chemical metallisation of the second shot polymer was possible (without any metal being deposited on the first shot polymer). This is a compromise between establishing a strong adhesion between the two polymers and securing a selective metallisation. These two characteristics are acting in opposite directions [21]. The new design consists of 4 main elements compared to the 6 of the original design. With the proposed process chain, complexity in manufacturing was moved from the assembly steps to the injection moulding step. Furthermore, an extra metallisation step was introduced compared to the traditional solution.

In the following section, an example of tolerance verification of a micro polymer part is given. The work is partly based on [22]. The part under investigation is a polymer part as illustrated in figure 7. It has four measurands of interest: the inner diameter of the centre hole (d, 1.550 ± 0.020), the outer diameter (D, 5.400 ± 0.030), the concentricity between the two circles (C, 0.020) and the height of the pillar in the bottom of the picture (H, 0.380 ± 0.030).

The dimensions are not all sub-mm, but the tolerances are all in the µm- range. The tolerance verification in an industrial environment was based on methods and equipment where the uncertainty to tolerance ratio was ranging from 20% to 70%. This made the verification virtually impossible.

The challenge in this situation is to establish a basis for quality assurance that gives a reasonable conformance zone. By experience the golden rule of the gauge maker (stating that the measurement uncertainty should not be more than 10% of the tolerance zone) cannot be met at this scale. Figure 8 illustrates three main sources of variations that contribute to the overall variation. Enough space should be left for being able to detect process variations, and the only way this can be obtained is by reducing (as much as possible) the variations introduced by the instruments and the metrology procedure.

In the present case the metrology procedure was changed from the traditional approach based on the use of calibrated instrumentation to a substitution method [22]. Figure 9 illustrates the two principles. The first method, in this case, resulted in too large uncertainties compared to...
the tolerance intervals. By choosing the substitution approach, the measuring instruments were “only” used as comparators. The main source of uncertainty would come from the calibration of the reference artifact/workpiece. In this case, a high precision tactile coordinate measuring machine (TCMM) was employed. These results yielded extremely good U/T values (see figure 10). However, the TCMM is slow, so an optical coordinate measuring machine (OCMM) was employed using the substitution method close to the production. Figure 10 shows two results for the OCMM: one result obtained “as is”, and one result obtained after compensating for a systematic error. The systematic error is introduced by the fact that the reference workpiece is calibrated using the TCMM and the measurements related to the production obtained using the OCMM. The two different measuring principles yield different results. It is clear that the choice of the correct metrology procedure highly influences the capability of verifying tolerances on micro scale.

Figure 9: Metrology procedures applied in micro manufacturing. Top: method based on calibration of instrument. Bottom: method based on comparison with calibrated artefact.

<table>
<thead>
<tr>
<th>Measured</th>
<th>Measured instrument</th>
<th>Workpiece</th>
<th>Calibration artifact</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>12.4</td>
<td>15</td>
<td>2.0</td>
<td>4.6</td>
</tr>
<tr>
<td>D</td>
<td>46.5</td>
<td>10.5</td>
<td>123.8</td>
<td>40.9</td>
</tr>
<tr>
<td>H</td>
<td>17.3</td>
<td>4.6</td>
<td>33.4</td>
<td>14.9</td>
</tr>
</tbody>
</table>

Figure 10: U/T values obtained using TCMM, OCMM without compensation of systematic errors, on OCMM after compensation of systematic errors.

7 SUMMARY

The design and manufacture of micro mechanical systems is characterised by a multi-disciplinary environment. The use of established design approaches such as axiomatic design, functional analysis and biomimetic design can be used in the conceptual design phases. The concept of micro engineering is introduced in this paper to identify important characteristics of the design and secure a coupling to manufacturing possibilities at an early stage. The paper discusses the different design approaches applied to micro mechanical systems and illustrates some issues based on two specific cases.

8 REFERENCES

Random Design: A Design of Experiment Method to Integrate Geometrical Deviations throughout the Product Lifecycle into Performance Simulation

F. Vignat, D. S. Nguyen, D. Brissaud
G-SCOP Laboratory, 46 avenue Félix Viallet, 38031 Grenoble Cedex 1, France
Frederic.Vignat, Daniel.Brissaud@g-scop.inpg.fr, nguyen.dson@gmail.com

Abstract
Robust performance is one of the most important keys in product design when many variation sources exist throughout the product lifecycle such as material defects, machining errors, and use conditions of the product. However, most of the performance simulation of the product is traditionally carried out by using the numerical model created in the CAD system. This model only represents the nominal information of the product. Thus, it is difficult to determine the relationship between the performance of the product and the variation sources, especially geometrical variation sources. A method is proposed in this paper to allow modelling the geometrical variations generated during the product lifecycle and integrating ones into the product performance simulation. As a result, the relationship between the performance and the geometrical variation sources can be established.

Keywords:
Product Lifecycle, Geometrical deviations, Performance Simulation, Robust performance

1 INTRODUCTION
Robust performance is one of the most important keys in product design since many variation sources exist in the real environment, including the manufacturing operations, material properties and operating environment. If these variations are not considered in the design process, they will prevent the performance of the designed product from meeting the requirements of clients. Furthermore, most of the product designers can easily create a numerical model of the product and the manufacturing process by a variety of computer aided tools, such as CAD/CAM system, CAE systems, product data management (PDM) and product lifecycle management (PLM). The current computer aided tools are actually effective to reduce the time for designing and developing a new product. However, the numerical model of the product is a nominal representation that does not take into account the variations generated all along the product lifecycle such as geometrical variations generated during the manufacturing and assembly stage. Moreover, most of the product performance simulations are carried out on the nominal model of the product created in the CAD system. The real performance of the designed product is different from the nominal performance and the geometrical variability obviously has an influence on it. The risk is then that the designed product doesn’t satisfy the requirements of the users. In this case, the product-process design has to be considered as not good or at the least not robust. There are therefore important issues that have to be considered:

- How to model the geometrical deviations of a product generated in the successive stages its lifecycle?
- How to manage the causes and consequences of geometrical deviations at the design stage?
- How to establish the relationship between the performance and geometrical deviations of the product?

2 LITERATURE REVIEWS
2.1 Manufacturing stage
Much research for modelling geometrical deviations generated during manufacturing stage has been done. These projects can be classified into two different
The first approach is based on the concept of state space model that is, firstly, mentioned by [3]. The models to describe dimensional variation propagation along multistage machining processes using this approach are proposed by [4-5]. A state vector \( x(k) \) is used to describe workpiece deviations at \( k^\text{th} \) stage. These deviations are accumulated on the workpiece from the previous stages \((1,...,k-1)\) of the multistage machining process. The set of state vectors \( x(k) \) describes the workpiece deviations relative to the nominal one resulting from the whole machining processes. This model provides a quantitative relationship between the fixture locator errors and the final workpiece geometrical error and has great potential to be applied to fault diagnosis and process design evaluation for complicated machining processes.

The second approach is based on the concept of small displacement torsor (SDT) that is proposed by [6]. The method based on the SDT approach to perform 3D manufacturing tolerancing for mechanical parts is, firstly, proposed by [7]. Thereafter, a model of manufactured parts (MMP) based on the SDT for simulating and storing the manufacturing defects in 3D has been developed by [8]. It permits the collection of deviations generated during a virtual manufacturing process. The defects generated by a machining process are considered to be the result of two independent phenomena: the positioning and the machining deviations accumulated over the successive set-ups. The positioning deviation is the deviation of the nominal part relative to the nominal machine. The positioning operation of the part on the part-holder is realized by a set of hierarchically organized elementary connections. The manufactured deviations of the surface relative to its nominal position in MMP are expressed by the parameters of the SDT. [9] proposed a graph representation of the manufacturing process. This graph models the successive set-up and for each set-up the positioning surface and their hierarchy and the machined surfaces. This graph makes it possible to highlight the influential paths. They propose then two analysis methods. The first one is based on the concept of SDT. The second one is based on the use of CAD software in which they model a manufacturing process with defects. They then virtually measure the realized part and check its conformity.

### 2.2 Assembly stage
A product is made up of parts assembled by the way of connections. Each part has already passed through the manufacturing stage where geometrical deviations have been generated. Then the product passes through the assembly stage of its life cycle. The assembly stage of the product life cycle is an essential stage, and it obviously brings its share of deviations to the product. In general, models for mechanical assembly can be categorized into two different approaches. The first one is based on the state space model. Some models representing this approach are proposed by [10-11] with Stream-of-Variation Model (SOVA) for 3D rigid assemblies dimensional variation propagation analysis in multi-station processes. The deviations accumulated at \( P^i \) assembly station are described by a vector \( X(i) \in R^{\text{dim}^i} \). The state space model of a multistage assembly process is represented by equations (1).

\[
\begin{align*}
X(i) - A(i)X(i-1) &\in B(i)U(i) + W(i) \\
Y(i) &\in C(i)X(i) + V(i)
\end{align*}
\]

Where \( U(i) \in R^{m(i)-d} \) is the fixture/part deviation contribution from station \( i \), \( Y(i) \in R^{d(i)-d} \) is the measurement obtained on station \( i \), \( W(i), V(i) \) are mutually independent noise and \( A(i), B(i), C(i) \) are transformation matrix.

The second approach is based on the SDT concept. [12] introduced the geometrical behavioural laws based on the SDT for modelling geometrical deviations in the mechanism. The reafer, [13] proposed a model to analyse part deviation in assembly. The positioning variation of the part relative to its nominal position in the global coordinate system is expressed by equation (2).

\[
D(AR) = E(AS) + T(S) + E(S) + D(R)
\]

Where:

\[
D(AR) \text{ is the variation of part } A \text{ relative to its nominal position in global coordinate system.}
E(AS) \text{ is the variation of surface } S_a \text{ of the part } A \text{ relative to its nominal position.}
T(S) \text{ is the variation of link between the } S_a \text{ of the part } A \text{ and the surface } S_b \text{ of the part } B.
E(S) \text{ is the variation of surface } S_b \text{ of the part } B \text{ relative to its nominal position.}
D(R) \text{ is the variation of part } B \text{ relative to its initial nominal position in global coordinate system.}
\]

A linear system of equations is created from the contribution of each connection between part \( A \) and part \( B \). The positioning variation of part \( A \) is determined by the resolution of the linear system of equations based on the Gauss-elimination method.

The main principle of the models proposed by [10-11,13] is to model the variation of the part at each stage along the assembly process. However, they do not link the geometrical deviations of the surfaces of the parts to the parameters of the manufacturing process. These parameters are those of the MMP [8] that can be measured from the manufacturing process as proposed by [9]. That is the reason why the authors have already proposed a model using the small displacement torsor in order to model geometrical deviations generated during the manufacturing stage and accumulated in the assembly stage.

### 2.3 Design stage
The geometrical deviations generated and accumulated during the product life cycle affect the performance of the product. It is thus necessary to manage their causes and consequences at the design stage in order to reduce their effect on the product performance all along its life cycle.

[14] proposed to use engineering models for developing robust design in order to reduce the variance of the design under variation of sources, such as manufacturing operations, variation in material properties, and operating environment. [15] addressed the impact of the manufacturing errors on the performance of the product. He defined the Manufacturing Variation Pattern (MVP) to represent the manufacturing characteristics and investigated its effects on the performance of the product. [16] presented a theory that offers an analytical and geometrical description of the performance sensitivity distribution of a product in the variation space. The theory can be applied to find a robust design that is less sensitive to the dimensional variation due to manufacturing errors or product wear. [17] proposed a new Probabilistic Sensitivity Analysis (PSA) approach for design under uncertainty based on the concept of relative entropy. This approach provides valuable information about the impact of the design variables on the performance of the product and whole range or a partial range of the performance distribution. [18] proposed a statistical approach in order to evaluate the impact of the geometrical variations on the...
angular rotational velocity between two bevel gears. Monte-Carlo simulation method is used to consider the geometrical behaviour simulation and tooth contact analysis. [19] proposed to integrate material and manufacturing process uncertainties in the design in order to consider their impacts on the performance of the product. They developed a procedure for uncertainty propagation from the material random field to the end product performance based on the product finite-element mesh.

These studies examine the impact of the variation sources or the geometrical variations on the product performance variation. Their approaches are, however, based on the nominal model of the product. They do not take into account the real variations coming from whole stage of the product life cycle. In addition, the concurrent product modelling technology, such as CAD/CAM, PLM, etc., is unable to model geometrical deviations generated during the product life cycle. Most of the relationship approximations between the performance and the design variables are only based on the nominal model of the product. Thus, a method that models the geometrical deviations of a product during its life cycle and integrates them into the performance simulation is proposed in this paper. The aim is to manage the geometrical variability throughout the product life cycle and its impact on the performance.

3 SIMULATION OF PRODUCT PERFORMANCE WITH GEOMETRICAL DEVIATIONS

The geometrical deviations of each surface of the product are generated and accumulated along its lifecycle. They will obviously have an influence on its performance. The geometrical deviation model, as presented in [1], can model them. The Monte-Carlo simulation method is used to create a set of m products with geometrical deviations. The product designers can be aware of distribution and variation of each surface of the product from the simulation result [2]. It is, however, unable to integrate all of them into a performance simulation of the product. Thus, we propose in this section, to use the design of experiments (DOE) approach to integrate more geometrical deviation parameters into the performance simulation and to determine the relationship between the performance and the geometrical deviation parameters of the product (see figure 2).

![Figure 2: Performance simulation of the product with geometrical deviations](image)

Full factorial designs are used to establish the relationship between the performance and the geometrical deviation parameters of the product, as presented in [20]. However, the experimental runs will be very large when the numbers of the selected factors and the levels have more than two. For example, there are \(3^n = 81\) runs in the case of four factors and three levels.

Another solution, proposed by [21], is to use Taguchi orthogonal arrays. Taguchi's orthogonal arrays are highly fractional orthogonal designs proposed by Taguchi. They are used to estimate main effects by only a few experiments. They are applicable to investigate main effects with many factors and levels. The factors are similarly selected by the expert knowledge in order to eliminate the factors that have small effects on the performance of the product. The numbers of levels of each factor are also defined by the result of geometrical deviation simulation [2]. The numbers of new nominal models of the product will be created by using Taguchi's orthogonal arrays. The number of runs is chosen corresponding to the number of factors and levels. For example, it is necessary to create 27 new nominal models of the product for the performance simulation in the case of 10 factors and 3 levels. In comparison with the factorial design, we need to create \(3^{10} = 59049\) nominal models for simulating the performance of the product. This is one of the advantages of the method.

In the case of increasing complexity alongside with factors number, the number of necessary simulations to determine the relationship between performance and deviations can become too large and thus time consuming even using Taguchi design. Moreover, the use of expert knowledge to determine key factors filters the deviation sources and can reach to the lost of some influential factors. Factorial design and Taguchi's orthogonal arrays are not effective in this case. Thus, we propose a random design method in order to address these issues. All product geometrical deviations parameters, called factors \(\{p_j\}_{i=1}^{n}\), that have small effects on the performance of the product are taken into account. The variation range of each factor is determined based on the results of geometrical deviations Monte-Carlo simulations, as presented in [2].

The random design approach is realized in 4 steps (see figure 3):

![Figure 3: Algorithm diagram of random design method](image)

- Step 1. Draw randomly a product with geometrical deviations in the set of products collected from the Monte-Carlo simulation.
- A \(k^\text{th}\) product with geometrical deviations is randomly drawn in a set of M products collected from the results of the geometrical deviation simulation. The value of each factor \(p_j\) is calculated based on the drawn product deviation parameters values. The set of values of the factors \(\{p_j\}_{i=1}^{n}\) is added into the \(k^\text{th}\) row of the design matrix P, as given in equation (3).

\[
P = \begin{bmatrix}
p_{11} & p_{21} & \cdots & p_{m1} \\
p_{12} & p_{22} & \cdots & p_{m2} \\
\vdots & \vdots & \ddots & \vdots \\
p_{1k} & p_{2k} & \cdots & p_{mk}
\end{bmatrix}
\] (3)
Step 2. Create the deviated CAD model.
The deviated model of kth product will be created in the CAD software corresponding to the value of each geometrical deviation parameter \( \{r_i\}_{i=1}^n \).

Step 3. Simulate the performance of the product.
The deviated model created in step 2 is used to simulate the performance of the product in order to determine the performance of the representative product. The result is appended into the response vector \( R \), as described by equation (4):

\[
R = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}
\]

Step 4. Eliminate the drawn product in the set of M products.
Repeat step 1. In this step, it is necessary to eliminate the product that has been drawn in step 1. The loop is repeated until the number of the drawn products is equal to N products.

Then the relationship between the performance of the product and the parameters of geometrical deviations \( \{p_i\}_{i=1}^n \) is obtained by using the linear or non-linear model based on the design matrix \( P \) and the response vector \( R \).

Then the relationship between the performance of the product and the factors is established by using the linear or non-linear model based on the realised simulations.

### 3.1 Linear regression model

To reduce the number of factors, key geometrical deviation parameters are defined based on expert knowledge. These key parameters are measured on the product and are functions of the elementary deviation parameters. The value of these factors can be calculated for the m products. Then the number of levels for these factors has to be defined depending on a compromise between the desired precision and the calculation time.

The value of each level of each factor is calculated according to its range of variation. The N values of the n parameters are then gathered in a design matrix \( P \), as shown in equation (5).

Two cases are encountered to determine the performance of the N products. In the simple case, the function performance is known and the performance value is obtained by the value of each parameter \( p_1 \) of each line of the design matrix into the analytical formula. In the complicated case, simulation tools as FEA, CFD, etc., are used to calculate the performance of the product. Thus a set of N deviated models of the product, corresponding to the geometrical parameters \( p_i \) of each line of the design matrix, has to be created in the CAD system. Each model is used to simulate the performance of the product in order to determine the performance of the N products that are called response vector \( R \). The response vector \( R \) corresponding to the design matrix \( P \) can then be filled, as expressed in equation (4).

The relationship between the performance of the product and the selected factors \( \{p_1, p_2, p_3, \ldots, p_n\} \) is established by the design of experiment method. This relationship can be expressed by equation (5).

\[
\text{Performance} = f(p_1, p_2, p_3, \ldots, p_n)
\]

For example, the linear least square fit model is used to establish the relationship in the case of n factors at 2 levels. From the result of 2*\( N \) simulations, the relationship is expressed by the function as shown in equation (6).

\[
f = \hat{\beta} p + \varepsilon
\]

Where:

\[
p = [p_1, p_2, p_3, \ldots, p_n]^T
\]

\[\hat{\beta} \]

is a coefficient vector of the model. It is calculated by equation (7).

\[
\hat{\beta} = (P^T P)^{-1} P^T R
\]

\[
R = [r_1, r_2, r_3, \ldots, r_N]^T
\]

\[\varepsilon \]

is a residual vector describing the deviation. It is calculated by equation (8).

\[
\varepsilon = E(R) = \frac{1}{N} \sum_{i=1}^{N} f_i
\]

The performance population of the product (for example, a million products) will be generated by replacing the value of the selected factors \( [p_1, p_2, p_3, \ldots, p_n] \) based on the collected data of the Monte-Carlo simulation in equation (5).

### 4 A CASE STUDY

A geometrical deviation model of a product during its lifecycle has been presented in the above section. Then the Monte-Carlo simulation method is used to estimate the probability distributions of the geometrical deviations of the product. As a result, the relationship between the performance and the product is established by the design of experiment method. Then a population of the product performance is generated by using the collected data in the Monte-Carlo simulation stage. In order to illustrate the method, we propose, in this paper, a simple example of a spring system using three proposed methods.

#### 4.1 Spring system

**Geometrical deviation and frequency model**

The numerical model of the spring system is modelled in CAD software. It is represented in figure 4. The frequency of the spring system is investigated in this case.

The frequency of the spring system is expressed by equation (9).

\[
f = \frac{1}{2\pi} \sqrt{\frac{k}{m}}
\]

Where:

\[
m = \rho V
\]

\[k\]

is called spring constant or spring stiffness. It is calculated by equation (11).

\[
k = \frac{Gd^4}{8nD^3}
\]
Where:

- \( E \) - Young's modulus
- \( d \) - Spring wire diameter
- \( D \) - Spring outer diameter
- \( n \) - Number of active windings
- \( \nu \) - Poisson ratio

\[
G = \frac{E}{2(1+\nu)}
\]

From the expert knowledge that has been presented above, the frequency of the spring system depends on many parameters, especially the geometrical deviations such as deviations of the cylinder and plan shaping the load, spring wire diameter and spring outer diameter.

The load of the spring system is manufactured by the manufacturing process shown in figure 5.

The geometrical deviations of all surfaces of the load are modelled by the manufactured part model (MMP). For example, those of the plan \( P_1 \) of the load are expressed by following tensor:

\[
T_{P_1, P_1} = \begin{bmatrix}
-60 r_{x_{1,251}} - r_{x_{1,2}} & -60 r_{y_{1,251}} - r_{x_{1,251}} + 60 r_{y_{1,251}} - 60 r_{y_{1,251}} \\
+ r_{x_{1,4}} - r_{x_{1,251}} & + 50 r_{y_{1,2}} - 60 r_{y_{1,251}} + r_{x_{1,2}} + r_{x_{1,4}} - r_{x_{1,251}} \\
+ r_{y_{1,4}} - r_{y_{1,251}} & + 50 r_{x_{1,2}} + 60 r_{x_{1,251}} + r_{y_{1,2}} + r_{y_{1,4}} - r_{y_{1,251}} \\
0 & 0
\end{bmatrix}
\]

The Monte-Carlo simulation is used to simulate the geometrical deviations of all surfaces of the spring system. For example, the distributions of the cylinder's deviations of 100000 loads are described in figure 6.

The histogram of the rotational deviation \((R_x, R_y)\) and the distribution of the translational deviation \((T_x, T_y)\) of the cylinder relative to the axis \(X\) and \(Y\) are shown in figure 6a, 6b and 6c respectively. From this result, the variation of the load mass is calculated by the geometrical deviations of two plans and a cylinder of the load. It is described in figure 6d.

From the expert knowledge presented above, the relationship between the frequency of the spring system and its geometrical deviations is expressed by equation (13).

\[
f = \frac{1}{4\pi^2} \sqrt{\frac{G}{\pi n_f(D+A)^3}} \left[ \frac{G (d+\delta)^4}{2h + R^2} \left( \frac{h + R^2}{2} + h + R^2 + h + R^2 \right) \right]
\]

Where:

- \( \Delta, \delta \) - Dimensional deviation of the outer and wire diameter of the spring.
- \( R, h \) - Radius and height of the cylinder of the load.
- \( T_{x_1}, T_{x_3} \) - Translational deviation of two plans of the load.

The frequency simulation of this system can, in fact, be realised by using equation (13) directly. The histogram of one hundred thousand frequencies is easily shown in figure 7, because the time for calculation is small in this case. The design of experiment method is, however,
preferred to illustrate realisation of the proposed method. We can, in addition, compare the accuracy among them.

Random design

This method is realized by the 4 steps as presented in the above section. Ten spring systems are randomly drawn from ten thousand spring systems. The ten frequencies are calculated by equation (13). Then the linear regression fit model is used to establish the relationship between the frequency and all parameters of geometrical deviations of the spring system. It is expressed by equation (14).

\[ f = 4.91592 + 2.53885\delta - 0.1232\Delta - 0.201746\text{dr} - 0.0387945 Tz_1 - 0.0448108Tz_3 \]  
(14)

In comparison of the impact coefficient of each geometrical deviation parameters, there is not too much difference between the random design approach and the real one. However, it is necessary to realise 100,000 runs if using the real approach. In the complex case, the random design method is obviously better then the real approach in terms of time and cost.

<table>
<thead>
<tr>
<th></th>
<th>Exact model</th>
<th>Random design method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of frequency</td>
<td>4.99506</td>
<td>5.0719</td>
</tr>
<tr>
<td>Standard deviation of frequency</td>
<td>0.589</td>
<td>0.596</td>
</tr>
<tr>
<td>Deviation parameters</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Number of runs</td>
<td>100000</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1: Summary of the proposed methods.

5 CONCLUSIONS

This paper proposes a new random design of experiment method to establish the relationship between product performance and the geometrical deviations of the product surfaces. This relationship is used to overtake the limitations of the concurrent product modelling technology such as CAD/CAM, CAE, PLM, etc. by integrating the geometrical deviations generated during the production stage of the product lifecycle into performance simulation. The geometrical deviation model already proposed by the authors [1-2] gives an image of the population of produced parts with geometrical deviations. Using this relationship, an image of the population of the product performance is calculated.

In previous papers, two others design of experiment methods have been proposed to establish this relationship: full factorial design [2] and Taguchi design [21]. Random design of experiment, proposed in this paper, is more effective when it is difficult to determine the key factors that have a strong influence on the performance of the product, or when the number of the factors is considerably large.

6 REFERENCES


Abstract

GeoSpelling as the basis of the Geometrical Product Specifications (GPS) standard [1] enables a comprehensive modeling framework and an unambiguous language to describe geometrical variations during the overall product life cycle. This is accomplished by providing a set of concepts and operations based on the fundamental concept of the “Skin Model”. However, the definition of GeoSpelling has not been successfully completed. This paper presents a novel approach for a formal description of GeoSpelling concepts. In addition to mapping fundamental concepts and operations to discrete geometry objects, we investigate the use of Monte Carlo Simulation techniques for skin model simulation when considering geometrical specifications. The results of the skin model simulations and visualizations are shown and the performances of the described simulation methods are compared to each other.

Keywords:

Geometrical Product Specifications, GeoSpelling, Discrete geometry, Skin model, Multi-Gaussian distribution, Gibbs sampling

1 INTRODUCTION

The control of product geometrical variations during the whole development process is an important issue for cost reduction, quality improvement and company competitiveness in the global manufacturing era [2].

During the design phase, geometric functional requirements and tolerances are derived from the design intent. The modeling of product shapes and dimensions is now largely supported by geometric modeling tools. However, permissible geometrical variations cannot be intuitively assessed using existing modeling tools, and this results in the specification uncertainty.

In addition, the manufacturing and measurement stages are the main geometrical variations generators according to the two following axioms [3-4]:

• Axiom of manufacturing imprecision: all manufacturing processes are inherently imprecise and produce parts that vary.

• Axiom of measurement uncertainty: no measurement can be absolutely accurate and with every measurement there is some uncertainty about the measured value or measured attribute.

To reduce the total uncertainty, the product geometrical variations should be considered during the whole product life cycle (figure 1).

In order to evaluate product geometrical variations and ensure that the fabricated product can satisfy the functional requirements, designers should determine the limited values that constrain product geometrical variations. This process is now well known as specification.

In the context of Digital Mock-Ups (DMUs), the design process is supported by modeling, simulation and visualization tools such as CAD systems. The Digital Mock-Up, as a “digital” alternative to constructing physical parts, should be enriched by geometrical variation models to allow testing of design errors on assemblies and realistic visualization of the product. At the manufacturing level, multiple representations based on smooth surfaces and discrete representations (triangular meshes) are considered. Moreover, an ordered or unordered set of points resulting from manufactured part acquisition is processed for the purpose of product inspection.

Figure 1: Geometrical variations during the product life cycle.

A comprehensive view of Geometrical Product Specifications should consider multiple geometric representations, and as well as suitable processing techniques and algorithms. The discrete geometry theory can offer a great support in this area, since discrete geometry is a mathematical research field related to geometrical objects whose nature or property is discrete. Therefore, it can provide the theory to handle both point and polyhedral mesh based descriptions.

The organization of this paper is as follows. After a comprehensive review of Geometrical Product Specifications and geometric tolerancing approaches (section 2) we show their limitations in considering multiple geometric representations and non-ideal entities. The principle of discrete geometry for GeoSpelling is described in section 3. Skin model simulation and visualization are highlighted in section 4. Afterwards, the skin model simulations which consider geometric
tolerance and results comparison are developed in section 5. The conclusion is given in section 6.

2 RELATED WORK

Many efforts to build specification models for geometrical tolerancing have been attempted in recent years. The existing approaches can be mainly classified into standard-based and mathematical models for tolerancing. The standard-based methods rely on technical drawings, and are based on the concepts of tolerance features, tolerance zones and datum. This geometrical tolerancing representation was adopted by ISO 1101-2004, ISO 5459-2000, and ASME Y14.5-2009, and it was the most popular way to describe tolerance requirements in the past years. However, this method cannot keep up with current tolerance requirements, since it is based on human interpretation and is not convenient to transfer the data what is now a digitally-based industry.

Mathematical models for tolerancing can be classified into several groups. The offset zone approach proposed by Requicha [5] obtains the tolerance zone by offsetting the ideal feature a certain distance and this method is suitable to geometrical models with simple shape representations. Jayaraman and Srinivasan introduced the Virtual Boundary Requirements (VBRs) method [6] to improve the offset zone method and to define the virtual boundary by mathematical foundations. The VBRs method considers assembly and material volume requirements. However, its shortcoming is that the results are not compatible with GPS standards and cannot describe all kinds of tolerances. Hoffman [7] and Turner [8] defined tolerancing models in different dimension spaces, and Fortini [9] introduced the vector tolerancing concept in parameter space, and then Wirtz [10] argued that vector tolerancing should be included in the ISO standards. The shortcoming of the vector tolerancing method is that it is not able to describe the tolerance features and the geometrical variation requirements. Bourdet and Clement [11] proposed the Small Displacement Torsor (SDT) theory, which can describe the tolerancing types by the small rigid displacement movement of geometric features. In contrast, this method is only appropriate for ideal features. Clement and Riviere [12] introduced the Technologically and Topologically Related Surfaces (TTRS) theory. According to TTRS, three-dimensional surfaces or features are classified according to their respective degree of invariance under the action of rigid motions. Basically, seven main features equivalent to kinematic lower pairs are identified: planar feature, cylindrical feature, revolution feature, spherical feature, prismatic feature, helicoidal feature and complex feature. Each main feature is then described by a unique minimum geometrical reference element (MGRE) that allows positioning in Euclidean space. An MGRE is set as a combination of elementary geometrical objects: point, line and plane. TTRS Theory has been adopted by ISO TC213 and successfully implemented in the CATIA v5 CAD system to manage assembly constraints and tolerance annotations.

All of the methods described above cannot consider non-ideal features, and some of them even lead to ambiguous interpretations. The model of GeoSpelling [13] adopted by ISO-17450 allows a unified description of ideal and non-ideal features and permits a unique expression of mathematical parameterization of geometric features.

3 DISCRETE GEOMETRY FUNDAMENTALS OF GEOSPELLING

GeoSpelling proposed by Mathieu and Ballu [14] is used to describe both ideal and non-ideal geometric features [1]. Indeed, it allows the expression of product specifications from function to verification with a common language. This model is based on geometrical operations which are applied not only to ideal features defined by CAD systems, but also to the non-ideal features which can represent a real part. These operations include partition, extraction, filtration, association, collection and construction items.

Discrete geometry research focuses on basic discrete geometrical objects, such as points, segments, triangles and other convex discrete shapes, and it is quite efficient to implement digital discrete processing techniques. Therefore, discrete geometry theories and techniques are suited to enhance the data processing capabilities of GeoSpelling. Based on the standard [4], a specification is defined as a condition on a characteristic defined from geometric features which are created from a skin model by different operations. The concepts of "characteristic," "feature," and "operation" are then mapped to their underlying discrete geometry mathematical concepts as summarized in table 1.

![Table 1: Concepts mapping between GeoSpelling and discrete geometry.](attachment:image)

### 3.1 Features

In GeoSpelling, features include non-ideal features and ideal features. In discrete geometry, non-ideal features are discrete shapes, such as points, segments, triangles, point sets, polylines, and polyhedral meshes. Ideal features are derived from the classification of 3D surfaces based on their invariance under the action of rigid motions. Ideal features can be obtained by association operations.

### 3.2 Characteristics

In GeoSpelling, characteristics include distances and angles. While in discrete geometry, distances are defined between discrete shapes: point-point, point-segment, point-triangle, segment-segment, segment-triangle and in a general case to Hausdorff distances. The angles are related to the three well-known cases: angles between segment-segment, segment-plane, and plane-plane.
3.3 Operations

Partition operation is used to identify bounded features [1]. In discrete geometry this kind of operation is called segmentation. The majority of point set segmentation methods can be classified into three categories: edge-detection method, region-growing method and hybrid method [15]. The main problem of the edge-based method is that when the points are near sharp edges they are quite unreliable. This problem means that the edge-based method has a relatively high sensitivity to occasional spurious points. The advantage of face-based techniques is that they work on a larger number of points to reduce the risk of sensitivity to occasional spurious points, and they can identify the points that belong to each surface, but the main disadvantage is time processing. The hybrid method has been developed by combining the edge-based and region-based methods together to overcome the limitations involved in the original methods.

Extraction operations are used to identify a finite number of points from a feature with specific rules [1]. In discrete geometry these rules are equivalent to sampling techniques. Zhang et al. [16] classified the extraction strategy into four categories: grid extraction, stratified extraction, special curve extraction and point extraction. Depending on the invariant class of the surface, users can determine the prior extraction strategy. Other extraction strategies were investigated in literature [17], such as Fanshawe sequence sampling, the Hallon-Zaaremba sequence, Aligned systematic sampling, and Systematic random sampling.

Extraction operations are used to distinguish roughness, waviness, form, and so on, by separating the different wavelength components into predefined bandwidths [1]. There are already some options in today’s GPS standards, such as polynomial fitting, 2RC filtering, Gaussian fitting, wavelet filtering, etc. [18]. In discrete geometry, signal processing filtering techniques and other techniques such as outlier removal, based on a certain criterion, are reported in [19].

Association operations are used to fit ideal feature(s) to non-ideal feature(s) according to specific rules [1]. In discrete geometry, association operations fit ideal feature(s) to discrete geometric feature(s) according to given criteria, such as the Moving Least Squares (MLS) method. MLS methods take the distance influence into account when calculating the association arithmetic [20].

Collection operations are used to consider some features together, which play a functional role, and construction is used to build ideal feature(s) from other ideal feature(s) [1]. In discrete geometry union-Boolean operations and intersection-Boolean operations [21] respectively have the same capability.

4 SKIN MODEL SIMULATION

The skin model is a non-ideal surface model. It is a virtual model imagined by designers when taking into account different kinds of geometric defects. The main originality of GeoSpelling is to build geometric models for tolerancing specification not from nominal models but from the skin model itself. It can also help designers to express specifications corresponding to manufacturing requirements. Few research studies have focused on the skin model simulation. Chiabert developed a shape identification method of the skin model using rigid body motion and Monte Carlo simulation [22]. Samper [23] proposed a finite element analysis method to simulate form defect expressions of skin models. However, there is no uniform way to express the skin model nowadays. Therefore, the skin model simulation method will be discussed here. Three different statistical methods are considered in this paper: 1D Gaussian method, multi-Gaussian distributions and Gibbs sampling. The details of each method are explained below.

4.1 1D-Gaussian method

In probability theory, the Gaussian distribution is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to cluster around a single mean value. The graph of the associated probability density function is “Bell” shaped as showed in figure 2.

![Figure 2: The principle of 1D-Gaussian method.](image)

The principle of the 1D-Gaussian method can be described in figure 2. The “bell” shape reflects the scope of the 1D-Gaussian distribution acting on one point. It is a random value and it can be calculated by the probability density function of the 1D-Gaussian (Formula 1). In this formula, the mean value is the input points’ coordinates, and the variance determines the width of the Gaussian distribution. This method uses the Gaussian variable as the deviation value in the direction of vertex normal (the vertex normal estimation method is explained in section 3.3), then it applies this calculation to each point, the distribution parameters yield formula 2, and then the result is illustrated in figure 3 with different views of the skin model of a plane.

\[
X \sim N(\mu, \sigma^2) : f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)
\]

\[
u \sim N(0,1)
\]

Where \(\mu\) is the mean value, \(\sigma^2\) is the variance, and \(x\) is the Gaussian variable.

![Figure 3: 1D-Gaussian skin model.](image)

4.2 Multi-Gaussian method

A multivariate Gaussian distribution is used to generate random vectors. The trivariate Gaussian distribution is considered in this work. A spatial random vector is defined as \(X = [X_1, X_2, X_3]^T\). The probability density function of multivariate Gaussian distribution can be expressed as formula 3.

\[
X \sim N(\mu, \Sigma) : f(x) = \frac{1}{(2\pi)^{3/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)\right) \quad (3)
\]
Where $\Sigma$ is the covariance matrix, $|\Sigma|$ is the determinant value, and $\mu$ is the mean vector.

The principle of this method is described in figure 4. The ellipsoid reflects the scope of 3D-Gaussian distribution acting on one point. It is a random vector and can be calculated by the probability density function of the 3D-Gaussian (formula 3). In this formula, the mean value is the point's coordinates, and the relationships among each axis are constrained by the covariance matrix. Figure 5 displays different views of the skin model of a plane when applying this calculation to each point.

![Figure 4: The principle of 3D-Gaussian method.](image)

<table>
<thead>
<tr>
<th>Point set</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D-Gaussian distribution</td>
<td></td>
</tr>
</tbody>
</table>

**4.3 Gibbs Method**

The Gibbs sampling algorithm is used to generate a sequence of samples from the joint probability distribution of two or more random variables. It is an iterative method based on Markov chain Monte Carlo (MCMC) algorithms. It aims to design a Markov chain whose stationary distribution is the target distribution. It requires an initial value of the parameters, and at each iteration, each parameter of interest is sampled a given value from the other parameters and data. Once all the parameters of interest are sampled, the nuisance parameters are sampled given the parameters of interest and the observed data [24-25]. This characteristic of the Gibbs method can make the random distribution of point set of skin models approximate a Gaussian distribution.

In order to determine the number of iterative runs, the probability distribution of the normality assumptions of Gibbs should be considered. The normal distributions are simulated using Minitab software for different iterative runs (figure 6). The results show a good convergence at 10000 iterations.

![Figure 6: Normality assumption of Gibbs method.](image)

![Figure 5: 3D-Gaussian skin model.](image)

(a) skin model (b) front view (c) left view

**4.4 Skin model visualization**

Tolerance values are much smaller than features' sizes, so it is difficult to visualize the skin model with multi-scale geometry. This section proposes to use RGB color scale mapping technique to visualize the geometrical deviations on the vertex normal direction. The skin model can be reflected by a continuous color strip.

**Vertex normal estimation**

A normal vector is a local geometric property of a 3D surface, which is specific to a given point or a planar facet. Many attempts have already been made for reliable estimation of normal vectors from discrete point data [15]. Given a polyhedral mesh surface, the normal vector at a vertex can be estimated as the weighted average of the normal vectors of the adjacent triangle facets around it. Considering an arbitrary vertex $p$ in a discrete mesh surface $\Sigma$, assuming its neighbor contains $N$ triangles, then the normal vector at $p$ could be estimated using formula 4.

\[
\hat{n}(p) = \frac{\sum_{i=1}^{N} \omega_i \cdot \hat{n}_i}{\left| \sum_{i=1}^{N} \omega_i \cdot \hat{n}_i \right|} \tag{4}
\]

Where, $n_i$ ($i = 1, ..., N$) indicate the unit normal vector of the $i$th triangle facet. $\omega_i$ ($i = 1, ..., N$) are the weight coefficients corresponding to the normal vectors of facets $f_i$. 

Suppose $X = (X_1, ..., X_d)$ is the random variable in $\mathbb{R}^d$, and $X_{i,j} = (X_{1,j}, ..., X_{d,j})$ is the random value of $d - 1$ dimension. Note that the conditional density function of $X_i \mid X_{-i}$ is $f(X \mid X_{-i})$. Therefore, the Gibbs sampler select candidate points from this dimension conditional distribution. The related process is that, at beginning, the time is equal to zero ($t = 0$) and it has an initial value $X(0)$. When $t$ is increasing ($t = 1, 2, ..., T$), then $X(t)$ follows a certain function to generate new point to replace old one and iterative calculation is performed until it converges to the target value. The corresponding pseudo-code is described as follow.

1. Let $x_i = X_i(t - 1)$
2. Let $j$ is a variable between $[1, d]$. For $j = 1, 2, ..., d$, using $f(X_j \mid x_{-j})$ to get candidate point $X'_j(t)$, and then update $X'_j(t)$.
3. Let $X(t) = (X'_1(t), ..., X'_d(t))$ and then increase $t$.

This iterative process generates random variables, which follow the bivariate normal distribution, can simulate the skin model. The result is showed in figure 7.

![Figure 7: Gibbs Sampling Skin model.](image)

(a) skin model (b) front view (c) left view
The method used here for the weight coefficients computation considers the influence of the area of each adjacent triangle facet and the distance between the given vertex and the barycenter of each adjacent facet. Parameter $\omega_i$ can be calculated by formula 5.

$$\omega_i = \frac{A_i}{d_i^2}$$  \hspace{1cm} (5)

Where, $A_i$ ($i = 1, \ldots, N$) represents the area of the $i$th triangle facet. $d_i$ ($i = 1, \ldots, N$) are the distances between the vertex $p$ and the barycenter of the $i$th triangle facet. Parameter $N$ is the number of all the triangle facets adjacent to the given vertex.

The notations mentioned in the above formula are described in figure 8. Figure 9 shows examples of vertex normal estimation considering planar and cylindrical shapes.

**5 CONSTRAINT-BASED SKIN MODEL**

After creating the random point set to simulate the unconstrained skin model, geometrical and dimensional tolerances should be considered to satisfy the specification requirements. The following section mainly discusses the form, orientation and position tolerance considerations to enhance the skin model simulation.

**Form specification**

To estimate form specification, the first step is to determine the tolerance zone direction using the Principal Component Analysis (PCA) method. Then all of the point set should belong to the tolerance zone. The principle of this method can be described by figure 11, where $n$ is the vector of principle direction of the point set, $M_1$ and $M_2$ are two arbitrary points, and $d$ is the distance between these two points in the direction of the vector $n$.

**Orientation and position specification**

Besides form constraints, the orientation and position constraints must also be considered. For these two constraints, the tolerance direction is the same as the
normal direction of the datum plane and all the points should be within the tolerance zone. The parallelism specification (figure 12) satisfies the constraints in formula 9.

\[ d = \left| \text{Max}(m \cdot n) - \text{Min}(m \cdot n) \right| \leq t_{\text{parallelism}} \]  

(9)

![Figure 12: Principle of parallelism specification.](image)

Where \( n \) is the normal direction of the datum plane, \( M_i \) is an arbitrary point, and \( d \) is the distance between two points in the direction of vector \( n \).

For position specification (figure 13), the related constraint is described by formula 10.

\[ d = \text{Dist}(m, P_A) \in [a - \frac{t_{\text{position}}}{2}, a + \frac{t_{\text{position}}}{2}] \]  

(10)

![Figure 13: Principle of position specification.](image)

Comparisons

The skin model is created using the three different simulation methods proposed in this paper. A point cloud of a plane composed by 273 points is the reference test, and the specification constraints are flatness, parallelism and position tolerances which are equal to 0.01 mm, 0.02 mm and 0.05 mm respectively. The 50 skin models are generated and the main statistical characteristics are computed for comparison. In the Gibbs sampling method, the iterative time is equal to 10,000.

The comparison items include the average deviation value, the limit value, and the processing time. The distribution of the deviation values is computed using Minitab statistical software. The result is shown in figure 14. Corresponding standard deviations are summarized in table 2. From this table, it can be deduced that the Gibbs method offers the closest simulation result to the target value, but it is much more time consuming.

<table>
<thead>
<tr>
<th>Items</th>
<th>1D-Gaussian</th>
<th>3D-Gaussian</th>
<th>Gibbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Value</td>
<td>0.008666</td>
<td>0.008705</td>
<td>0.008883</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.000567</td>
<td>0.001287</td>
<td>0.000521</td>
</tr>
<tr>
<td>Maximum value</td>
<td>0.009869</td>
<td>0.009831</td>
<td>0.009778</td>
</tr>
<tr>
<td>Minimum value</td>
<td>0.006943</td>
<td>0.000510</td>
<td>0.007488</td>
</tr>
<tr>
<td>Time</td>
<td>&lt;0.001 s</td>
<td>&lt;0.001 s</td>
<td>0.297 s</td>
</tr>
</tbody>
</table>

![Figure 14: Statistical distribution results.](image)

6 SUMMARY AND CONCLUDING REMARKS

In this paper, we saw that discrete geometry for GeoSpelling provides a new mathematical framework for Geometrical Product Specifications. Starting from the fundamental concepts of the skin model and non-ideal features, skin model simulation and visualization using 1D Gaussian, 3D-Gaussian and Gibbs method is developed and compared.

With new foundations for Geometrical Product Specifications, this paper concludes that discrete
geometry and statistical shape techniques are promising approaches towards skin model consideration during the product life cycle.

7 ACKNOWLEDGMENTS
The first author extends the sincere thanks to the China Scholarship Council (CSC) for their financial support for this work.

8 REFERENCES
Individual Contour Adapted Functional Implant Structures in Titanium

C. Schoene¹, R. Stelzer¹, P. Sembdner¹, J. Markwardt², B. Reitemeier³, G. Engel⁴
¹ Department of Mechanical Engineering, Institute of Machine Elements and Machine Design
Dresden University of Technology, 01062 Dresden, Germany
² Department of Oral and Maxillofacial Surgery, ³ Department of Prosthethetic Dentistry
Dresden University of Technology, 01062 Dresden, Germany
⁴ Product Development Company Hofmann and Engel GmbH in Boxdorf near Dresden
christine.schoene@tu-dresden.de, ralph.stelzer@tu-dresden.de, philipp.sembdner@tu-dresden.de

Abstract
Apart from applications in mechanical engineering, now even the domain of medicine may benefit from the option of using metallic materials for Direct Manufacturing. In the medical domain, the use of biocompatible materials, such as titanium or titanium alloys is essential to produce individual implants. As a result of this development, it is now possible to generate new patient-specific geometries fitted to the contour. This paper elucidates the process chain to derive individual design variants and to produce patient-specific bone replacement implants for the lower jaw-bone regions by using innovative reverse engineering and manufacturing methods. For this interdisciplinary project, technical scientists, medical scientists at the university hospital and engineers from a product development firm work together.

Keywords:
Rapid Product Development, Direct Manufacturing, Reverse Engineering

1 INTRODUCTION
Scientific studies of CAD/CAM applications in medicine and dental prosthetics, which have been ongoing for approximately 15 years, focus on models, tooth crowns and bridges [1-3]. The use of CAD/CAM technology with CNC-milling and rapid manufacturing in dental industry is now very common. The majority of scientific approaches to use rapid manufacturing since 2006 have been aimed at endoprostheses in CoCr and titanium alloys [4-8].

In cases when jaw implants are required due to disruption of continuity in the lower jaw bone, the relevant anatomical regions are commonly represented by means of imaging techniques, such as computer tomography (CT). Based on these data, which are specific to each patient, we analyse the range of available standard care programs, choose the appropriate reconstruction plates for the jaw region and implant them. As a rule, the reconstruction plates currently in use are characterised by great variations in stiffness between implant and bone (Figure 1).

Figure 1: Standard reconstruction plate for jaw region

In Figure 2 a patient is illustrated with aesthetic and clinical deficits as a result of the damage of a standard reconstruction plate after surgery and tumour resection on the left facial side. Additionally, since the geometry is not significantly individualised, we also see obvious functional and structural-mechanical deficits, as well as aesthetic disadvantages. Consequently, an application that protects the tissue and is also highly stable, which is a necessity for optimal treatment, cannot be provided by any of the methods currently available [9-11].

As a result of ongoing globalisation, the greatly expanding market for medical implants made of biocompatible high-performance materials is under ever-increasing pressure from competitors. In this context, the reconstruction of bone defects, in particular in the oral, jaw and facial region, by means of osteosynthetic plates is regarded as

Figure 2: Patient after damage of standard reconstruction plate
a great challenge. Here, special advantages may accrue to a new implant design whose contour and stiffness are tailored to specific geometric and elastic conditions, since in this way it is possible to reduce complications during ingrowth.

The LaserCUSING® method [12-13] provides the first technological approach to manufacturing new filigree implants that are perfectly aligned with the contour and gradually modified in stiffness. LaserCUSING® is an innovative technique, following a generative approach, which is able to realise structures according to the direction of force action.

One objective of the planned research project is aimed at the development of a process chain that extends the way from CT layer images of a diseased patient up to the manufacturing of individual bone substitute implants for the patient while taking into consideration a Rapid Manufacturing technique. Thus, the rapid manufacturing of individual implants that repair defects is primarily emphasized in order to keep the waiting periods for patients as short as possible.

2 APPROACH

The project consortium began with mandibular implants in 2006.

The work we are doing in Dresden is unique in that it features close interdisciplinary co-operation among radiologists and oral and maxillofacial surgeons, dentists, and engineers. The study includes an ethics proposal for animal experiments and a patent.

The CT data required for diagnosis are also used to generate the virtual 3D model of the jaw bone. Individual design and modifications of the implant are performed based on the 3D model. The fundamental steps necessary to generate individual implants are listed below:

- Imaging of the diseased area and surrounding regions by means of CT
- Creation of a discrete surface model from the CT image stacks
- Alignment of the lower jaw model for design in a defined co-ordinate system
- Definition of cutting planes to isolate the defective regions
- Generation of a mathematical surface representation of the lower jaw contour in the affected area and the surrounding regions
- Definition of the positions of the holes that will later secure the implant in the residual bone
- Implant design with CAD system (Computer Aided Design)
- Design of the cutting templates, which are applied to the jaw before resecting the diseased bone
- Production planning for the LaserCUSING® system incl. placement in the working space
- Creation of the support structure taking into account the building layer configuration
- Manufacturing of the implant and the cutting templates with LaserCUSING®
- Removal of the implant from the building plate by means of erosion
- Manual removal of the support structure
- Corundum blast finishing of the implant surfaces.

2.1 From the CT image stack to the CAD solid Model

When generating the individual implant, the first step consists of mapping the defective bone regions and the surrounding soft part tissues by means of CT techniques. The result of this data acquisition procedure is made available in single images in the DICOM format. The next step is to read these images by means of VoXim® [14]. The following step is 3D soft tissue segmentation, wherein materials of different density, such as soft tissue and bone, are separated from each other. Afterwards, a faceted 3D model of the segmented bone regions is output in the STL format. This model is important for ongoing design.

Thus, discrete data is first made available. Now it is necessary to reverse these data into a solid model of mathematically correct representation for CAD modelling. To do this, the faceted data are processed with the Geomagic Studio [15]. The polygons are subject to various repair- and filtering measures. Here, the following steps are to be run in order:

- Diversify (deletes stand-alone object regions)
- Remove peaks
- Repair cuts
- Smooth
- Manually rework

At this point, we also define a uniform co-ordinate system and the cutting planes in co-operation with the surgeon. These cutting layers define the region that is to be removed in surgery and replaced with an implant. In this cutting region, the contour of the inner bone structure, the cancellous bone, is reversed as a Spline curve and simultaneoulsy stored separate from the lower jaw model. This information is significant, among other things, for the definition of the positions of the holes that will later hold the implant in place. These positions should be in the region of maximal layer thickness of the corticalis (outer bone region).

In the strictest sense, Reverse Engineering describes the procedure of 3D digitising of workpieces with sculptured surfaces, conditioning the 3D point data and converting them to CAD models [16]. Positive results will only be obtained by 3D data recording in conjunction with qualified and problem-oriented data conditioning and application in the follow-up computer-aided strategy. Another equally important fact is that the CAD representation successfully withstands production planning, manufacturing and quality inspection [17].

Reverse engineering is performed then to generate a parametric solid model. The solid model is stored in the STEP format and is now available for design. The path from physical object to CAD solid model is represented in Figure 3.

2.2 Design of the individual implant

Current designs of implants that are identical in contour orient themselves to the organic bone structure configuration. They consist of an outer mounting shell design and an inner filigree tissue structure. The designed outer shell of the implant correspondingly follows the contour of the removed jaw region. Manufacturing technology allows this to be very thin in shape in order to reproduce the stiffness and strength of the bone. Thin-walled envelope geometry of about 0.3 mm thickness is realized; thickness in the areas attached to the residual bone is 0.4 mm.
In cases that the tumour destroyed the bone, there is no useful geometry. Therefore, a database, containing characteristic curves, is planned (Figure 4).

![Figure 4: left: Registration of 5 different jaws, right: curves after virtual jaws cutting in position between tooth 31 and 41](image)

It is also possible to intentionally include discontinuities in the enveloping geometry.

The shell design makes it possible to implement an inner filigree structure as well. These structures inside the implant are expected to offer greater reliability in the bone regeneration. In this approach, these structures may have a different geometric shape, as well as stochastic discontinuities and different dimensioning.

Design is carried out considering the positive contact with the residual jaw on both sides. In general, we are investigating two different ways to attach the implant to the residual jaw (Figure 5). The first connecting type (Figure 5, left) is based on a variant in which the implant is secured to the outer lower jaw contour. In the second variant (Figure 5, right), the implant is shifted into the bone and cemented with a suitable bone substitute material. The choice of variant must be based on the individual case.

The principal design procedure for an individual implant is described below using the variant in which the implant is attached to the residual bone:

- Insert section layers in lower jaw model
- Section the jaw model
- Prepare lower jaw stubs of residual bone (remove milling region)
- Transfer surface information of the milled overlapping regions
- Offset surfaces by wall thickness value
- Define length of the overlapping regions for the implant
- Fill in enveloping geometry according to the cut section of the jaw regions
- Insert inner filigree structure.

Before repairing the defect with an implant, it is necessary to remove the corresponding jaw region in a surgical operation. The section layers required have already been defined virtually in an earlier step. Since the operating team does not have access to these virtual layers during implantation, the position of the section layers is predefined in the form of cutting templates (Figure 6). These templates guarantee an unambiguously positive contact at the jaw. Figure 7 demonstrates one variant of the attached implant to the pig cadaver jaw.

The cutting templates are also designed with the CAD model of the lower jaws. Beginning with the parting planes where the resection will be performed, the adjacent surfaces are derived. From this, we create a two-piece body. One lateral surface of the body is used as the cutting surface along which the medical doctor moves the saw. Two-piece performance is necessary in order to secure the cutting templates to the jaw and to prevent undercuts.

Afterwards, the design results are also subject to finite element analysis to evaluate the implant's stability. We calculate using various extreme values due to the different strength values for the bone that are necessary for the computations. Finally, mechanical strength trends can be abstracted.
Figure 6: Cutting patterns, made by Direct Manufacturing, here in the model test carried out on a model of the lower jaw.

Figure 7: Implant was attached to jaw of pig cadaver, connecting type is based on a variant in which the implant is secured to the outer lower jaw contour.

2.3 Production planning and manufacturing of the implant

The implants are produced by means of LaserCUSING®, which is a generative technique based on amorphous material, such as powder. In this way, LaserCUSING® is able to produce functional models. The material characteristics obtained are commensurable with those of the series material and make it possible to use the parts thus produced even under the conditions of production. LaserCUSING® is a technology that works using a layer-by-layer technique, wherein layer thickness values vary from 30 µm to 50 µm.

Depending on the technology and the material, it is very difficult or simply impossible to produce surfaces lying under an angle of 45° to the building plate. For these surfaces, we need special supporting structures which have to be generated in the CAD system and later on by means of Magics®, which is a type of Rapid Prototyping software.

First, the implant is placed in the CAD system just as it is to be built in the LaserCUSING® system. Then it is shifted in Z direction by 0.5 mm so that is can be removed from the building platform by means of wire erosion later on. Thus it is possible to add the supporting structure to the overhanging surfaces. After the building procedure, this supporting structure has to be removed again. For this reason, this additional structure should be kept as small as possible in order to reduce necessary rework. Afterwards, the implant model is exported as an STL file. Magics® is used to generate the remaining supporting geometry. The parameters for this supporting structure have to be dimensioned and modified as a function of the shape type and position. As a function of the focus diameter, the supporting structure is only fused in the building process as a line structure. Consequently, it may be easily removed afterwards. After this step, these generated data are virtually cut into layers. The LaserCUSING® system is filled with pure titanium powder and fitted with a titanium plate intended to be use as a building plate. Manufacturing of the mandibular implant using pure titanium, is a technological challenge since it requires inert gas. In contrast to other body regions, however, for the oral and maxillofacial zones, pure titanium is preferred due to allergic reactions.

Next, the layer data are entered into the software of the machine, and the implant is positioned on a virtual building plate. The implant and the support structure are assigned the corresponding manufacturing parameters. Thereby, laser power and rate are defined, among other parameters. The entire process, from setup to removal, is performed in an inert gas atmosphere to guarantee manufacturing free of oxidation.

In the first step, the building plate is lowered down by one layer element, and new powder is introduced. In the next step, the powder is surfaced with a lamination plate (coating). In the last step, the deposited powder coating, which has a constant thickness, is selectively fused by laser (exposing). This procedure is repeated until the component is complete.

After completion of the building procedure, the building plate is removed, and the implant erodes from the plate. After that, the support geometry is removed and the implant is cleaned.

Building of the cutting patterns is performed analogously, with the difference that stainless steel, processed in a nitrogen atmosphere, is used as material.

Figure 8 elucidates examples for an implant made of titanium and the corresponding drilling patterns made of stainless steel to be secured to the bone.

3 SUMMARY AND OUTLOOK

The process chain introduced here shows the path from the CT image of a diseased patient via design of individual implants to the production of titanium implants by means of generative manufacturing techniques. This approach has been tested in eight lower jaws of pig cadavers, one human model lower jaw and two macerated human lower jaws up to now. Further operations were performed on the jaws of 10 living test animals (miniature pigs). So it was carried out, how the titanium implant is attached to the residual bone, the positions of the number of screws and the usability of the tools (Figure 9).
Figure 9: Titanium implant for outer attachment and cutting patterns to the residual bone of living animal

The results of these experiments show that we have succeeded in achieving a general fitting accuracy. At present, we are testing the process chain in animal experiments and are verifying the suitability of the implants in living beings.

It takes about 32 hours to carry out the entire process to produce individual implants. This span includes 7 working hours to prepare the CT data for the solid model of the jaw region. 13 hours are required to design the implant and the cutting patterns, while 12 hours are allotted for production planning and manufacturing.

In the future, the process may be optimised in the field of CT layered image processing. Process time should be positively influenced by segmentation and creation of the 3D model of the lower jaw. Adequate interpolation and filtering methods should contribute to higher data quality.

Design of implants and cutting patterns should be improved by the generation of new software tools. To keep lead times to a minimum, we are currently optimising the support structures and their process parameters for production planning and manufacturing. The coating system is also subject to continuous improvement.

Other recent research topics focus on the computer-aided modelling of inner filigree support structures that stimulate growth. In the future it is expected that such structures will be efficiently created in an automated manner using a CAD system. A stable attachment to the residual bone is essential to the function of the individual implant. The presentation outlines possible design variants. In design, strength, biocompatibility and operating conditions are to be considered. Future tests will determine the stability of these connections. The authors are currently developing a test bench for investigations of the jaw model. The test bed is also used to validate the results obtained in the FE analyses.

4 ACKNOWLEDGMENTS

We acknowledge Dipl.-Ing. Gerd Engel, managing director of the product development company Hofmann & Engel Produktentwicklung GmbH in Boxdorf near Dresden, for his innovative ideas and his forward-looking decision to invest in an advanced laserCUSING® system at an early date. We also acknowledge Dipl.-Ing. Thomas Jahn, project manager in this scientific project in the company. Gerd Engel also contributed to the distribution of this modern technology in an industrial environment – and now for applications in the medical domain as well. The research topic is sponsored by the Saxon Bank for Reconstruction and Development SAB Dresden (promotional ident:13363/2273).

5 REFERENCES


A CAD Based Tool for the Support of Modular Design

J. Pandremenos, G. Chryssolouris*
Laboratory for Manufacturing Systems & Automation, University of Patras, Rio, Patras 26500, Greece
xrisol@lms.mech.upatras.gr

Abstract
The objective of this paper is to present a novel CAD based software tool for the support of modular product design. The tool is developed in MS Visual Basic and works as an add-on to CATIAv5 CAD software. Its main functions are the automatic Design Structure Matrix (DSM) generation from the product's CAD model, the calculation of modularity performance indicators that assess the product's design architecture, the facilitation of the product's components clustering and finally, the visualization in a CAD form of a clustered DSM. A case study is presented in order for the application of the tool to be demonstrated and evaluated.

Keywords:
Design, DSM, modularity, clustering

1 INTRODUCTION
Today's market is characterised by customers with varied demands. In order for these advanced customer needs to be addressed, a large number of alternative products are produced by different companies, having as a consequence the blowup of competition. Furthermore, the tailoring of the product, according to the customer's needs, is being made. The development of modular design architectures is considered being the most efficient current practice for coping with this challenge. Currently, there are many efficient and successful design methods and methodologies capable of controlling a product's design architecture, performing the product's components clustering into chunks and generating product families. However, although efficient methods there exist, their implementation in industrial environments is most of the times difficult, since a supporting tool executing these tasks is missing.

The objective of this paper is to integrate some of the authors' developments, dealing with design architecture control and product design clustering, into a CAD based software tool in assisting designers to the generation of modular designs. The main functions of the tool are: (a) the DSM generation from the product's CAD, (b) the calculation of modularity indexes, (c) the facilitation of clustering and (d) the representation of a clustered DSM in CAD form. The tool is applied in a real design case study, stemming from the automotive industry, in order for its applicability to be investigated.

1.1 Methods for modular design
The tool presented in this paper, to a certain extent, was created for the facilitation of the methods for modular design, developed and published by the authors. These two methods are briefly described hereafter.

Modularity Performance
Pandremenos et al. [1] have developed an index for the quantification of a product's design architecture, at a part level. The index was called “Modularity Performance” (MP) and was given by the following equation:

\[ MP = 1 - \frac{I - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}} \]  (1)

Where \( I \) the number of interactions within a product's parts, \( I_{\text{max}} = (n^2 - n)/2 \) and \( I_{\text{min}} = n - 1 \), where \( n \) the number of parts. \( MP \) derives from the normalization of \( I_{\text{max}} \) and \( I_{\text{min}} \) values in order for the same boundaries, between 0 and 1 (\( I_{\text{max}} \) and \( I_{\text{min}} \) values depend on \( n \) and therefore, are not constant) to be always maintained. Obviously, when \( I = I_{\text{max}} \), \( MP = 0 \) (Integral architecture) and when \( I = I_{\text{min}} \), \( MP = 1 \) (Modular architecture).

This index could be employed by designers so as to assess a product's design architecture and thus, decide whether or not the clustering of this product is worthy. The closer the \( MP \) value would be to the boundaries (0 or 1) the less worthy would be this product's clustering.

Clustering algorithm
A novel clustering method utilizing Neural Network algorithms and Design Structure Matrices (DSMs) has been introduced in [2]. This method is capable of reorganizing a product's components in clusters, in order for the interactions to be maximized inside and minimized outside the clusters. Additionally, a multi-criteria decision making approach is used, in order for the efficiency of the different clustering alternatives, derived by the network, to be evaluated. In more details, the algorithm is based on Self Organizing Neural Networks (SONNs) trained with unsupervised competitive learning. The idea behind this method is to represent the interactions of each product's components, through the DSM, with vectors, which are afterwards inserted into the network for clustering. The implementation of this method has been performed in MATLAB.
2 LITERATURE REVIEW

Many of the existing software tools, supporting the design process, focus on the Axiomatic Design Theory (ADT) [3]. Over the last 20 years there have been numerous ADT industrial and academic applications. It was proven that in order to efficiently design large systems with ADT, high-level support tools, helping designers with developing complex products and large systems in a concurrent manner, are required [4]. Currently, there are software tools which: enable the designer to decompose Functional Requirements (FRs) and Design Parameters (DPs) in accordance with the rules of ADT, help him on the decision making process, assist on the requirements management and product hierarchies and evaluate the design or propose modifications.

Acclaro is a general purpose software tool, which makes use of ADT. This software enables the designer to enter the top-level FRs and DPs and decompose and map those in two tree hierarchies and associated Design Matrices (DMs). The software's ultimate output is a design that satisfies the FRs and Constraints (Cs). The designer provides inputs such as FRs, DPs and Process Variables (PVs) to the software and afterwards, he answers questions on the relations among these characteristic vectors, prompted by the software. Based on these inputs, Acclaro creates DMs and helps the designer to make correct design decisions by performing partitioning, clustering and tearing [5-6].

SLATE, part of the EDS Team Center suite of the software, is a tool that performs requirements management. It helps the designer to construct requirements and product hierarchies. Moreover, allows the designer to attach constraints and text fragments within each layer of the system's decomposition and also provides a good level of traceability of non-FRs throughout the design decomposition. Despite the assistance provided by the software, the results rely on the user's experience and efficiency [7-9].

Manager’s Aid for Intelligent Decomposition (DeMAID), developed by NASA, is a software tool, which adds the designer to make decisions that take advantage of decomposition, concurrent engineering and parallel processing techniques. The original version of DeMAID was released to the public in 1989. This version is a knowledge-based software tool for minimizing the feedback couplings, sequencing the design processes, grouping processes into iterative sub cycles and displaying the sequence of processes in a DSM format. The major enhancement of DeMAID/GA is the addition of a Genetic Algorithm (GA), which sequences the processes within each iterative sub cycle to minimize the time and cost to converge upon the design solution [10].

Design Evaluation Machine (DEM) is a software tool that helps the designer to select an optimal design satisfying the FRs. At first, the DM is used for checking any violations of the independence axiom. In order to apply the information axiom, the software makes use of an evaluation method that measures the information. This method adds intensity weight to the measuring process solving the problem, when the information is the same [11].

The benefits of integrating both the principles of AD and DSM, brings out the Computer-aided Conceptual Design System. The system after requesting the FR definition and through a design matrix, it follows the repository of solution principles, and performs the decomposition. DSM is generated in order to deal with physical integration, decomposition into elemental DPs or components and finds any undesired interactions between them. A re-engineered DSM (RDSM) provides feedback to the construction of a new DM (RDM) and finally to the desired DSM. The last DSM provides guidance to the designer to build the real DSM [12]. Furthermore, a lot of tools for the clustering of a product's components into modules are available. The majority of them are DSM based and utilize GAs [13-16], however, there are some others employing Artificial Neural Networks (ANNs) [2], King's and the Modified Minimum Degree algorithms [17] as well as k-means for partitive clustering [18]. Additionally, Lattix Inc's Dependency Manager (LDM) was the first commercially available implementation of the DSM analysis for software, offering algorithms that may perform partitioning of DSMs, a task that is quite similar to clustering [19].

2.1 Outcome

Most of the tools found for the support of modular design, are based on ADT and thus, assist the design only from the functions-to-parts mapping aspect of the design architecture. No tool has been discovered, to be helping a designer to develop modular products in terms of the interactions-between-parts point of view. Furthermore, several tools are currently available for the clustering of a product's components with the use of DSM and a clustering algorithm, however, none of them can automatically generate the DSM of a product. Finally, there is no existence of a tool capable of representing a clustered DSM in CAD form.

3 DESCRIPTION OF THE TOOL

The main idea behind the development of this tool is the facilitation of the product design process for the generation of modular architectures. The concept of the tool is shown in the flowchart of Figure 1. The designer firstly generates the DSM of the product he is developing, from its own CAD. Afterwards, he evaluates his design for its modularity by using the MP index and he decides whether the design needs to be clustered or not. If the product needs clustering, he imports the previously generated DSM into the clustering algorithm, from which he receives a new clustered DSM that is finally imported back to the CAD tool in order to visualize the clusters derived. More precisely, the tasks that the tool can perform, towards the design process assistance, are the followings:

- Automatic generation of the DSM from a product's CAD model, by using the constraints defined among the parts in the Product tree
- Calculation of the $l_{max}$, $l_{min}$ and $MP$ indexes for the assessment of the product's architecture (at part level) as well as investigation of the clustering necessity and efficiency
- Visualization in CAD form of a clustered product, by importing the clustered DSM into the CAD software in Excel format
- Facilitation of the DSM clustering using ANNs with MATLAB through the compatibility of the latter with the developed tool

155
The tool was developed within the environment of CATIAv5 CAD package. The Visual Basic programming language for CATIA was employed for the creation of its different functions. Moreover, a new toolbar, named DSM tools, was created by having incorporated all the tools developed (Figure 2).

Figure 1: CAD tool's concept.

3.1 DSM generation
The manual DSM generation for products, consisting of many parts, is considered being a very time consuming and effortful process for a designer. This is one of the most important reasons that many DSM based design methods failed to be utilized in real industrial applications. With the tool developed, the production of such large DSMs is carried out in only a few seconds, both automatically and very easily.

In CATIA, all the information about the design of a product is stored in the Product tree. Such information comprise the parts of the product, the interfaces between them, the design steps and functions utilized for the creation of each part etc. In order for the product's DSM to be generated, the information regarding the parts and interfaces is extracted from the tree, with the help of Visual Basic. Then, based on this information, the DSM is created and stored in an Excel sheet of the PC's predefined folder.

3.2 Calculation of indexes
The tool developed is capable of calculating the \(I_{\text{max}}\), \(I_{\text{min}}\), and \(MP\) indexes. This is accomplished by taking into consideration the interactions of the product's parts, in a manner, described in the previous section (information from product tree etc.). Through this facility, the designer can control the architecture of the product under design, in real time. Furthermore, he can easily assess whether his design needs to be clustered as well as how efficient the clustering will be.

3.3 CAD representation of clustered DSM
With this function, a clustered DSM, in Excel format, is imported in CATIA and therefore, the product's clusters can be easily illustrated. This is accomplished by altering the structure of the product tree. Figure 3 demonstrates that change. The tree in its initial form, has two main hierarchical levels: the product and the parts. On the other hand, the transformed tree of a clustered product has three main levels: those of the product, the clusters and the parts. Thus, the designer may hide, set apart or color the different clusters very easily, in order to visualize and get familiar with them.

3.4 Clustering facilitation
The CAD based tool was structured to be compatible with MATLAB and therefore, the execution of the ANN algorithm, developed by the authors, is considerably facilitated. The compatibility lies to importing in MATLAB the DSM, produced by the tool, as well as to importing back in the tool, the DSM exported by the clustering algorithm in MATLAB.

Figure 3: Product tree of a clustered product.
4 CASE STUDY

4.1 DSM formulation of a car's BiW

In this case study, the DSM is generated from the design of a "state of the art" car BiW. The design consists of 38 main components (\(n\)) having 108 interactions (\(I\)) among them. The interactions in this case, are meant as interfaces. The DSM was created and stored in an Excel sheet with the help of the tool, in just a few seconds. In Figure 4 the BiW design and the DSM are illustrated.

4.2 Car's BiW design architecture assessment

The DSM of the car BiW was afterwards assessed with the method described in Section 1.1. The CAD based tool was once more employed for the calculation of these indexes (Figure 5). As it can be noticed, the calculation of the MP revealed a fairly modular design architecture of this product (0.89 out of 1).

4.3 Visualization of clustered car's BiW DSM

In order to demonstrate the tool's clustered DSM visualization facility, the BiW's DSM Excel file was imported into MATLAB, clustered, using the algorithm described in Section 1.1 and exported back to a new Excel file. This file was finally imported into CATIA for the visualization of the clusters. The different clusters derived are shown with different colors in the "exploded" view as well as in the product tree of Figure 6.

5 CONCLUSIONS

The CAD based tool introduced, came to fill an existing gap in the control of product design architecture from the interactions-between-parts point of view. Through this tool, a designer may automatically generate a product's DSM from its CAD, reducing significantly the effort required for this task (especially for products with a large number of components). Additionally, he is able to quickly assess his designs in terms of their architecture (through the calculation of the MP) and to be facilitated in the execution of the clustering process. The application of the tool to a...
real case study, that of the BiW of a commercial vehicle, has demonstrated its capabilities and shown its efficiency to handle quickly products consisting of many parts. Furthermore, since the tool was developed as an add-on to a commercial CAD software (CATIAv5) and at the same time, is user-friendly, it can be very easily employed by the industry.

As a future work, the authors' intention is to create a second generation version of the tool, integrating the ANN clustering algorithm, through the Visual Basic programming language. Thus, the clustering of the product's components will become more direct. Additionally, the tool will be tested for its efficiency, on products with even a larger number of components.

6 ACKNOWLEDGMENT

The work was partially supported by the EU funded project: FP6-2004-NMP-NI-4-026621, Multi-Functional maTerials and related prodUction technologies integRated into the Automotive industry of the future –FUTURA”.

7 REFERENCES


Modeling the Structure and Complexity of Engineering Routine Design Problems

J. M. Jauregui-Becker, W. W. Wits, F. J. A. M. Van Houten
Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands
j.m.jaureguibecker@ctw.utwente.nl

Abstract
This paper proposes a model to structure routine design problems as well as a model of its design complexity. The idea is that having a proper model of the structure of such problems enables understanding its complexity, and likewise, a proper understanding of its complexity enables the development of systematic approaches to solve them. The end goal is to develop computer systems capable of taking over routine design tasks based on generic and systematic solving approaches. It is proposed to structure routine design in three main states: problem class, problem instance, and problem solution. Design complexity is related to the degree of uncertainty in knowing how to move a design problem from one state to another. Axiomatic Design Theory is used as reference for understanding complexity in routine design.

Keywords:
Routine design, Complexity, Axiomatic Design Theory.

1 INTRODUCTION
Advances in technology and competitive markets are driving the development of products with shorter times to market. In addition to this, products are becoming more complex, with more functionality, and yet lower prices. To cope with this trend, engineers often redesign existing products to meet new requirements and optimize performances based on a steady concept. This type of design problems can be regarded as routine.

Although routine design occurs within a well defined domain of knowledge, such problems may be of very complex natures. Consider for example the design of injection molds. The first injection molds were designed and developed in 1868 by John Wesley Hyatt, who injected hot cellulose into a mold for producing billiard balls [1]. Much later, in 1946, James Hendry built the first screw injection molding machine, giving birth to the machines and processes we know nowadays. Since then, much knowledge on injection mold design has emerged and been formalized in books (e.g. [1] and [2]), expert systems (e.g. [3] and [4]) and the Internet. However, given the amount of components, physical phenomena and processes involved, the design of injection molds is still considered a complex task.

If such problems are modeled following the pyramid model of Gerrit Muller [5], the result is a large hierarchical multi-layered network of design functions, components and variables. As one might envision, determining strategies and methods for solving such complex problems is not a trivial task. Furthermore, if one considers that around 80% of design in industry is routine, it can be concluded that counting with a proper understanding of its complexity is a relevant topic in the field of design theory and methodology.

Having the previous as motivation, this paper presents a model to structure and determine the complexity of routine design problems. The future goal of the research this paper forms part of, is to effectively and systematically manage design complexity as a means for automating the generation of candidate solutions to complex routine design problems. The concepts of design complexity as stated in Axiomatic Design Theory (ADT) are used to assemble a specific model complexity for routine design.

The remainder of this paper is organized as follows:
• Section 2 present a classification of design problems and defines what routine design is.
• Section 3 describes the rationales of the synthesis process in routine design.
• Section 4 presents a new model for structuring routine design problems.
• Section 5 describes the main concepts stated in Axiomatic Design Theory.
• Section 6 presents a model of complexity in routine design by mapping ADT complexity theory onto the structuring framework presented in Section 4.
• Section 7 presents some general guidelines for dealing with design complexity.

To finish, Section 8 presents a discussion and final remarks.

2 DESIGN PROBLEM CLASSIFICATION
The scope of this paper lies within the field of artifactual engineering routine design. This section explains what is meant by this at the hand of the FBS [6] model.

2.1 FBS model
FBS models a design artifact by distinguishing the following levels of object representation: Function, Behavior/State and Structure, as shown in Figure 1. The basis of the FBS model is that the transition from function to structure is performed via the synthesis of physical behaviors.
Therefore, behaviors allow characterizing the implementation of a function. As many different views of the FBS model have been developed and researched, this paper adopts the unified FBPSS model presented by Zhang et al [6]. This model is based on the analysis and generalization of the Japanese [7] [8], European [9], American [10] and Australian [11] schools of design modeling.

The FBPSS model uses the following definitions:

- **Structure**: Is a set of entities and relations among entities connected in a meaningful way. Entities are perceived in the form of their attributes when the system is in operation. For example, in Figure 1 the Structure is represented by an electric motor and a crank mechanism. Here, the two possible entities (structures) are the lengths of the bars $L_1$ and $L_2$.

- **States**: Are quantities (numerical or categorical) of the Behavioral domain (e.g. heat transfer, fluid dynamics, psychology). States change with respect to time, implying the dynamics of the system. For example, in Figure 1, the states of the structure are represented by the distance $L_2$ between the electric motor and the piston, the torque $T_F$ of the electric motor, or the displacement of the piston $s$.

- **Principle**: Is the fundamental law that allows the development of a quantitative relation of the States variables. It governs Behavior as the relationships among a set of State variables. For the example in Figure 1, two possible principles are electromagnetism ruling the operation of the electric motor, and solid mechanics ruling the function of the crank mechanism.

- **Behavior**: Represents the response of the structure when it receives stimuli. Since the Structure is represented by States and Structure variables, Behaviors are quantified by the values of these variables. In the case presented in Figure 1, the two Behaviors are Generate torque and Convert torque into force.

- **Function**: It is about the usefulness of a system. For example, in Figure 1, one possible function of this system is to compress gas.

---

### 3 SYNTHESIS IN ROUTINE DESIGN

#### 3.1 Modeling Design Artifacts

Artifacts, e.g. an injection mold, can be modeled as a hierarchical multi-layered network of interconnected components and parameters that resemble the structure of the pyramid of Gerrit Muller [5], as shown in Figure 3(a). In this model, the top layers represent functional requirements, the in-between levels represent components, and the lower levels represent design parameters of these components. Functional requirements specify the characteristics of an artifact’s function, as for example the power of an electric engine. Furthermore, in this model components are composed of networks of other sub-components, and so forth. For example, sliders in injection molds are composed of mechanical linkages, which are simultaneously composed of rigid links and joints. It is characteristic to complex artifacts to have a large number of interconnected networks of components, as well as a large number of parameters, relations and constraints.

#### 3.2 Modeling Design Problems

An artifactual design problem can be modeled as an incomplete description of an artifact, as it is shown in Figure 3(b). The descriptions known beforehand are regarded as the design requirements, and these must be satisfied by candidate solutions. Design requirements can be functional requirements, components, parameter values, or combinations thereof. Creative, innovative, and routine design problems can be represented using this
model, as the differences among them reside in the type of knowledge available for generating candidate solutions. In routine design there is knowledge available about:
• the types of components that can be used to generate candidate solutions,
• how components can be connected with each others, parametric descriptions of each component, and,
• relations and constraints that relate parameters and components to functional requirements.

Furthermore, designing one type of artifact can be the subject of different types of problems, as several combinations of design requirements can be formulated.

3.3 Synthesis in Routine Design
As Figure 4 indicates, moving from an incomplete representation to a complete description is done by a synthesis process. Synthesis processes in routine design are performed by two types of tasks: (1) generating networks of components and (2) attributing values to unknown parameters. The exact strategy determining how these tasks are performed depends on the distribution of requirements throughout the different levels of detail of the problem, e.g. only on top, only at the bottom or as a mix. However, this relation is not known a priori and is different for different distributions of design requirements. As it will be shown in Section 6, complexity in routine design relates to the uncertainty in determining this relation.

4 THE STRUCTURE OF ROUTINE DESIGN
In design, structure and complexity are two closely related concepts. Complexity is a property related to the degree of difficulty or uncertainty for finding a solution to a design problem [12], whereas structure is a property related to the organization of the variables and relations describing the design problem itself [13]. In order to analyze design complexity, it is necessary to understand the structure of the problem. The structure of a design problem has three important aspects to be studied:
• The consistency of the design variables: all design variables have to be related to each other’s by relations that are ultimately used to determine the performances of the problem.
• The distribution of design parameters along the problem model: all parameters concentrated in one element vs. several parameters scattered through several elements; one vs. several levels of detail.
• The relation between what is known (design requirements) and what is unknown (unattributed structure variables): scattered along the problem model, concentrated in problem chunks, at the top (only functional requirements), at the bottom (only design parameters) or as mix of all these possibilities.

In order to develop a standard model of complexity in routine design, a standard way of structuring such problems is first required. Therefore, this section describes a framework that has been developed to structure design problems. The framework is inspired in the structure of analysis problems. The analysis of design complexity to be presented in Section 6 is based on this framework.

4.1 Structure in Analysis
Physicists model natural phenomena through differential and integral equations. Specific problems are solved by setting boundary conditions on the differential and integral equations, and applying solving procedures to obtain analytical expressions. The resulting expression can then be used to calculate values of variables by specifying the values of the input parameters. Consider for example the law of heat conduction shown in Equation 1. This differential equation models the phenomena of heat transfer through matter from a region of high temperature to a region of a low temperature. As this is done independent of geometry, material properties or temperature distributions, the equation is generic. To model a specific case of heat transfer the equation is rearranged by introducing boundary conditions, canceling unnecessary terms and performing mathematical manipulations. For example, heat conduction in one dimension between two flat plates results, after rearranging Equation 1, in Equation 2. The obtained equation can now be used to introduce known values and calculate the values of the required parameters, as for example in equation (3) the time required to get temperature \( \psi = T \) at a point \( \zeta = x \) is \( \xi = t \).
\[ \frac{\delta T}{\delta t} = a \frac{\delta^2 T}{\delta x^2} \]  
(1)

\[ f(x, T) = \frac{x^2}{\pi^2 \alpha} \cdot \ln \left( \frac{8}{\pi^2} \cdot \frac{T - T_w}{T_{fo} - T_w} \right) \]  
(2)

\[ f(x - \zeta, T - \psi) = -\zeta \]  
(3)

If one would generalize this structure, one would notice that physicists model natural phenomena at the hand of tree phases which are solved by two types of procedures, as shown in Figure 5. On the one hand, the three phases are: the differential/integral equation, the analytical expression, and the solution. On the other hand, the procedures are: differential and integral calculus to transform the differential equation (phase 1) into an analytical expression (phase 2), and algebra or numeric methods to transform the analytical expression (phase 2) into values of unknown parameters (phase 3).

Figure 5: Structure of problems in modeling natural phenomena

From a research perspective, this problem structure has allowed:

- defining the types of representations required for modeling each domain (i.e. operators),
- studying the complexity of each domain by analyzing the configuration of the used representations (i.e. equation order, equation linearity),
- developing procedures and operations to solve each problem domain (i.e. Laplace Transformations, Newton-Raphson method)
- identifying common problem structures (i.e. Poisson's equation, Laplace equation)

On the other hand, some of the major advantages from an application perspective are:

- reuse of existing problem formulation,
- utilization of standard solving methods,
- development of computer based simulation tools.

Structuring design problems as done in natural phenomena is likely to drive the automation of design problems toward more generic approaches, with advantages in both research and application. For this research, such a structure would allow the identification of features causing complexity, and do so independent from the problem semantics. Furthermore, strategies for managing design complexity can be formulated as function of problem structures.

4.2 Definitions

The structure here presented is based on the definitions presented in [14], where the different types of information contents and models used in routine design are presented. The definitions are:

- **Element**: is a class description of a component.
- **Descriptions**: characterize an element class by representing its attributes in the form of variables.
- **Relations**: are considered class descriptions and are of different types, namely, topology, physical coherence, design rules, analysis relations and objective functions [14]. Their descriptions can be declared within the scope of the relation or by pointing towards descriptions of embodiment and scenario elements.
- **Cardinality**: Is a parameter that models the number of elements in a design solution. Its value can be unknown, known or determined by an algebraic relation.
- **Embodiment**: is the subset of descriptions of an element upon which instances are created to generate design solutions.
- **Scenario**: is the subset of environment variables, attributed to elements in the natural world and considered in measuring a design artifact’s ability to accomplish its function.
- **Performances**: are descriptions used to express and assess the artifact’s behavior.
- **Analysis relations**: use known theories (e.g. the laws of physics or economics) to model the interaction of the design artifact with its environment and predict its behavior. Determine the performances.
- **Topology relations**: define the configuration of embodiment and scenario elements by means of relations expressing belonging and connectedness.
- **Objective function**: weighs and adds the performances into one general indicator.

4.3 Structuring Framework

It is proposed to structure routine design problems at three different states: problem class, problem instance and problem solution. This is shown in Figure 6. Problem classes are transformed into problem instances by specifying its requirements. Requirements are specified by instantiating descriptions. Problem instances are transformed into problem solutions by algorithms that generate instances to the unknown descriptions. Therefore, one problem class can represent many problem instances, and one problem instance can have many problem solutions, as indicated in Figure 7. Under this view, solving routine design is analogues to solving problems with known differential equations.

Figure 6: Framework to structure design problems

This structure can also be used to structure innovative and creative design problems. Solving innovative design problems is analogous to combining different differential equations to model various interrelated physical phenomena. Solving creative design problems is analogous to developing new differential equations. However, these two types of design problems are outside the scope of this paper.

**Problem Class**

A problem class is structured in:

- **Elements**: are considered class descriptions, and are used to represent both, embodiment and scenario elements. Elements can also be differentiated by assessing the functions and types of descriptors modeling a component.
- **Relations**: are considered class descriptions and are of different types, namely, topology, physical coherence, design rules, analysis relations and objective functions [14]. Their descriptions can be declared within the scope of the relation or by pointing towards descriptions of embodiment and scenario elements.
• Descriptions: are variables that characterize elements and relations by mathematic models. These can also be of different types: parameters, shapes, fields, topology and spatial, as described in [14].

**Problem Instance**

A problem instance is structured by:
- Instantiated scenario: represent scenario specifications,
- Partially instantiated embodiment elements or parameters: represent embodiment requirements, and impose constraints to the space of possible solutions.
- Partially instantiated performance parameters: represent the performance specifications the embodiment has to meet.

**Problem Solution**

The problem solution consists of fully instantiated elements, relations and parameters. For under-constrained problem-instances, many solutions may exist. This depends on how constrained the problem is. An under constrained will allow for multiple solutions, while a systems of equation type of problem will have a limited number of solutions.

4.4 Example: Spring Design

Consider for example the spring design formulation shown in Figure 8. Figure 8(b) shows that the problem class corresponds to the declaration of the different types of parameters of the problem formulation. By setting requirements, the problem class is transformed into one (1) problem instance. However, by setting requirements to other parameters, other problem instances can be obtained. The figure shows one possible problem solution. Nevertheless, for this problem instance, many possible problem solutions are possible.

5 COMPLEXITY IN AXIOMATIC DESIGN THEORY

As stated in [15], the complexity of systems has been studied from two different perspectives: the physical domain and the functional domain. On the one hand, complexity in the physical domain is seen as an inherent characteristic of physical things, including algorithms and products. According to this view, systems with many parts are more complex than those with less. Examples of such studies are computational complexity [16] and complex emergent systems [17]. On the other hand, complexity in the functional domain is seen as a relative concept that evaluates how well we can satisfy “what we want to achieve” with “what actually is achievable”. From this perspective, axiomatic design theory [15], incompleteness of information [18] and multi-disciplinary complexity [19] are some of the different models to understand the complexity of systems.

This paper is based on the notions of design complexity stated in Axiomatic Design Theory (ADT) [15]. The basic idea of this model is that without difficulty in understanding (or making, operating, etc.), a system is not complex. In this sense, complexity is the property of a system that makes it difficult to understand with the available knowledge about its constituents parts. Tomiyama further elaborates this view, by stating that complexity can be studied from the view point of knowledge structure [19], identifying two types of complexity: complexity by design and intrinsic complexity of multi-disciplinarily. The former is attributed to the structure of the design problem, while the latter deals with behavioral characteristics.

This paper adopts the ADT definition of complexity, and focuses on complexity by design. The three main reasons why this model has been chosen as framework in this research are: (1) complexity is regarded as a relative property, (2) complexity is the consequence of engineering activities, and (3) it is assumed that complexity can be managed. The ADT model of complexity is used to identify different types of complexities in routine design. This section presents a summary of ADT’s model of design complexity.

5.1 Axiomatic Design Theory

ADT is based on the hypothesis that there are fundamental principles that govern good designs [20]. Its two founding axioms are: (1) maintain the independence of the
Functional Requirements (FRs) and (2) minimize the information of the Design Parameters (DPs).

FRs are the set of requirements that characterize the needs of the artifact in the functional domain, while DPs are the variables that characterize the design in the physical domain. The relation between the FRs and the DPs is represented in equation form as:

$$[FR] = [A][DP]$$  \hspace{1cm} (4)

where $A$ is the Design Matrix (DM) of the problem.

Depending on the DM, a design can be coupled, decoupled or uncoupled. Consider for example a problem with two FRs and two DPs. When the design is coupled, the FRs cannot be satisfied independently because of the interdependence of both DPs, as shown in Equation 5. In a decoupled design, shown in Equation 6, the DPs have to be solved in a particular order so that FRs are achieved. In uncoupled designs (Equation 7), the FRs are independent from each other, and no particular order is required for solving the DPs.

<table>
<thead>
<tr>
<th>Coupled</th>
<th>Decoupled</th>
<th>Uncoupled</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[FR]_1 = [x \ x][DP_1]$</td>
<td>$[FR]_1 = [x \ 0][DP_1]$</td>
<td>$[FR]_1 = [x \ 0][DP_1]$</td>
</tr>
<tr>
<td>$[FR]_2 = [x \ x][DP_2]$</td>
<td>$[FR]_2 = [x \ x][DP_2]$</td>
<td>$[FR]_2 = [0 \ x][DP_2]$</td>
</tr>
</tbody>
</table>

$$[FR]_1 = [x \ 0][DP_1], \quad [FR]_2 = [0 \ x][DP_2]$$  \hspace{1cm} (5)

$$[FR]_1 = [x \ 0][DP_1], \quad [FR]_2 = [x \ 0][DP_2]$$  \hspace{1cm} (6)

$$[FR]_1 = [x \ 0][DP_1], \quad [FR]_2 = [0 \ x][DP_2]$$  \hspace{1cm} (7)

5.2 Complexity model

In ADT, complexity is defined as “the measure of uncertainty in achieving the functional requirements of a system within their specified design range”. When the range of a system changes as a function of time, it is regarded as a system with time-dependent complexity. When the range does not change as a function of time, it has a time-independent complexity. “Time” is used in a general sense, signifying the progression of “events”.

Time-independent complexity is classified into time-independent real complexity and time-independent imaginary complexity. The former is a consequence of the system range not being inside the design range. The latter occurs when there are many FRs and the design is a decoupled design. It is called imaginary because this corresponds to a situation in which the different orders in solving the design matrix have different attributed levels of difficulty. A system with imaginary complexity can satisfy the FRs at all times if we vary DPs in the right order.

Time-dependent complexity is the uncertainty caused by the increase or decrease of the number and types of DPs during the design process itself. ADT classifies time-dependent complexity into combinatorial and periodic complexity. Design problems with combinatorial complexity experience a continued growth of their DPs in time. For example, constructing a sentence by the combination of words has combinatorial complexity. As the number of words (the DPs in this case) increases, keeping semantic and syntactic consistency among them becomes more difficult. On the other hand, periodic complexity is the case in which the increase of parameters is restarted after a period of time (or succession of actions). An example described in [19] is air traffic control. Air traffic in large airports follows a wave pattern that depends on the time of the day. When the traffic is at its peak, air controllers deal with very complex situations. However, at low traffic times their task becomes significantly simpler.

ADT suggests three main strategies for managing design complexity: (1) minimize the number of FRs, (2) eliminate time-independent real complexity and time-independent imaginary complexity, and (3) transform a system with time-dependent combinatorial complexity into a system with time-dependent periodic complexity. This paper adopts this model of design complexity, and explores its characteristics for routine design problems.

5.3 Translating ADT terminology

ADT terminology is translated into the design structuring framework (see Section 4.2) as follows:

- **Functional Requirements (FRs):** correspond to the functions, performances and scenario descriptions as described in Section 4.
- **Design Parameters (DPs):** correspond to the embodiment descriptions in the routine design formulation, and model elements and parameters.
- **Design Matrix (DM):** is formed by the analysis, topology and physical coherence constraints.

6 MODEL OF COMPLEXITY IN ROUTINE DESIGN

The model of complexity obtained in this research is the result of mapping the ADT complexity model presented in Section 5 onto the structuring framework presented in Section 4. The result is a set of complexity types in problem classes and another for problem instances. Complexity of problem classes deals with incorrect problem formulations, while complexity in problem instances deals with deriving strategies for solving it.

6.1 Complexity of Problem Classes

Time-independent complexity captures the complexity of a system in which the time dimension does not limit its ability to achieve its functional requirements. In other words, the range of the system does not change over time. In routine design, the process of moving from problem classes to problem instances is not a synthesis process by itself. This is rather a human process that involves specifying which are the parameters and elements characterizing the input of the problem. Therefore, complexity here is related to how well the problem has been formulated, and regards two types of uncertainties. One is the uncertainty of having all of the required information in the problem formulation. The second is the lack of differentiation between problem chunks and its interrelations. Figure 9 illustrates this idea by showing an inconsistent and unstructured problem in Figure 9(a) and a consistent and structure one in Figure 9(b).

6.2 Complexity of Problem Instances

**Time-independent Real Complexity**

Real complexity appears in multi-objective problems, where one parameter has to satisfy contradicting objectives. This is caused when several disciplines determine an artifact’s behavior. For example, the more lanes a highway has, the more traffic it can accommodate. At the same time, the number of lanes increases, the number of accidents also increases. Designing highways with the objectives of traffic maximization and accidents minimization has a contradiction, and therefore is a problem with time-independent real complexity.

**Time-independent Imaginary Complexity**

Imaginary complexity originates from the fact that design requirements are set at different combinations of parameters and elements. As consequence, it is not known a priori in which order the problem will be solved. Furthermore, when the DM is decoupled, a particular order is required for solving the problem. Imaginary complexity
depends on the relations between known and unknown cardinalities as well as on the relation between instantiated and non-instantiated elements and parameters. The former is regarded in this work as knowledge distribution, while the latter is regarded as requirements distribution.

From a knowledge distribution viewpoint, the more cardinalities that are known in a problem instance, the lower its uncertainty is regarding the number of elements to instantiate. Moreover, as the cardinalities of elements are often interrelated among each other, modifying one automatically leads to the modification of others. In this sense, knowledge distribution refers to the distribution of known cardinalities among the elements of the problem.

Requirements distribution refers to the distribution of instantiated and non-instantiated elements and parameters. For elements, complexity regards the uncertainty of having to apply bottom-up or top-down approaches, while for parameters it regards the uncertainty of knowing in which order to solve the constrained system.

**Time-dependent Combinatorial Complexity**

For problem instances, combinatorial complexity occurs when the design problem consists in generating complex topologies or shapes in which embodiment elements are instantiated several times within one design solution (for example, the number of gears required in a gear box). This results in a DM with time-dependent varying size and terms. When no knowledge is available about the number of instances required to satisfy the FRs, the problem presents time-dependent combinatorial complexity. One way of recognizing this type of complexity is by assessing the cardinalities of the elements in the problem formulation. When the ranges of cardinalities are not known, or cannot be written as function of other parameters, the design problem has combinatorial complexity.

### 7 COMPLEXITY MANAGEMENT

#### 7.1 Problem classes

Ignorance of FRs and DPs is related to the failure to properly understand them in a design task. This is caused by a faulty or incomplete description of the functions, performances and scenarios in the problem formulation. As a result, one would be addressing a wrong problem. Complexity emerging from incomplete problem formulations is time-dependent imaginary complexity. In ADT, imaginary complexity is defined as the uncertainty that arises because of the designer's lack of knowledge and understanding of a specific design itself. In order to solve this, the FRs of the problem have to be identified and related to the problem's DPs. Then, the problem can be reformulated in terms of the emerging relations between FRs and DPs. By doing so, the imaginary component of the complexity can be managed. Two methods can be used to manage such complexities:

- **FBS based formulation**: Described in [21], this method aids the exploration of the information required to formulate a given design problem. The goal of the method is to obtain a consistent mapping between a problem function, its behaviors, and the parameters and relation in its formulation.

- **ADT based decomposition**: Described in [22], this method deals with the decomposition of the problem into smaller problem chunks. The goal is to distribute the information of the problem among different abstraction levels.

#### 7.2 Problem instances

Time-independent real complexity can be managed by using constraint satisfaction and optimization techniques. In the field of Multidisciplinary Design Optimization [23] several techniques have been developed to cope with such problems, as for example Collaborative Optimization (CO), Bi-Level Integrated System Synthesis (BLISS) and Analytic Target Cascading (ATC).

Time-independent imaginary complexity can be managed by determining strategies that establish in which order to solve the design problem. Notable techniques are based on Design Structuring Matrix manipulations [24].

Time-dependent combinatorial complexity can be managed by making parametric models of the topology of the system. The field of Computational Design Synthesis (CDS) has developed grammar based approaches that can be used to cope with this type of complexity [25].

#### 7.3 Integration challenge

Having this overview, one can say that the real challenge lies in the development of a framework to integrate these techniques. Such a framework requires representations that can be used for modeling design knowledge (elements, parameters, and relations) in a generic fashion while supporting the utilization of aforementioned techniques simultaneously. Furthermore, such representations should allow the development of methods for determining solving strategies as a function of the organization of the building blocks used to represent a design problem and permit the independent modeling of abstraction levels.

### 8 DISCUSSION

In general terms, solving routine design can be performed in two different fashions: developing specific methods for specific problems, or developing generic methods for problem families. It is the second approach that can enable the future development of software that automates this process, and where the focus of this paper lies. Such an approach is confronted with three main challenges. Firstly, design problems have to be translated into terms that allow their study independent of its semantic contents, thus to define a common language. Secondly, basic problem
characteristics have to be identified which consists of finding the common ground of these problems. And thirdly, general problem solving approaches have to be developed.

The first challenge was researched by the authors and presented in [14]. The second challenge is what this paper focuses on: investigating different types of complexity in design. Its result is a generic description of the basic components of design complexity that can be found in routine design. Further research is required to develop generic methods for managing these complexities, which is the third challenge. It is expected that these approaches will lead to robust standard methods that are capable of automating the generation of candidate solutions to routine design problems.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Innovation-Oriented Research Programme 'Integral Product Creation and Realization (IOP IPCR)' of the Netherlands Ministry of Economic Affairs, Agriculture and Innovation.

REFERENCES

Lead Time Reduction and Efficiency Enhancement Show Strong Interference with Customer Constraints in Banking Service Process Design

M. Nakao¹, M. Okano² and D. Dvivedi²
¹ The University of Tokyo, Dept. of Engineering Synthesis, Faculty of Engineering, Tokyo, Japan
² Shinsei Bank, Ltd., Japan

Abstract
This paper reports our results of applying engineering process design to the banking service of housing loan credit approval. Process design of manufacturing digitalizes the entire process and stacks independent judgments one by one; and process design for banking service has the same nature. A major difference between the two is that lead time reduction and work efficiency enhancement have strong interference with customer constraints in banking service, and weak in manufacturing. The interference is hard to eliminate because we cannot force the customer to always fill out the application forms in a complete manner.

Keywords:
Scheduling, Service, Bank

1 INTRODUCTION
Service is a process that satisfies the customer expectation. The service provider should clarify what the customer expectation is before starting the process, and at the same time should confirm the customer constraints and then start finding the optimum design solution process for the customer. In real service, however, the process starts before the customer constraints are clarified and towards the end, the service provider adjusts the design solution to offer the best satisfaction to the customer. The solution often turns out to be a compromise between the customer and the service provider. A good example is a housewife shopping for groceries to prepare dinner. In this case, even if she had prepared a list of items to shop for the expected dinner, depending on what a good buy was at the store on that day, she would change the expectation drastically by changing the original constraints of freshness or cost that she had.

In modern manufacturing processes, functional requirements and constraints are set before production starts and remain unchanged until the production at the first run ends. Information technologies these days have tools for process management such as axiomatic design theory [1] and once an order is confirmed, obtaining material to manufacturing processes are determined immediately to the optimum conditions. For example, even with a single part production with injection metal mold [2-3], once the manufacturer has the drawing of the final plastic product the customer wants, he can use 3D CAD to invert the part to determine the mold shape, apply CAM to generate tool paths, and cut the material with NC machining to produce the part as originally designed. In this series of processes, a decision is made based on the independence axiom [1], no interference should take place, for example, fitting two parts that are machined separately without any adjustment.

One of the authors Nakao has reduced the lead time of a metal mold [2] by 86% from 352 hours down to 49.8 hours by dividing up the mold production into 583 digitalized and independent sub-processes, and at the same time replaced operator judgments with machine measurements and CAD calculations to limit the number of human judgments. Of the 583 sub-processes only 17% of 77 remained. Allowing judgment by operators should be avoided [3] because it will introduce exceptions. The result is a large number of options to complicate the processes like a coupled design and this will eventually cause a long lead time.

So far, many engineering researchers have reported production scheduling problems. The approach should include social aspects, not only technical aspects [4-6]. Precise simulations of the digitalized scheduling were also introduced [7-8]. This report applies these production scheduling methods to a banking service process to clarify its difference from production processes and discuss how to shorten lead time and enhance work efficiency of the service.

2 BANKING SERVICE PROCESS
2.1 Credit approval process for a housing loan
We need to select the service process for the analysis of this paper. The large number of banking processes can be categorized based on the uncertainty of the aforementioned constraints [9]. For example, the service that expects to "withdraw cash from an ATM" proceeds as designed without the influence of constraints. The customer inputs his personal identification number (PIN), specifies the amount to withdraw and the ATM checks the ATM card and balance after the transaction to complete. On the other hand, the customer expectation for the service to "manage the balance of a trust deposit" is constantly affected by the uncertain economy and politics, and processing it to the original design usually leaves a customer dissatisfied.
Our study chose the service of credit approval (CA) for a housing loan (HL), one that stands between the above two services in terms of uncertainty with the constraints. The customer has established the expectation to “borrow money to buy a house,” however, the constraints of income and housing price fluctuate. Banks nowadays have their own formulae to automatically judge the possibility of the customer making the mortgage payments, however, inputs to the formulae like income, other loans, land price, or insurance have to be double checked.

At first, we digitalized the CA process. The method is complicated, but familiar to engineers [7-8]. Figure 1 shows the entire process. The CA process of HL is roughly separated into three process groups.

The first is the digitalization process group that inputs personal information for the application in the computer. Japanese banks have a long history of following instructions by the Ministry of Finance and cannot change the manual process easily. The customer has to fill out the application on paper, not on a network, and confirm the application with a personal seal. The bank we worked with for this research is aiming to digitalize the process by first scanning the application form to input it as digital data, whereas the analog application form is still stored in a big warehouse in the suburbs.

The processes are separated into one for each judgment and the sequence of the approximately 30 sub-processes looks the same as a mold making process [2]. Since the passing criteria are clearly specified, the service operation needs no special skills. Each operator can pick up any lead of the open sub-process of any customer without any interference. Some important sub-processes, however, employ a dual system. Two operators work one sub-process separately, and the computer checks the difference of the result.

The computer requires the form to be completely filled out. For example, if the form is not checked, where it says full time employment, the system judges that the income is unstable and declines the application. However, some customers simply forget to place the checkmark and the process may require sending a letter urging the customer to reapply for the loan.

The second is the review process where income and land are checked. Other loans and insurance are automatically searched through companies that offer such services. However, reviewing the income requires someone to read the tax return documents and checking the land requires land appraisers to be hired. The matter is simple for salaried people working for large companies. However, self-employed applicants usually have a number of income sources and loans, and purchasing an office house for living would further complicate the process. Reading the pile of documents alone takes hours. When such efforts still do not meet the criteria, the bank would write letters or make phone calls to have further documents attached to the application.

The third group is completing the contract. The bank explains the conditions over the phone and then mails the contract to the customer. The customer signs the contract in front of a bank person or a judicial scrivener. However, some applicants lose interest in making a purchase at the time of signing.

2.2 Process design for digitalization

Just like the metal mold process, the digitalization process is divided up into sufficiently fine digitalized and independent sub-processes. The judgments are not only automatic but the processes themselves are visible.

Figure 2 explains the differences before and after digitalization. At this time, the processes are analyzed by axiomatic design theory [1]. Before digitalization, the process was a coupled design as shown in Figure 2 (a). The person in charge checked his own customer with the help of specialists or supervisors. Their expertise created interference in the process. The file of the document of each customer was a kind of process tag, but the progress on the work was not visible. Some files were stocked in his cabinet, making a long lead time. The design matrix might not be square because the number of customers is generally larger than that of operators.

After digitalization, the process became a decoupled
The system controls the lead of the open sub-processes. Any operator can pick up the application of any customer. The judgment of the sub-process is decided using information only in the processing sub-process. The process should not have any feedback loops. The design matrix is a lower triangular, square matrix, so it is a decoupled design. Axiomatic design theory says that the decoupled design of the digitalized process should be more effective for optimizing the job.

This design matrix can be optimized for shortening the lead time using the design structure matrix (DSM) method. DSM is also familiar to production engineers for designing hardware or software products. The tool of DSM, partitioning or clustering, can change into a triangle or quasi-diagonal matrix. In this paper, the change from the set of functional requirements in Figure 2 (a) to that in 2 (b) presented an essential improvement. This change was induced by axiomatic design for us but also could be done by the quality function deployment (QFD) method.

The operator has only to process the customer transaction and the system automatically records the process of who worked on which customer request. Even if an operator dozed off in front of the terminal that time is also counted as work hours so the count includes idle time. When a base process time exceeds 15 minutes, the system judges that an unexpected trouble occurred and closes the original base process. At this point the operator has to consult with the supervisor (escalation) or manually move on to the other means (additional activity) to revive the process as shown in Figure 1. Such remark, revise or rescheduling is familiar to production engineers. We could add the rework sub-processes like a manufacturing process.

In the second review process group and the third contract group, the process shifts from computer-based to manually-based with the help from specialists and supervisors. The first shift is to the process of land and building evaluation by an appraiser. The second is the process of closing the contract with a witness from the bank or a paralegal. The third is working with a supervisor to verify matters that are unclear. The time for this process is counted as work hours in bank.

These processes branched out from the review process and waiting takes place before and after the process steps. The longest waiting time, however, is in delays by the customer not filling out forms to resubmit. Half of the work of about 800 calls per day at call centers is taking inquiries from customers and the other half is reminding customers to resubmit applications. Without these efforts, customers usually do not bother resubmitting the forms again. These play-catch steps between the bank and the customer are unfamiliar to engineers because the week-span delay, not a minute-span one, happens in the scheduling as mentioned in the next section.

### 3 ANALYSIS RESULTS OF HOUSING LOANS

#### 3.1 Analysis results of lead time

Figure 3 shows the distribution of the lead times of digitalization and review processes combined as well as the lead time for contract for 9,323 cases that closed during a certain time period. The mode of the former is about 850 hours, but with a long tail in the long lead time area. The average was 1,177 hours (49.0 days). The mode of the later is about 450 hours with an average of 804 hours. If we combine the two, the mode is about 1,100 hours (46 days) and the average is as long as 1,981 hours (82.5 days).

Figure 4 splits the processes into some sub-processes and shows their average lead time values. In the former digitalization and review group, 81.5% of the 49.0 days, 40.0 days were spent in waiting for customer reply. If we look at outsourced processes, land appraisals took as long as 5,173 minutes (3.6 days, 7.3% of the entire lead time). Land appraisers receive data in digital form and complete the appraisals before their deadlines. Judging from the fee, the time for appraisal is probably about 1 hour. Waiting time among base processes in the bank was long as well. It was 7,667 minutes (5.3 days, 10.8% of the entire lead time). For example, the processes start every morning after mail delivery and a wave of processes is passed on to the following processes. When an operator cannot complete all of the applications during that day, the remainder will be left on the desk for at least 12 hours to add to the lead time. Lastly, the net process time without the idle time was only 243 minutes (4 hours), i.e., only 0.3% of the entire process time.
The later contract process group showed the same trend. Lead time waiting for customer reply was 56.7%, and that inside the bank was 43.1%, and the net process time was only 0.2% (105 minutes). In other words, most of the lead time was spent in waiting.

Figure 5 shows the distribution of net work time. The aforementioned combined digitalization and review processes have a mode at about 130 minutes, and the later contract at about 15 minutes. Similar to Figure 3, the number of cases tail long in the longer lead time area.

As we discussed above, many customers will be asked about their applications with unclear entries, however, a straightforward application that zips through the system about their applications with unclear entries, however, a straightforward application that zips through the system about their applications with unclear entries, however, a straightforward application that zips through the system about their applications with unclear entries, however, a straightforward application that zips through the system about their applications with unclear entries, however, a straightforward application that zips through the system about their applications with unclear entries, however, a straightforward application that zips through the system about their applications with unclear entries.

3.2 Distribution of Work Efficiency
The previous section analyzed the lead time from the customer standpoint. This section explains the analysis results from the standpoint of a bank clerk.

The amount of review for an HL varies largely with the economics. During our study, the period from the fall of 2007 to the summer of 2008 enjoyed a good economy and there was plenty of work. The Lehman shock that followed turned the economy down but it has started to slowly come back from the fall of 2009. Employees are hired in response to the amount of work. However, the time delay in hiring and terminations cause busy or relaxed work during such times of big economic change.

The work includes not only closed transactions, but also those that did not close. The rate of closing is also affected by the economy with an average of about 60%. Among the transactions in Figure 5, those with a short lead time consisted about 40%, thus among all of the transactions, 60% x 40% = 24% enjoyed short lead process times to go through the processes. In contrast, among those that did not close, about 50% were immediately turned down and thus, 40% x 50% = 20% rejections did not take long. The remaining half (20%) of those that did not close were, for example, cases where the customer wanted to refinance but lacked a mortgage due to a drop in land price. An accurate estimate for the land price required much information from the customer and the lead time was long. So 60% of the 60% closed and 50% of the 40% that did not close, a total of 56%, were customers that caused long work times for the bank and whether closed or not, the waiting time for the customer reply caused a long lead time.

Figure 6 shows the work by employees in the digitalization and review process group. The net work time captured by the computer was 41%, and that measured manually was 42%. The net time included 16% escalation that involved the supervisor, and 26% additional activity when a process step took 15 minutes or longer. The transactions that had to go through these extended time processes were 61% for escalation and 35% for additional activity. We traced the transactions to find that these customers were in the 56% of troublesome customers that caused long process time to the bank.
Other times totaled 16% including 4% attending seminars and meetings, and 10% waiting in between processes. When we measured the work time of metal mold operators [2-3], the time for attending seminars and meetings, waiting, and bathroom breaks totaled about 20% and thus was not very different from this bank employee time. This scheduling system could optimize 20% and thus was not very different from this bank.

When we measured the work time of metal mold operators [2-3], the time for attending seminars and meetings, waiting, and bathroom breaks totaled about 20% and thus was not very different from this bank employee time. This scheduling system could optimize 20% and thus was not very different from this bank.

Figure 6 shows breakdown of working time in the digitalization and review processes.

Other times totaled 16% including 4% attending seminars and meetings, and 10% waiting in between processes. When we measured the work time of metal mold operators [2-3], the time for attending seminars and meetings, waiting, and bathroom breaks totaled about 20% and thus was not very different from this bank employee time. This scheduling system could optimize 20% and thus was not very different from this bank.

When we measured the work time of metal mold operators [2-3], the time for attending seminars and meetings, waiting, and bathroom breaks totaled about 20% and thus was not very different from this bank employee time. This scheduling system could optimize 20% and thus was not very different from this bank.

Figure 7 shows examples of reasons for moving from automatic measurement to manual. All of them could have been avoided if the customer took time in preparing perfect applications. The fact was even a single entry that is unclear caused the operator to ask the supervisor or send additional questions to the customer.

4 DISCUSSION ON THE SERVICE PROCESS

4.1 Differences in processes for service and manufacturing

We found that the digitalization and scheduling method of manufacturing processes could be applied to the bank process effectively. Resource optimization could also be realized like a factory. Sequential sub-processes with independent judgments one by one satisfied the independent axiom of axiomatic design theory [1]. Each

operator could individually take the lead of the open sub-processes of any customer without any interference. The design theory also works well in the service process.

The shortening of the lead time, however, needed a totally different method from manufacturing. Figure 8 shows the results of comparing the CA for HL process lead time to those of mold manufacturing. The mold was a single piece production thus the involved work varied with each order and could not be arranged for a synchronized flow production. This caused 55.4% waiting time for CAD/CAM design or cutting tools. On the other hand, HL processes had to wait as long as 71.4% for customer replies. This long waiting time is not due to inefficiency of the workers. Comparing the net work-hours, HL was 0.3%, whereas that of molds was 44.8%. This shows the large influence of customer constraints is on the HL process.

In other words, the lead time consists of 71% waiting time for customer reply, and in terms of work efficiency, 42% occupies additional process time for customers with complex information. As a result, the main reason for poor efficiency is the customers, not the workers. The Japanese customers do not complain thinking that long review processes of 46 days by banks or government are natural. Thus, shortening the lead time would not gain customer satisfaction or excitement, and the service provider loses interest in such efforts. Banks, however, are different from the government and have the business motivation to reach a contract even with incomplete application forms from the customer. This motivation made the 42% net work-time of additional activity for the 56% special complex customers, and solving this part will lead to great improvement in productivity.

Forcing the customer, however, to always provides complete information and to return immediate replies is not easy. This is a large difference between service and manufacturing. Note that even in manufacturing, however, preprocesses of market research and interaction with the customer including presales, defining the specification or requirements are there, and often, the work proceeds with constraints not fully defined. This situation is similar to the bank service.

4.2 Improving the service process

As we showed in this report of HL process analysis, the shortening of the lead time of this service had a trade-off problem to the customer's satisfaction. Drastic reduction in process lead time and enhancement of work efficiency require more detailed and quantified information about the customer from the beginning, so the computer of the bank can make automatic judgments. Banks have worked hard in improving the application form so it is easier to extract the necessary information from the applicants. When these efforts succeed, escalations and additional activity for special customers will reduce as well as lead time and the amount of process work. We estimate that the improvement of lead time or necessary human resources will be reduced to about 70%. This solution, however, is shifting the work of eliminating vagueness in the information from the bank to the customer.

Figure 7: Reasons for escalation process.
The customer may want to pass the work to a consultant. In Japan, however, it is customary not to provide monetary compensation to consultants whose accomplishments do not have physical form. For example, consulting a customer who wants to build a custom design house would not accept an invoice for consultation, thus the fee is added onto the construction fee. The consultation appears to be a free service. Banks indirectly have been paying for it so far.

Overseas companies are now entering the Japanese market and like the case of life insurance business, Japanese banks will suddenly have to enter the competition of good service to the customer. In order to stay competitive in the race, banks should at least make the preparation of dividing the work, digitalizing, and individualizing them. The bank in this study did so and now they can at least visualize the work processes with the computer.

5 Conclusion

We applied the conventional scheduling method of the manufacturing process into the service process. We took the credit approval process of housing loans by a bank, analyzed the service processes, and aimed at shortening the lead time for the customer and enhancing the work efficiency for the bank.

Sequential sub-processes with independent judgment one by one based on axiomatic design theory was effective for optimizing scheduling for each customer or each operator. Each operator could pick up the lead of open sub-processes of any customer. The digitalization, scheduling and optimization method and design theory of the axiomatic design in the manufacturing could work well even in the bank.

But the shorting method of the lead time does not work sufficiently because the cause of the long lead time was related with the customer's constraints. Almost half of customers resubmit the application due to insufficient information in the application, and that caused the longer lead time, and affected the additional work time. Filling out the application forms with complete information, however, may reduce the customer's satisfaction because the customer has to do a troublesome job. Engineers should consider the scheduling of the service in the social aspect.

6 REFERENCES

Axiomatic Design for Understanding Manufacturing Engineering as a Science

C. A. Brown
Mechanical Engineering Department, Worcester Polytechnic Institute
Worcester, Massachusetts 01609 USA
brown@wpi.edu

Abstract
The objective of this work is to advance the understanding of manufacturing engineering as a scientific discipline in the context of axiomatic design. Scientific disciplines have a few, simple self consistent rules that can be applied to a wide range of problems. Manufacturing engineering includes a wide range of processes and systems without a clear unifying foundation. Two top FRs for the design of any manufacturing process or system are proposed: maximize value added, then minimize costs. The similarities with value engineering are acknowledged. The application of maximize and minimize FRs are discussed in process, business, and societal contexts.

Keywords:
Axiomatic Design, Manufacturing Engineering, Manufacturing Science

1 INTRODUCTION
The objective of this work is to discuss the development of manufacturing design as a scientific discipline through the application of axiomatic design. Scientific disciplines are distinguished by a few simple laws or rules that can be applied to solve a wide variety of problems that define that discipline. Manufacturing is, of course, a technological discipline. It relies on scientific findings and the scientific method. Nonetheless manufacturing is lacking a common, simple technological foundation, composed of a few, simple, recognized laws, which would make it a true scientific discipline. This discussion is aimed primarily at the design of manufacturing processes. The discussion could be extended to the design of manufacturing systems as well.

The establishment of manufacturing as a scientific discipline with rules that are common to all manufacturing processes would be important. Such a scientific discipline would be easier to understand, learn, teach, and apply to solving manufacturing problems [1].

Text books on manufacturing processes tend to be like encyclopedias e.g., [2-3]. The text books contain sequential descriptions of individual manufacturing processes, arranged by some kind of taxonomy, which is usually based on what is being done to the material, e.g., remove, add, deform, etc. The text books include different approaches to solving problems related to the processes. However, there is usually no attempt to connect the approaches used to solve the different kinds of problems. There appear to be no laws that are common to the process descriptions.

The application of lean principles [4-5] to manufacturing systems promotes an approach to the design of manufacturing systems based on value stream mapping and cost reduction that has been effective in improving the return on manufacturing operations. Lean, which was based on the Toyota production system (TPS), has been applied to many production and product systems (e.g., [6]), regardless of how similar their situation might be to Toyota's. Lean is largely constrained to system level design.

Suh [7] (p. 318) discusses a generalized design for flexible manufacturing systems, stating that the top FR is to maximize return on investment (ROI). Two DPs are proposed: (a) a system to produce components at a minimum cost, and (b) a system to provide a high-quality product that meets customer needs. Solution (a) works well when there is unlimited demand, as for Ford's early cars. Solution (b) applies when there is an overcapacity of production facilities; therefore it is the desirability of the product that determines success. This approach explicitly recognizes different environments for manufacturing. It also recognizes a high level general objective: ROI. Suh's decomposition leads to many of the same principles common to Lean manufacturing. Suh's discussion is limited to the system level of manufacturing.

The objective of this work is to advance the understanding of manufacturing engineering as a scientific discipline through the application of axiomatic design. Scientific disciplines have a few, simple self consistent rules that can be applied to a wide range of problems. Manufacturing engineering includes a wide range of processes and systems without a clear unifying foundation. Two top FRs for the design of any manufacturing process or system are proposed: maximize value added, then minimize costs. The similarities with value engineering are acknowledged. The application of maximize and minimize FRs are discussed in process, business, and societal contexts.

Keywords:
Axiomatic Design, Manufacturing Engineering, Manufacturing Science
2 METHODS
The method discussed here for the development of manufacturing axioms is to identify commonalities and generalize. Manufacturing can be defined as the process of transforming resources to produce a product that meets human needs. Human needs equates to value. To be effective, manufacturing must also provide a return on investment. The commonalities should relate to value and return. The development should include extending these principles to the details of the process level.

2.1 Process perspective
The most interesting problems in text books on manufacturing processes address force calculations and material behavior. The approaches to solving these problems do have some common roots, such as, force balance, analytical geometry and mechanics of materials. These solutions are certainly useful for the design of machines and tools that can transmit the loads necessary to perform the processes.

Usually in text books on manufacturing processes there is no recognition of the commonalities in the approaches to the solutions of manufacturing process problems. More importantly there is no discussion of the larger context in which problems in process design should be solved. Engineers working on process design problems are machine builders. Filling the need to solve these kinds of problems was a part of the motivation for the development of mechanical engineering which traces its origins to the industrial revolution. The industrial revolution largely starts in Germany, where the term for mechanical engineer is Maschinenbauer, literally translated: machine building engineer.

Mechanical engineers, it could be argued, were the first manufacturing engineers. The collection of subjects that currently comprises mechanical engineering, including thermodynamics, heat transfer, fluid mechanics, stress analysis and mechanics, are what an engineer would need in a factory in the first decades of the industrial revolution. The engineer focused on solving detailed technical problems associated with supplying power and making reliable machinery. Taylor is well known for developing the methods to study efficiency at the beginning of the twentieth century and developing Industrial Engineering [8-9]. Technical and efficiency problems continue to be issues that need to be addressed. Early in the industrial revolution some products were in such demand that all cars could be black. The human need was keen and market competition not so fierce. Black cars had enough value that product differentiation was not necessary for success. In such a business climate certain metrics can develop, such machine tool usage, which are not applicable in other business climates. These kinds of metrics are not sufficiently universal to be the basis of manufacturing axioms.

2.2 Recognizing useful commonalities
Consider the nature of scientific laws. The science of mechanics is governed by Newton’s laws. Newton’s laws cannot be proven. Examples of consistency with the laws do not constitute a proof. The laws are recognized because no one has seen a violation of them, at least within an accepted field of applicability. Newton’s laws are obeyed by mechanical systems. They apply to planets and to objects on earth. Therefore, there is no evaluation of goodness of mechanical systems based on how well they obey Newton’s laws. All Mechanical Systems obey Newton’s laws perfectly. The same thing cannot be said about manufacturing. Manufacturing relies on people who make choices. Manufacturing is not constrained by unbreakable laws as is mechanics.

Because a spectrum of manufacturing from good to bad exists, manufacturing laws would need to be different than Newton’s laws. Manufacturing laws would be closer to the axioms that Suh [10] has proposed for design. A spectrum of good to bad designs can and does exist. Suh states that good designs conform to two axioms, maximize the independence of the functional elements and minimize the information content.

Suh’s Axiomatic Design [10], in addition to defining the best design axiomatically, suggests how to solve design problems. Designing the process for the solution to design problems is itself a design problem. The top functional requirement (FR0) is to create the best design. The corresponding design parameter (DP) is a system to apply axiomatic design. FR1 and FR2 would be to maximize the independence of the functional elements (1st axiom) and to minimize the information content (2nd axiom). DPs 1 and 2 could be the applications of axiom one and axiom two, respectively. This approach could be applied at all levels of the design decomposition.

Suh states that the first axiom must be satisfied before the second [10]. That suggests that the design matrix for the design process is a lower triangular matrix. This means that the application of axiom one can influence the compliance with axiom two. Several of the theorems address this point of solutions to axiom one influencing axiom two ([7] p. 60-64).

In developing axiomatic design, Suh [10] studied existing designs to find the commonalities, proposed candidates and eventually narrowed them to the two axioms mentioned above. Womak et al. [4-5] essentially followed that route to develop lean manufacturing by studying Toyota. The development method here is to consider manufacturing in the context of axiomatic design.

Similar to design, manufacturing is a human creation with good and bad instances. Unlike mechanics and like design, manufacturing is not required to follow manufacturing laws. The laws would be what distinguish good manufacturing from bad. The general top FRs for systems for designing manufacturing process should include lean [4-5] and return on investment [7] and be applicable at the detailed process level. The number of FRs should be a minimum ([7] p. 60, Corollary 2).

3 RESULTS
This work proposes two FRs that can be applied to all manufacturing problems from the details of the process to the abstraction of manufacturing systems.

FR1 – maximize the value added to the product
FR2 – minimize the cost in the production process

Similar to the design axioms, FR1 must be applied first. There are legitimate costs that are necessary for the value adding processes. If minimizing the cost were to be the satisfied first, then the potential to add value could be eliminated. The corresponding DPs could be stated generally as systems to control the value added and systems to control costs.

4 DISCUSSION
An important question should relate to cost and the independence of the FRs. Cost is usually a constraint in product design because it cannot be decoupled from the other FRs [10]. Certainly there are costs that are necessary for producing value. Costs are therefore implied in FR1, so it might be argued that the FRs are not independent and cannot be independently satisfied and
therefore this decomposition violates Suh’s axiom one for independence. We might then restate FR2 as minimize the cost of the non-value-adding activities. The risks with this restatement are that value-adding activities, like conventional machining, discussed below, would not be scrutinized sufficiently, and that larger kinds of costs, like the cost to society of non-sustainability, would be ignored.

The terms maximize and minimize can be criticized in a design process because there are no clear limits, which means that there is no clear end to the process. FRs should define the objectives, or targets, for the design solution. No-cost manufacturing is not a realistic objective. “Control” or “provide” could be an alternative terms for both “maximize” and “minimize” in these and similar situations. Target values and tolerances could be set based on business analysis. Providing targets and tolerances would remove the frustration of never reaching infinite value or zero cost. Stating that something is an FR implies that it should be provided and controllable. Minimize and maximize provide a clear sense of the desirable direction. With any of the terms, a reasonable approach is to indicate target values and functional tolerances for the FRs as part of the design process. These targets and tolerances could be adjusted with each product and process design cycle.

What is proposed here, on an abstract level, has similarities with what can be found in the literature on value engineering and target costing. The emphasis in this paper is that the upper level should be applied to the smallest details of the process as well as to the manufacturing organization in a decomposition hierarchy based on axiomatic design. This important literature on value engineering and cost targeting provides useful themes for further hierarchical decompositions in the application of these principles to axiomatic design. Details of these applications are left to future work.

The intent here is to use the domains and decompositions inherent to axiomatic design in order to link value to the customer, starting with customer needs, through functional requirements to design parameters (e.g., physical features) and the processes to create them. These value links establish a two dimensional value chain that extends horizontally across the domains and vertically through the hierarchical decomposition.

Lean uses a process of value stream mapping that is essentially material based and follows the transformation of resources. Within axiomatic design a value chain analysis is established traceable to customer needs and functional requirements.

Costs are meant to include waste and investment. Clearly a means for reducing cost is to reduce waste and investment, provided that the desired value can be achieved.

Waste as a cost could be considered in a broad sense. Consider byproducts of manufacturing, such as carbon emissions. Legally, there are constraints on what can be discharged. Ethically, there could be constraints that go beyond the legal constraints. The first canon of engineering ethics is to hold paramount the health, safety and welfare of the public. Certain emissions resulting from certain production processes are harmful to the environment, and are damaging to the health, safety and welfare of the public. What are the ethical obligations of engineers involved with the design of these processes?

4.1 Application to Processes

The value in machining is often creating a surface with the desired roughness in the desired location. The costs in machining have little to do with creating this surface. Most of the cost in energy and tooling is in the formation of chips. The processes referred to as material removal are appropriately identified from the cost perspective, although perhaps not so appropriately from the value perspective.

The application of the proposed FRs could lead to precision castings and forging, which require less material removal. Another avenue suggested by the proposed FRs would be to minimize the energy required for material removal.

A kind of value chain analysis could be applied to processes. This would link value at the customer level, or customer needs with specific features created by the processes. The aspects of the process, which require resources, should be compared with the features at the detailed end of the value chain. The resource expenditures should be consistent with the creation of value.

5 CONCLUSIONS

Two top level FRs have been proposed for application to all manufacturing processes and systems:

1) Maximize the value added
2) Minimize the cost

The proposition is that these could also be axioms to be used as a foundation for manufacturing science.

6 REFERENCES

Axiomatic Design of Carbon Composite Bipolar Plates for PEMFC Vehicles

H. N. Yu 1, J. W. Lim 1, M. K. Kim 1, J. D. Suh 2, and D. G. Lee 1

1 Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology
Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea

2 Fuel Cell Vehicle Team, Eco-Technology Center, Research & Development Division,
Hyundai-KIA Motors, 104, Mabuk-dong, Giheung-gu, Yongin-si, Gyeonggi-do, 446-912, Republic of Korea
dglee@kaist.ac.kr

Abstract
A PEMFC (Polymer Electrolyte Membrane Fuel Cell or Proton Exchange Membrane Fuel Cell) stack is composed of Gas Diffusion Layers (GDLs), Membrane Electrode Assemblies (MEAs), endplates and bipolar plates. The bipolar plates are multifunctional components as they collect and conduct the current from cell to cell, separate the fuel and oxidant gases and provide flow channels in the plates to deliver the reacting gases to the fuel cell electrodes. The electrical resistance of bipolar plates should be very low in order to conduct the electricity generated in the fuel cell with minimum electrical loss. In this study, the concept of the continuous carbon fiber composite bipolar plate coated with graphite foil is suggested by Axiomatic Design Theory and verified experimentally. Consequently, the total resistance in the through-thickness direction of composite bipolar plate was reduced by 86% compared to that of conventional composite bipolar plates through the employment of a graphite coating layer.

Keywords:
PEMFC, Bipolar plate, Axiomatic Design, Graphite coating method

1 INTRODUCTION
A polymer electrolyte membrane fuel cell or proton exchange membrane fuel cell (PEMFC) is an electrochemical energy converter that converts the chemical energy of fuel directly into DC electricity when hydrogen and oxygen are supplied to the anode and cathode, respectively, producing water and electricity through the electrochemical reaction without making any pollutants. Therefore, the PEMFC is a promising power source candidate for automobiles, small-scale stationary power generations, and portable power applications [1-3]. A PEMFC stack is composed of Membrane Electrode Assembly (MEAs), Gas Diffusion Layers (GDLs), endplates and bipolar plates as shown in Figure 1. The bipolar plates have several functions in the fuel cell stack. They provide flow channels for reactant gases to maintain a proper pressure distribution over the whole active area of membrane electrode assembly, transmit electrons from an anode to its adjacent cathode in a unit cell, transfer the reaction heat from active area to coolant and provides coolant flow paths as shown in Figure 2 [1]. Therefore, there have been many research projects about the development of bipolar plates using various materials such as graphite, metal, and composites. However, graphite is not at all suited to the levels of mass production required for the full-scale commercialization of fuel cells because of its sophisticated machinery manufacturing process. Metallic bipolar plates have corrosion problems because the inside of a fuel cell is a very corrosive environment due to the presence of strong acids ranging from pH 2 to 3. From an Axiomatic Design perspective, both the graphite and metallic bipolar plate designs are typical coupled-designs where the number of design parameters (DPs) is less than the number of functional requirements (FRs) [4]. Continuous carbon fiber reinforced composite bipolar plates are plausible options for PEMFCs because they not only offer the advantages of low cost, high corrosion resistivity, lower weight and greater ease of manufacture than traditional graphite, but also their mechanical properties are strong enough to meet the DOE target [5,6].

However, the high interfacial contact electrical resistance of the composite material in the through-thickness direction is an obstacle to overcome. The high electrical conductivity of bipolar plate materials in the through-thickness direction is very important for the efficiency of the PEMFCs. The lower the electrical resistance of bipolar plates for the PEMFC, the less electrical loss or higher energy conversion efficiency will be. The resistance of the bipolar plate of the PEMFC consists of bulk material resistance and interfacial contact resistance (ICR). The bulk resistance is not a significant source of voltage loss in fuel cells, even for relatively high-resistivity plates. Although the electrical conductivity of the composite is several orders of magnitude lower than the conductivity of the metallic plates, the bulk resistive losses are on the order of several millivolts. Much higher resistance results from the interfacial contacts, such as between the bipolar plate and GDL [1, 7].

Figure 1: Proton exchange membrane fuel cell stack components.

In this study, this coupled-design problem of the PEMFC was solved based on Axiomatic Design Theory. Since the ICR between the GDL and composite bipolar plate is a function of the surface hardness of the composite, graphite was coated on the composite bipolar plate in order to lower the surface hardness while keeping high
mechanical strength and stiffness of the composite bipolar plate. Then, the effectiveness of this design method was experimentally verified.

4H + 2H₂O = 2H₂ + O₂

Water, Heat
Coolant
Bipolar plate (Cathode)
Bipolar plate (Anode)
Polymer membrane (Teflon+HSO₃)
GDL (Gas Diffusion Layer, Carbon fiber mat)

Figure 2: Unit cell of the PEMFC stack.

2 AXIOMATIC DESIGN OF THE COMPOSITE BIPOLAR PLATE

The functional requirements of bipolar plates can be summarized using Axiomatic Design Theory as follows:

FR₁ = Reduce the electrical resistance of cells in series connection.
FR₂ = Provide the structural support for the stack.
FR₃ = Separate the fuel and oxygen gases in adjacent cells.
FR₄ = Have corrosion resistance.
FR₅ = Provide flow channels.

In order to satisfy the functional requirements (FRs), in this work, the following design parameters (DPs) were developed and a design matrix was established as follows:

DP₁ = Carbon composite materials with high electrical conductivity
DP₂ = High strength and stiffness carbon composites
DP₃ = Consolidation of composites
DP₄ = Resin with high corrosion resistivity
DP₅ = Formability of carbon fibers

As shown in Figure 3, the design matrix reveals that the bipolar plate design is an acceptable decoupled-design by using the continuous carbon fiber reinforced composite with high electrical conductivity and high structural strength [8].

<table>
<thead>
<tr>
<th></th>
<th>DP₁</th>
<th>DP₂</th>
<th>DP₃</th>
<th>DP₄</th>
<th>DP₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR₁</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FR₂</td>
<td>0</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FR₃</td>
<td>0</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FR₄</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>FR₅</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 3: Design matrix of the carbon composite bipolar plate.

3 AXIOMATIC DESIGN FOR THE ELECTRICAL CHARACTERISTICS OF THE COMPOSITE BIPOLAR PLATE

3.1 Decomposition of FR of the composite bipolar plate

At the highest level of design, it was found that the continuous carbon fiber composite could satisfy the functional requirements of the bipolar plate. However, a voltage loss occurs due to the bulk resistance of the bipolar plate material and the interfacial contact resistance between the bipolar plate and the GDL. Therefore, the FR₁ was decomposed as follows:

FR₁ = Reduce the electrical resistance of cells in series connection.
FR₁₁ = Have low bulk resistance.
FR₁₂ = Have low interfacial contact resistance.

The bulk resistance of the composite bipolar plate depends on the volume fraction of the carbon fiber while the interfacial contact resistance is related to the contact surface conditions. The ICR is affected by not only the surface electrical conductivity of the materials but also the contact surface area and contact pressure. The lower surface hardness is suitable for reducing the interfacial contact resistance, however, it does not meet the requirement of the mechanical strength for the bipolar plate. Decomposing the FR₁ as FR₁₁ and FR₁₂, composite bipolar plate design is coupled as shown in Figure 4 [4,8].

3.2 Decoupled-design of the composite bipolar plate

In this study, the surface treatment method of the composite plate was proposed as the new design parameter to decouple the composite bipolar plate design as shown in Figure 5. As a result, DPs and FRs equalized in number, and the DP₁₁ and DP₁₂ could be written as follows:

DP₁₁ = Carbon fiber volume fraction of composites
DP₁₂ = Surface characteristics of composites

Figure 4: Coupled-design by decomposing the FR₁.

Figure 5: Decoupled-design by creating the DP₁₂.
It was known that the graphite foil has very low contact resistance due to its conformability. The interfacial contact resistance is affected by the surface condition of the GDL, which provides pathways for reactant gases from the flow field channels to the catalyst layer, allowing them access to the entire active area as well as the electrical connections from the catalyst layer to the bipolar plate [9]. As shown in Figure 6, the GDL consists of a randomly oriented carbon fiber mat. The randomly oriented carbon fiber side is the contact surface with the bipolar plate. Considering the surface condition of the GDL, a graphite layer with low hardness on the composite surface could increase the contact area effectively as shown in Figure 7. Therefore, the composite bipolar plate is a decoupled-design without compromising its mechanical strength and stiffness by creating a design parameter which is a graphite foil coating method.

4 COMPOSITE BIPOLAR PLATE WITH THIN GRAPHITE LAYER

The composite bipolar plate design was verified experimentally. A thin layer of graphite could be coated on the carbon/epoxy composite plate surfaces by using the prepregs coated with the graphite foil. Therefore, the graphite coating method did not require any post process after demolding the composite plate from the mold. Figure 8 shows the process of the graphite coating method, where the graphite foils with a backup film were placed on the both surfaces of the stacked prepregs.

It was laminated using the hot roller of 80°C to make the sticky resin of prepregs hold the graphite foil, and then the backup films were removed. Finally, it was found that a thin graphite layer was transferred on the surfaces of the stacked prepregs.

4.1 Experiment

The materials used in this study were carbon/epoxy prepreg USN 020 (SK Chemicals, Korea) and the thin graphite foil was BD-100 (Samjung CNG, Korea). The properties of USN 020 and BD-100 are listed in Table 1. The graphite layers were coated on the stacked composite prepregs surfaces. Then with the stacked prepregs, the composite bipolar plate was manufactured by a compression molding method under a pressure of 11 MPa at 160°C for 40 minutes. The stacking sequence of the specimen was [0]₈, and the thickness of the carbon/epoxy composite plate was 0.15 mm.

The morphology of the thin composite bipolar plates with the graphite coating layer of 2 µm was observed with a scanning electron microscope (SEM) (Sirion, FEI, Netherlands).

Also, the electrical resistance in the through-thickness direction of the thin composite plate was measured using a specimen size of 100 mm × 100 mm with the experiment setup as shown in Figure 9. The electrical resistance depends on several resistances in series, such as the resistance of the two copper electrodes (Rₑₑ), two GDLs (Rₐₐ), the bulk resistance of the specimen (Rₛₛ), and more significantly the contact resistances between the GDL and the specimen (Rₐₐ₊₊). They were measured under the four bipolar compaction pressures of 0.5 MPa, 0.8 MPa, 1.0 MPa and 1.5 MPa [10]. Also the electrical resistances of the graphite plate (0.4 mm) and the composite plate without treatment were measured to compare with that of the composite plate with graphite layer.

4.2 Result

Figure 10 shows the cross section of the composite plate surface with the graphite layer of 2 µm. In Figure 11, The electrical resistance in the through-thickness direction of the composite bipolar plate with the graphite coating layer of 2 µm decreased 86% compared to that of the composite bipolar plate without surface treatment under the compaction pressure of 1 MPa. Therefore, it was found that the thin graphite layer could reduce the interfacial contact resistance between the composite and the GDL effectively.
Table 1: Properties of prepreg (USN 020) and graphite foil (BD-100).

<table>
<thead>
<tr>
<th>Material</th>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon prepreg</td>
<td>Modulus (GPa)</td>
<td>230</td>
</tr>
<tr>
<td>(USN 020, SK chemical, Korea)</td>
<td>Strength (GPa)</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Ply thickness (mm)</td>
<td>0.02</td>
</tr>
<tr>
<td>Graphite foil</td>
<td>Density (kg/m³)</td>
<td>1.0</td>
</tr>
<tr>
<td>(BD-100, Samjung CNG, Korea)</td>
<td>Thickness (µm)</td>
<td>100</td>
</tr>
</tbody>
</table>

5 CONCLUSION

In this study, the concept of a continuous carbon fiber composite bipolar plate coated with graphite foil was suggested by Axiomatic Design Theory. Carbon composites satisfied the functional requirements of the bipolar plate at the highest design level. However, in the process of decomposing the functional requirement on the electrical resistance, the design became a coupled-design because the soft materials to reduce the interfacial contact resistance, and the requirement of enough mechanical strength and stiffness to support the stack components were incompatible. Therefore, a new design parameter was developed to make the composite bipolar plate design decoupled. To this end, a thin graphite layer coating method was devised to lower the surface hardness of the composite plate.

The decoupled design was verified by measuring the electrical resistance in the through-thickness direction of the composite plate. As a result, the total resistance in through thickness direction of the composite bipolar plate with the graphite coating layer was reduced 86% compared to that of the conventional composite bipolar plate without compromising its mechanical properties. It was found that the continuous carbon composite coated with the graphite layer was a promising material for the bipolar plate of PEMFC by Axiomatic Design.

6 ACKNOWLEDGMENTS

This work was supported by the WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10045-0), Hyundai KIA Motors and partly BK21 project.

7 REFERENCES

Which Virtual Reality Environment for Usage Evaluation of Innovative Surgical Instrument in Minimally Invasive Surgery?

G. Thomann\(^1\), Q. V. Dang\(^2\), J. Tonetti\(^3\), F. Villeneuve\(^1\)
\(^1\) G-SCOP Laboratory, Grenoble Institute of Technology, France
\(\{\text{guillaume.thomann, francois.villeneuve}\}@g-scop.inpg.fr\)
\(^2\) LAMIH, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes, France
dangquocviet_pfiev_00clc@yahoo.com
\(^3\) Orthopaedic and traumatology centre, Michallon Hospital, Grenoble, France
JTonetti@chu-grenoble.fr

Abstract
In the domain of designing innovative products in the medical field, investigations are oriented towards communication between actors and needs comprehension. In the DESTIN Project, User Centred Design methodology with concrete experiments is applied. Researchers propose emulations in the operating room for the co-evaluation of innovative products and new adapted surgical procedures. In this paper, they intend to evaluate the usage of the product in a virtual environment using a 3D haptic feedback system. They model the operating room and the surgical operative conditions and finally discuss the possible use of this virtual environment system for product and surgical procedure validations.

Keywords:
User Centred Design (UCD), Minimally Invasive Surgery, Virtual Reality, 3D-Haptic feedback system

1 INTRODUCTION
The development of new technologies in medicine can significantly improve the effectiveness. On the contrary, the use of more complex systems tends to make the practice of medicine more difficult. In particular, this complexity reinforces the importance of preoperative planning and postoperative monitoring. New technologies in informatics and virtual reality allow physicians to better interpret the enormous amount of information that is provided by the imaging systems or therapy systems [1]. Specifically, virtual reality allows better understanding, better planning and to better work through visualization of three-dimensional images of anatomy and pathology. In addition, virtual reality supports the practitioner through the stages of diagnosis, therapy, and postoperative monitoring.

In this article, we firstly present applications and research in the domain of virtual reality technology in medicine. The presentation of the DESTIN project (Design of Surgical-Technological Innovation) allows us to focus on a new procedure of spinal minimally invasive surgery coupled with an innovative surgical instrument. To better understand the surgeons' needs in this case, we create a CATIA CAD model of the virtual operating room. It integrates the patient, medical equipment and surgical instruments. In this virtual environment, the surgeon has to manipulate the virtual surgical instrument on the virtual patient's spine. The goal of this exercise is to provide information to the designers to validate innovative surgical instruments during the design process. At the same time, it also allows surgeons to perform the operative procedures with haptic feedback as in the real operative case. Next, we present the results of manipulating the haptic devices: 3D mouse and haptic arm.

The difficulty in our research concerns the ability to sufficiently represent a virtual environment for the co-validation of the medical procedure and the innovative surgical instrument. Currently, researchers, designers and the medical staff regularly work in the real operating room. This work is very effective but necessitates heavy organisation and management (mainly in the hospital), the creation of mannequins, the manufacturing of many prototypes, etc.

Thus, the research questions of this article can be summarized as follow:

- Which is the sufficient level of representation of the operating room and the medical equipment?
- Is it necessary to represent all of the human tissues for the simulation of the operation and for the haptic sensation by the surgeons?
- What is the procedure to quickly model the spine in the CATIA environment and is it necessary to realise the entire operative procedure?

To answer these questions, the research methodology proposed is: to better understand the medical operative procedure with the equipment used by the medical staff. It is essential to know the exact situations of the others medical instruments manipulated on the patient at the same time.

The spine has to be modelled in a compatible format as the CATIA environment and integrated in a mannequin placed on the operative table. We will establish the modelling procedure (that can be applicable in the future on other parts of the human body). The others medical instruments manipulated on the patient at the same time must be modelled and also integrated on the 3D-spine. The connection between the medical instruments and the haptic device concretises the experiment environment proposed to the user.
2 USER CENTRED DESIGN, VIRTUAL REALITY

2.1 Towards the User Centred Design Methodology

The purpose of the research behind this article is to propose an adapted design methodology and to find the place of the specific user in design process, particularly when he has a specific position (as an expert or a handicapped person for example) and when the design is highly dealing with his expertise.

In this case, i.e. when the user is not only the client of a product but also the final specific user, the design process has to be adapted to the situation.

The great importance of human aspects in industrial environments have changed the viewpoints of designers and developed Human-Centered design approaches. One of the fundamentals of this approach is to consider human factors at all stages of the design process. The integration of human factors in design process phases requires the effective use of appropriate human models. Shahrokhi et al. present definitions of human models and their classification in industrial applications with emphasis on industrial design processes in [2]. The authors also focus on the application of human models in a human centred industrial system approach.

In the domain of advanced manufacturing system for example, researchers explain that human impacts can significantly impair system performance and the future capability of the company to react to market requirements [3]. As it is discussed, these authors propose the development of an alternative 'human-centred' approach: technical and human aspects of advanced systems can be considered in parallel from the start of the design process.

Advanced production technology is not only characterized by higher automation of production flow and control, but more and more measured at the level of the ergonomics of man-machine interaction. Although much effort has been devoted to user friendly design and improved interface techniques, today's systems do not take into account their individual user's problems and tasks [4]. One possible answer to this problem is the design of 'cooperative', adaptable or adaptive user interfaces. The idea proposed is to adapt interface behaviour (presentation and dialog control) on account of individual user differences or user problems by reasoning about user intentions in situational work contexts.

2.2 Virtual Reality and application to the medical field

Virtual Reality (VR) is an interactive immersive data-processing simulation in real or imaginary environments. In the 20th century, VR played an important role in the training of the pilots and in the entertainment industry. VR technology has been applied in many different fields such as: formation by simulator (driving vehicles, aerospace), design of products, the simulation of surgery, meteorology, etc.

The application of the VR technology in the product design processes offers more than a unique tool for analysis. It also offers opportunities for the future users to access and act upon all of the information created during the design process [5]. By respecting the constraints of the virtual environment fixed and managed by the designers, the users can define and test usage scenarios, and evaluate the prototype products or models directly by their confrontation within the scenarios. While observing and measuring the choices that the users make within the scenarios, and by combining these observations with oral or written information, the designers recover better feedback in the context of use of the proposed product.

In the surgical field, laparoscopy is a new procedure which requires surgeons to observe the operation on a monitor and requires acquisitions of new competences. This Minimally Invasive Surgery (MIS) differs from open surgery by the fact that the surgeon operates through small incisions (a diameter lower than 12mm) and uses specific instruments such as scalpel, grips, nets, etc. [6]. In spite of its many advantages (faster recovery of the patients, less damage to healthy tissue, smaller scars, less pain and less need for drugs), MIS requires a long time to train the surgeon's eyes-hand coordination.

A virtual surgical system was developed in [7]. This system is composed of a movement and a force manipulator. It is able to obtain the effort haptic feedback in real-time when the surface of the liver is deformed. This system is under development with the aim to carry out complicated surgical operations on various parts of the body. In [8], the design of a haptic system with scissors is composed of two principal components: the haptic interface and simulation software. It is used to simulate the cutting procedures of soft tissues and to evaluate the force feedback by the comparison of the simulation results to the real data.

Research developing the haptic control feedback device can be found in [9-12]. To follow the user's intentional movements, by interaction between hand and device, powerful haptic devices must be able to produce force feedback. Consequently, it is essential to closely examine the human touched interface and the constraints of application during the construction of these devices. A haptic interface with 4 degrees of freedom was designed by Guiatni et al. to compare it with devices commercially available [6]. This device has the capacity to offer force feedback in all of the degrees of freedom available during the MIS procedure.

The technology of VR plays an important role in many fields. The development of MIS procedures in the surgical domain pushed the researchers to concentrate on complex questions such as simulating the deformation of soft tissues, simulating the operations of cutting, retraction, etc. Currently, even during the surgical procedure, the use of haptic systems helps the surgeons (with training) to better control the feeling of force feedback. However, these studies are concentrating on the simulation of abdominal surgery - the majority realized on the model of the liver. Moreover, it is time consuming to transmit information related to these deformations in real-time.

In our case, researchers, designers and physicians work together on the development of a virtual environment to simulate a MIS operation on the spinal column. The goal is to create an integral virtual surgical environment with surgical instruments and haptic feedback in a model of the operating room.

3 WHAT IS THE DESTIN PROJECT?

The DESTIN Project consists of reflection about complexity of medical instruments design in the MIS and some about the Scenario Based Design (SBD) which constitutes another proposition for the better integration of the user in the Design Process [13]. Scientific progress in the last decades makes it more and more possible to
satisfy the needs of the surgeons in terms of surgical materials and more precisely of surgical tools. MIS has the main objectives of making the post-operative constraints less painful for the patient, mainly by modifying the operative process with the aim of introducing miniaturized or modified tools inside the human body.

In the DESTIN study, researchers try to propose an innovative process design method with a better integration of the user which allows the designer to design surgical tools more adapted to the surgical procedure and better supporting for the surgeon’s ideas. This proposition consists of a new idea on the integration of the Scenario Based Design Methodology and the confrontation of:

- Evolution of the surgical instrument prototypes,
- Modification of surgical procedure.

This situation will be able to decrease the differences between the surgeons’ needs and the designed tools and their usage. The proposal is to develop the application of a Co-Evolutive Design Process methodology, with an explanation of prototype and surgical procedure evolutions (figure 1).

In the next section, we explain the different steps that have been taken to create a virtual environment as realistic as possible for the user’s simulation. Thanks to the assistance of 3D-software, it is possible to create a model of the operating room. Moreover, haptic technology can simulate the biomechanical behaviours of the human body during collision with the surgical instruments. This method would allow evaluating a new designed prototype coupled with operating procedures with lower costs and less time consumed. It would not be necessary to manufacture mannequins and prototypes, schedule emulation in operating rooms, etc.

The creation of a virtual surgical environment allows surgeons to carry out the surgical procedure using the last surgical instrument prototype with a haptic feedback as in the real surgical conditions. The proposal is to integrate as soon as possible the surgeon in the design process. The proposed model shows the co-evolution of the instrument and the use.

4 VIRTUAL REALITY IN DESTIN PROJECT CASE

In the design process, it is necessary to make tests in order to validate the design choices of the product. To validate the design in the DESTIN project, a complete mannequin should be created (spinal column, fibrous tissues for the muscles, skin, etc). This method is very expensive and time consuming. Consequently, the application of VR technology in this case can be particularly appreciated. Thanks to the assistance of 3D-software, it is possible to create a model of the operating room. Moreover, haptic technology can simulate the biomechanical behaviours of the human body during collision with the surgical instruments. This method would allow evaluating a new designed prototype coupled with operating procedures with lower costs and less time consumed. It would not be necessary to manufacture mannequins and prototypes, schedule emulation in operating rooms, etc.

The creation of a virtual surgical environment allows surgeons to carry out the surgical procedure using the last surgical instrument prototype with a haptic feedback as in the real surgical conditions.

In the next section, we explain the different steps that have been taken to create a virtual environment as realistic as possible for the user's simulation.

2.3 Surgical instruments and the operating room

Currently, there exist numerous different surgical instruments that are used during the MIS procedure of the...
spinal column (fracture of the vertebrae). Participating in real interventions and discussions with surgeons allowed researchers to establish a list of the main surgical instruments that are always used during the surgical procedure. However, depending on the surgeons, other instruments (with similar functions) can be used.

The focus on the operative phase when the surgeon carries out the intervention with the innovative instrument called Protige allows the selection of three main instruments:

- Similar to other screws, the pedicular screws (Figure 3) can be used in instrumentation procedures to fix stems and plates to the spinal column. These screws are used to immobilize a part of the spinal column to help fusion by maintaining the structure's osseous unit. The screws are placed on two or three consecutive medullar segments (or vertebrae). A short stem is used to connect the screws. This construction prevents movement on the segments which are amalgamated.

Figure 3: Pedicular screw and corresponding 3D-Model

- The persuador allows the head of the pedicular screw to be firmly held during the tightening of the nut. The role of the persuador is also to guide the nuts. In addition, its end is equipped with parabolic cuttings at each side of its axis, to allow the stem passage. Discussions with surgeons permitted researchers to simplify the model of the persuador by fixing the persuador and screw in a unique mechanism. The 3D-model of persuador and screw is modelled using CATIA software (Figure 4).

Figure 4: Persuador and persuador and screw model (software CATIA)

- The dilating device is composed of a series of tubes that allows the human skin to be dilated without damage. From the first incision (Kirchner pin) to the larger one, the surgeon threads each tube in another to progressively dilate the human tissues (Figure 5).

Figure 5: Dilating device composed of a serial of tubes

For the model of the operating room, videos of surgical interventions (different physicians) constitute the support of the analysis. Meetings with users permitted the number of medical equipments that are necessary to be modelled in the virtual environment to be limited. We decided to concentrate our efforts on three main pieces of equipment described as follows:

- The imaging intensifier is a system which allows the transformation of an optical image to an electronic image. It allows the surgeon to see the surgical instruments through the human body. It is the biggest equipment near the surgeon and the one the surgeon often use during surgical interventions. To take a picture of the position of the instruments, the surgeon pushes on a pedal just inside the operating table. The 3D-model of the imaging intensifier is modelled in Figure 6.

Figure 6: 3D-model of the imaging intensifier

2.4 Operating situation: the rod interesting operative phase

Due to the fact that the design of the innovative surgical product only concerns the phase of insertion of the rod inside the head of the screws, only a part of the complete surgical instrument needs to be modelled. The six screws,
the dilating device and the persuader already have to be placed on the spine while the surgeon manipulates the innovative surgical instrument.

Unfortunately, each physician has his own manner of inserting the screws inside the vertebrae. For the first model of the operating environment, we choose to work with an experimental surgeon who proposes his own practice: to simplify, it is possible to consider the pedicles like cylindrical structures. Thus, the trajectory should allow the Kirchner pin or the screw to remain in the side wall of the medical pedicle (figure 7).

![Figure 7: Position of the screw in the vertebra chosen to model in the virtual environment](image7)

Finally, to be in accordance with the working situation on the user, six persuador and screw models are inserted with this orientation in six vertebrae.

### 2.5 3D-model of the spine and integration in the mannequin

To simulate the surgical procedures of the spine, it is necessary to reconstruct the 3D-model of the spinal column and integrate this model in the mannequin using CATIA software. In this section, we compare the two identified methods that can be used in our case to 3D-model of the spinal column.

The first proposal is to reconstruct the 3D-model of the spinal column using the 3D scanner equipment (from METRIS©) and the KUBE software. Even if the data formats are compatible from the 3D scanner to the CATIA software, the 3D-model obtained is too complicated. There are many errors in the mesh model and the surface model. For the reconstruction of the volume from the surface model, computing time is too long and too much memory is necessary for the data processing (figure 9).

![Figure 8: Scanning procedure of the spine with the 3D scanner equipment (Grenoble, France)](image8)

![Figure 9: Reconstructing the 3D-Model of the spinal column from the scan data](image9)

This first 3D-model reconstruction proposal has many disadvantages (computing time, aberrant points, etc.). Instead, a standard polyurethane physical model was used for this step.

The objective of a 3D-model reconstruction of the spine in the virtual environment is to perform a simulation of a surgical intervention in a real case. Thus, in collaboration with radiographists in the hospital, we discussed a second solution: reconstructing the 3D-model of the spinal column from the medical-scanner images (figure 10). It is also one of the solutions that is widely used in medical research today.

Although the second solution appears to be more complex than the first one, the results are very encouraging. It is possible not only to use real medical imaging from a patient, but also the 3D-model obtained is more representative of the physical one: each vertebra is completely independent from the others (essential recommendation from the surgeon to perform the surgical procedure) and no holes are visible in the reconstruction.
5 RESULTS: APPLICATION AND EVALUATION WITH THE USER

Finally, a realistic environment has been proposed to the expert user to perform the surgical procedure with the innovative surgical instrument. The surgeon has manipulated the 3D-Haption® haptic system in the GI-Nova platform in Grenoble, France (figures 11 and 12).

Using the 3D-vision, the surgeon manipulates the surgical tool in the virtual environment with the objective of inserting the rod through the three heads of the screws linked with the persuadors (figure 13). In the beginning of the tests, the mannequin was half transparent to guide the user performing the surgical procedure.

The results of the manipulation (user feedback) are listed as follows:

- The user manipulates the haptic system for the first time. In spite of his enthusiasm and concentration, the representation in the virtual environment was difficult. We propose to do specific exercises before the next experiments.

- In the same manner as the haptic sensation, the user has difficulties using the 3D-vision of the system. For the moment, it constitutes more an obstacle than an asset for the experiment!

- The non-concordance of the axes (physical user interface vs. virtual surgical instrument) causes difficulties of manipulation.

- The friction between instruments (rod vs persuadors-screws) doesn’t sufficiently reflect the reality.

- In parallel to the haptic system, a space mouse is connected to the computer to orient the space environment as the user needs. This adding function is not easy to use.

In the beginning of the project and in accordance with the user, we decided to detect only the collision between the surgical instrument manipulated and the rest of the scene (no detection or feedback sensation while the surgical instrument enters the body tissues). Because the objective of the experiment is to validate the main usage functions of the innovative instrument, we decided (in accordance with the surgeon) only to hide the operative space with the presence of the mannequin.

The surgeon doesn’t blame this essential technical experimental condition during the virtual usage.

During the surgical interventions, surgeons use the existing mobility between vertebrae to insert the rod easily. These mobilities allow the insertion of the rod in the head of the screws in spite of their bad alignments. Unfortunately, the current virtual model of the spine

Figure 10: Reconstructing the 3D model of the spinal column from the medical imaging.

Figure 11: Surgeon is manipulating the surgical tool using the 3D-Haption® haptic system under CATIA software.

Figure 12: With special glasses for the 3D-Vision.

Figure 13: Spine with persuador and screw model inserted.
constitutes one unique part. An evolution of the 3D spine model is essential for future experiments. The user is satisfied on many points:

- The realism of the virtual environment: operating room, sufficient number and placement of medical equipment, realism of the spine with lots of details, placement of the surgical instruments, etc.
- He appreciates the easy manipulation and the quality of the haptic robot system
- He quickly imagines lots of other medical applications with this kind of technical equipment!

6 CONCLUSION AND PERSPECTIVES

During this research project, research questions were oriented to the satisfaction of the user manipulating a 3D-haptic feedback system.

From observation of surgical interventions in operating rooms and discussions with expert users, it was possible to give answers. A simple but sufficient operating environment was created to place the physicians in a known situation. Two procedures for the modelling of the spine were compared. One has been chosen and the spine model was integrated in the virtual environment. Chosen surgical instruments have been modelled to complete the realism of the surgical experience.

Experiments with expert users using the 3D-Haption® haptic system to perform a surgical intervention were satisfactory. The CATIA environment is satisfying for the surgeon (manipulation, realistic 3D view of medical equipment and surgical instruments) but some efforts need to be made on 3D-visions, 3D space mouse and specific manipulation exercises for the surgeon's practice. Research on concordance of the axes and friction between parts must be made.

CATIA has been used for this research because of the easy connection with the 3D-Haption® haptic system. We can consider that it is the right solution for initial validation. Due to the recommendation to explore the integration of the human tissues mechanical behavior in the virtual environment, other packages would better meet the requirements of this research. Connections with the SOFA software are currently being studied.

For more realism, it can be interesting to correlate this virtual spine surgical intervention with radiography control. From the surgeons' point of view, this virtual manipulation procedure is interesting for the simulation surgical interventions in other parts of the human body.

7 ACKNOWLEDGEMENTS

The authors gratefully acknowledge all the professors, engineers and technicians of the University of Grenoble. They especially thank all the people working on the Gl-Nova Platform, AIP Primeca DS; in the AIP Primeca French Network (http://www.aip-primeca-ds.net/ Gl-Nova .html)

8 REFERENCES

Exploring the Potential of 3D Visualization Techniques for Usage in Collaborative Design

W. W. Wits¹, F. Noël², C. Masclet²

¹ Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands
w.w.wits@utwente.nl

² G-SCOP Laboratory, Grenoble, France
frederic.noel@g-scop.inpg.fr, cedric.masclet@g-scop.eu

1 INTRODUCTION

This paper is a positioning paper to investigate the use and perception of 3D visualization techniques for collaborative design including remote interaction. To boost the information exchange and collaboration between design team members, clear communication plays a vital role. 3D visualization techniques support design teams in their collaborative work, due to their capability to comprehensively present the artifacts being designed. This, in turn, makes product models more tangible and increases product awareness, stimulating better mutual communication.

On the consumer market today, one can already buy 3D systems ranging from stereoscopic systems (usually involving special glasses) to autostereoscopic systems, also known as auto 3D displays. Examples of the former are, for instance, anaglyphic imagery, polarized imagery or alternating imagery. For each of these techniques special glasses are required. Examples of the latter systems are holographic imagery or lenticular lenses. For these systems no optical aids are required. Currently many companies are announcing that they will have autostereoscopic displays on the market in the near future.

Any such system may support a design team locally; however, for remote collaborative design support multiple systems should interact through a network. This research envisions a system that enables the connection of local 3D visualization techniques to a global platform, thus supporting a design team with specialists at different geographical locations. On this platform, design team members can work conjointly and simultaneously on their design, independent of their location. Such a platform could lessen the amount of travel, increase productivity and ultimately speed up the product design process.

As the use and perception differs for the various visualization techniques, this must be well understood first in order to strengthen the mutual communication. As mentioned some techniques require optical aids, whereas others lack the ability to provide an independent view of the design artifact to each group member. This might cause a lack of clarity when collaborating, especially when you are on a remote location. Interaction techniques need to be researched as well, as these may also differ for the various visualization techniques. Altogether, the potential of 3D visualization for collaborative design is explored.

1.1 Research goal

This research aims to advance the current state-of-the-art in collaborative design support with the use of an interconnected 3D visualization platform. Using the internet or other frameworks, interaction between engineers and designers should not be limited to a specific location. An essential aspect of the research will be to develop a platform that is able to join multiple visualization devices at different geographical locations. As aforementioned, next to the technological aspects, a clear understanding about the use and perception of 3D visualization techniques for members of a design team needs to be gathered. This knowledge will enable the exploration of the full potential of such interconnected visualization techniques.

The impact of new interaction modes for collaborative design must be investigated. For instance, depending on the scenario or product to be designed, what would be the best combination of interconnected visualization techniques and their associated manipulation facilities? Studies about the trends in digital design studios are already rising up. For instance, Van Doorn and Horvath [1] are especially focusing on the possible scenarios for the entire design process. They involve different interacting technologies according to their level of maturity.

This study will provide guidelines for potential users about what visualization systems to procure. Finally, this research should give us an answer up to what level of collaboration – even at remote locations – can be
strengthened by the environment and whether we continue to need traditional local design meetings.

1.2 Research approach of this study
To better understand the usage, possibilities and (in)conveniences for design teams working on a platform with multiple viewing stations, several 3D visualization techniques are compared. This paper presents a bottom-up approach as shown in Figure 1, where: (1) a design artifact is displayed on a specific 3D system, (2) the perception of the design artifact is studied and (3) model interaction methods are researched. These issues require a complete qualification of the holistic process with respect to tangibility and acceptability for collaborative design.

Figure 1: Bottom-up analysis process.

For this study, based on a restricted range of technical devices, the available 3D visualization techniques were anaglyphic and alternating imagery as stereoscopic techniques and holographic imagery as an autostereoscopic technique. Other visualization techniques as CAVE-based 3D immersion or full immersion Head Mounted Displays (HMDs) were not tested in practice, due to cost and availability; however, they are considered from a theoretical point-of-view. To strengthen the perception and interaction with the artifact, a classical mouse, 3D mouse, haptic devices and data gloves are considered.

2 OVERVIEW OF SOME 3D VISUALIZATION TECHNIQUES

2.1 Anaglyphic imagery
Anaglyphic imagery has been used for more than a century. Even though it has been recently replaced by more competing technologies, it still stands as an easy and very affordable means of achieving stereoscopic vision. It relies on two views of complementary colors of the presented scene that are slightly offset. A pair of glasses with color filters causes the viewer to perceive the projected artifact in 3D, as each filter occludes one of the rendered images for one of the eyes. Various color schemes exist, but most commonly used color filters are red and cyan. Compared to old movies or antique still images, the computer has brought the possibility to provide animated scenes in real-time.

Figure 2 presents the usage of anaglyphic imagery on a classical computer display. The advantage of anaglyphic imagery is that no specific hardware is required. In fact, only a pair of color filtered glasses – costing less than 50 eurocents – is required. A normal desktop PC or laptop is able to render the images smoothly. Also, multiple experts may gather around a screen or projection wall and look at the same artifact simultaneously. The artifact can be manipulated by any input device connected to the computer. In Figure 2 a classical mouse is used. Disadvantages of anaglyphic imagery are the inability to use the full color spectrum and sometimes retinal rivalry causes discomfort for the viewer. Furthermore, the need for glasses make the swapping from the virtual to the real world uncomfortable because of the dual color filtering.

2.2 Alternating imagery
The second stereoscopic technique studied – alternating imagery – solves the previously mentioned disadvantages. Also, in this case, the computer renders two views: one for the left eye and one for the right eye. Both views are projected slightly offset on a screen with a beamer that alternates the projection of both images. In this case the user has to wear special glasses that are synchronized with the beamer to occlude one of images for one of the eyes. The 3D model is perceived in the same manner as the previous technique.

Stereoscopic perception
Both stereoscopic techniques share the disadvantage that all users present in front of the screen see the same 3D view simultaneously regardless of their position in front of the screen. However, they perceive the projected 3D object at different locations in front of them. For instance, the designer on the right hand side in Figure 3 may be pointing at one of the engines; however, this will never be clear to the other designer on the left hand side. As features on the projected artifact cannot be pinpointed with an object outside the image, e.g. a finger or a stick, collaboration is hampered. Hence, to enable clear communication often an avatar representing a person’s hand is introduced into the virtual scene. In other words,
there is a clear separation between the Virtual Reality (VR) world and the real world. Using a head tracking system both stereoscopic techniques can enable the viewer to virtually look around the artifact (the computer renders an updated view for the new head position). Obviously this works only for one person; the other users will perceive the same movement of the artifact(s) as the tracked viewer. As movement cannot be shared with these stereoscopic techniques, this may also disrupt clear communication.

2.3 Holographic imagery

Holographic imagery is an autostereoscopic technique; that is, neither special glasses nor any other optical aids are required to see the image in 3D. Also in this case, the computer renders different views for the left and right eyes, but they are displayed through a holographic optical element. This optical element reflects the computer generated images to a specific (narrow) viewing angle, separating the views for the left and right eyes, and thus producing a 3D image for the viewer. Recently, much progress has been made in this area [2-3] and promises new opportunities for collaborative design.

Figure 4 presents the usage of holographic imagery. Again the same airplane is used as the virtual product model, only this time manipulated by a different haptic device (Phantom Omni). The advantage of this technique is that multiple viewing angles can be defined around the holographic element, which allows each viewer to look around the artifact individually by shifting the position of their head. The available views of the artifact depend on the number of reflected viewing angles.

Holographic perception

Perception of the image is determined by the viewer’s position around the holographic optical element. However, features on the projected image can be pinpointed with an object outside the image and will be perceived by all viewers at the same location within the scene. As, in this case, there is no clear separation between the VR world and the real world, we will refer to this as Augmented Reality (AR). The concept of AR was first published by Feiner et al. in 1993 [5]. They differentiate AR from VR by “presenting a virtual world that enriches, rather than replaces, the real world.”

Often AR is dedicated to improving (enriching) a desktop environment with a see-through HMD. Here, it is important that the user is provided with a tangible interface to this Mixed Reality (MR) [6]. In the presented case of holographic AR this seems not to be an issue: the virtual model is already very tangible. Following our bottom-up approach of Figure 1, we must however still investigate how tangible input devices are in respect to the potential usage. Needless to say, tangibility only applies to local viewers. For design team members at a different geographical location, an avatar would need to be projected within the scene to show the remote interactions.

3 IMPACT ON COLLABORATIVE ACTIVITIES

3D visualization techniques combined with internet communication networks provide new opportunities for collaborative design. Here, cooperation refers to synchronous activities where one or more specialists must share some perspectives about a common artifact. With the development of network communication, remote interaction on various representations of artifacts is available. VR and AR devices, supporting this feature, can be combined to share models in remote locations. The current section discusses the differences between VR and AR technologies regarding their potential usage for remote cooperation.

Scheme (Figure 1) illustrates how we will focus our interest on new holographic technologies.

3.1 Perception with stereoscopic techniques

The difference between VR and AR technologies remains confusing in most cases. Here, we analyze how they differ in respect to user perception. Let us consider the perception of various 3D visualization systems. Most 3D displays are based on a "Trompe l’oeil" technique. Two levels of techniques to produce 3D perception are identified, where 3D images are mapped onto a 2D screen.

The first 3D representations were provided by wireframe drawings (Necker cube) [7]. People are able to perceive a 3D structure rather than a collection of 2D segments, but static views are ambiguous because a depth cue is absent: dual interpretation is possible. Regarding the information required, the observer is sometimes confused. For instance, in Figure 5a it is clear that one cube is presented, however it is not clear which face of the cube is in front. Also, significant time of acclimation to learn the visual code and to build the interpretation scheme is observed.

It has been demonstrated that motion cues can partially solve this problem. Wherever the relative movement originates from – observer changing the motion of the artifact or observer moving around the virtual artifact – a slight change in the perspective angle of the observer gives sufficient information to clear-up any doubt on the spatial orientation. A depth cue occurs when the view is rotated. Nowadays, we can combine the rules of perspective projection established by “Renaissance”
painters and the capacity of computers and graphic cards
to display the images real-time. Color, illumination, texture
and shading effects all improve the 3D perception of the
object, as illustrated in Figure 5b. Such rendering
techniques experienced great improvements in recent
years, in both the business and entertainment fields. They
are now quite popular and accessible with medium range
computers.

At the second level of 3D perception, the capacity to
mislead the observer's brain is increased by proposing
different images for each of the observer's eyes. The main
point is that the two images remain 2D images and the
brain is still in charge of building a 3D mental
representation of the scene, but it does not have to
interpret the visual code anymore. Stereoscopic
technology was already established at the beginning of
the 20th century. Once again computers are now able to
animate the images, thus extending the misleading
perception, as was shown in Sections 2.1 and 2.2.

Whatever the technology (anaglyph, alternation, etc.), 2D
images are projected on a surface. Figure 6 illustrates the
perception volume of the 3D scene on such a display. The
images are projected onto the screen (usually a plane),
while the perception of the object may be in front or even
behind the screen. In any case, the perception will be in
front of the observer and it will remain in the space
between the observer and the far limit behind the screen.
In addition, the maximum protrusion distance of the scene
perceived in front of the screen, depends on the last real
world visual reference between the observer and the
screen. The mind perceives a 3D object but any additional
visual reference between the observer and the projection
screen changes the point of view and pushes the mental
perception away from the observer. Hence, trying to grasp
the object with your hand pushes the front perception of
the object back towards the screen.

Figure 6: Perception of a 3D image projected using
stereoscopic techniques.

In CAVE based VR systems, the observer is immersed
between a set of projection walls. In this case, the
previous perception is improved by the fact that images
come from various directions simultaneously. This cancels
any real world references from an optical point of view,
fully immersing the user in the scene. With head tracking
systems the projected views may also be adapted to the
viewer's point of view. However, the principle remains the
same: the observer can approach the object but he cannot
touch it.

With both discussed levels of 3D techniques, the
observer's brain is variably in charge of building the 3D
scene. Most technical evolutions are directed to fool the
visual perception (sensitive stimulus) to enhance the
cognitive interpretation. The hypothesis is that providing
the most realistic view (eye stimulus) to the user helps
him to increase consistency with his internal
representation (what the brain knows it should look like).
The second level of 3D perception opens collaborative
activities to anyone not acquainted with 3D
representations and their conventions.

3.2 Perception with autostereoscopic techniques

With AR technologies, the image is mixed with the real
world. Some technologies that provide visualization for
AR are: (1) by tracking the observation direction, a
specific image may be mapped onto 3D glasses
overlaying information upon the real world (HMD), or (2)
with holographic technology an image can be constructed
in 3D space, removing the previously necessary
misleading of the brain's interpretation, as was shown in
Section 2.3.

Holographic technologies are also not new; however,
recent developments make them available with real-time
3D image processing [8]. In this case, the image is
projected in 3D space and it is not a surface image in
"Trompe l'oeil" anymore. Figure 7 describes its
installation and perception. The observer without any
dedicated optical aids sees a 3D image. This time the
perception zone is equal to the 3D location of the image
and the observer may enter the 3D perception volume
without interfering with the perception of the object as
long as he does not interfere with the projected views.

Figure 7: Perception of a 3D image projected using
holographic (autostereoscopic) techniques.

AR, as part of MR environments, tries to hybrid real life
artifacts with virtual artifacts [9]. However, the observer
may face several perceptual issues. These have been
analyzed theoretically in a study focusing on AR and MR
[7]. Among the various issues described, the
misperception of an object's location is a real concern
whenever direct interaction is required. Mixing references
from real and virtual worlds tends to mislead the observer.
If AR is used in a way that overlays information onto a
real scene, some shift between real and virtual objects is
acceptable. Conversely, accuracy is critical when objects
must come into contact with each other, especially when
the real object is part of the observer's body that wants to
interact with the scene.

Furthermore, not only the visual cues are involved in MR
perception, it is also a great challenge to keep
consistency between the visual sensations and others
3.3 VR and AR interaction methods

Real manipulation of virtual artifacts provides an intrinsic way of interacting with the model. Kinesthetic and visual modality to learn about and perceive the artifact will enhance the quality of acceptance by the actors. Kinesthetic feedback allows people to get information they would not have access to with visual means only, for instance, weight and gravity effects, dynamic properties of mechanisms (inertia), material properties (density, ductility, plasticity, etc.) or coarse textures. Kinesthetic feedback allows people to get information they would not have access to with visual means only, for instance, weight and gravity effects, dynamic properties of mechanisms (inertia), material properties (density, ductility, plasticity, etc.) or coarse textures. Anticipating technological progress, we could even imagine the perception of temperature, fine texture, sensual or pain sensation. This kind of interaction will be very important for designers because it gives valuable information to understand and evaluate the design intention. It extends the concept of material utterance as defined by Dearden [11] for digital materials. With VR technologies, the observer cannot interact directly with the scene. The interaction must be handled by indirect input devices. Many devices are proposed from classical mice to haptic arms. The observer handles a device that does not belong to the scene and a virtual artifact must be mapped into the scene to localize this interaction. This avatar helps the observer to enter the scene, but he must renew his interpretation to achieve full interaction. With AR technologies, the observer can interact directly with the scene. The perception does not change with respect to the position of external visual references as a hand or finger. It is truly AR since the produced image includes the real world. An avatar is not required within the scene and interactions can be applied directly to the 3D scene. Nevertheless, limitations exist that should be taken into account. Some specific studies have led to assess the use of holographic displays for the sake of product visualization [12]. The user perspective has been favored to obtain relevant quantitative observations. Logically, among the several heuristics taken into account, few elements gave evidence considering that the potential interaction with the virtual objects was satisfactory enough. Thus, such manipulation must be precisely foreseen to check for compatibility with the scenario and to choose the best interaction methods.

Table 1 presents an overview of the features and user benefits for various 3D visualization techniques.

<table>
<thead>
<tr>
<th>Features</th>
<th>User benefit / discomfort</th>
<th>Stereoscopic</th>
<th>Autostereoscopic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual aids</td>
<td>User has to wear special glasses</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Depth resolution</td>
<td>Depth of the 3D view</td>
<td>Good</td>
<td>Limited</td>
</tr>
<tr>
<td>Color resolution</td>
<td>Presentation of full color spectrum</td>
<td>Technology dependent</td>
<td>Good</td>
</tr>
<tr>
<td>Continuous motion / parallax</td>
<td>No discontinuity when moving around the screen</td>
<td>Only with head tracking (1 person only)</td>
<td>Not possible (set-up dependent)</td>
</tr>
<tr>
<td>Direct interaction</td>
<td>User can directly interact with the 3D scene</td>
<td>Not possible</td>
<td>Not possible</td>
</tr>
<tr>
<td>Collaboration</td>
<td>Can user interact with team members</td>
<td>Not possible</td>
<td>Not possible</td>
</tr>
</tbody>
</table>

Table 1: Overview of features and user benefits for various 3D visualization techniques.

senses (e.g. tactile, audio, etc.) in order to avoid MR sickness [7, 10].

4 FUTURE SCENARIOS FOR HOLOGRAPHIC AR

AR not only gives the opportunity to deal with geometrical dimensions (releasing size constraints for instance). It also helps to enhance reality by overruling time constraints. The user has the power to represent what has been (visualizing the past), what will be (forecasting the future) and what could be (simulation). This offers new opportunities in many situations. Hereafter 2 examples are given: remote surgery and cooperative product design.

Remote surgery

Imagine remote surgery: usually a surgeon must watch a screen where he perceives visual feedback of his actions; however, the actions are performed through actuators driven indirectly. Figure 8 shows a set-up where the visual cues are presented using stereoscopic techniques, and interaction and feedback is handled by two haptic devices. This virtual training set-up was developed by Vrest [13-14], which was formed as a result of a collaborative project between a local hospital and our research group.

Figure 8: Virtual surgical training.

A clear separation between both spaces (visualization and interaction) is required due to conflicting technology demands. For the surgeon this is inconvenient; he cannot mix real and virtual worlds directly, nor can he interact with another surgeon standing next or opposite of him. Holographic AR technologies would allow the display and actions of the user to be integrated in the same space. This would require some devices to track the position of the hands and tools handled by the user, for instance with data gloves. However, in the end, the surgeon can focus on his actions directly. Moreover, his actions become much more tangible because they are clearly connected to his real world perception. Also, the glasses are redundant for holographic systems.

Cooperative product design

In the field of cooperative product design, communication between remote experts is vital. In this case, the use of holographic AR technologies will provide a more tangible system to simulate a co-located meeting. New visualization platforms must be envisioned, constructed and verified in this direction [15-16]. A holographic display for each expert on a remote location allows him to share...
3D objects with other experts. This enhances the capacity of engineers to share their ideas and present their design models in 3D.

Every engineer may act directly on his own 3D virtual prototype for instance to annotate or to modify the product model real-time. The actions of the actors (hand movements, etc.) must also be tracked and dispatched as events on the shared 3D model. Design engineers in other locations can perceive the complete modification and all annotations of the model on a realistic perspective (real 3D) or they can participate as observers using a more “classical” 3D stereoscopic system. Actions of remote colleagues must be displayed as an avatar in their local scene.

Ultimately this could result in a new form of 3D video conferencing, where you would see your remote colleagues through a 3D screen and between the both of you the virtual model under discussion is presented. Such a 3D conference system is presented in Figure 9.

![Figure 9: 3D video conferencing with distant colleagues](image)

In situations where the scene is complex, requiring accuracy and correctness, 3D allows the improvement of the job quality and provides a substantial gain in time [18]. Current AR technologies allow natural cooperation on the same virtual object for co-located observers. There is no need to digitize or re-create the contextual world. Human interactions are possible in a natural way with actors standing face-to-face or side-by-side. Real artifacts (professional tools, spare parts, missing parts from archaeological pieces, etc.) can be put into the AR scene to interact with virtual artifacts.

Breen et al. [19] demonstrated that in order to improve tangibility for designers, engineers or other actors, models pass from the real world to the virtual world or vice versa. For instance, a static capture of the real world is integrated into digital models or v.v. digital models are materialized with rapid prototyping tools. As these are all static representations, one step further would be to completely mix virtual and real worlds in a dynamic mode. This is what we envision to achieve with holographic AR technologies.

Additionally, on a platform, holographic AR technologies must be able to combine interactions from several remote persons. You can see the avatars of your colleagues and simultaneously you can work on the same scene. In any case, both scenarios benefit from more tangible (real world) interaction methods with the virtual scene.

5 CONCLUSIONS

Augmented reality nowadays already provides so many applications that it could become an essential technology to assist in future everyday life activities, as well as in many business fields. However its acceptance remains an open issue because, as any other tool, it expects the user to adapt to new constraints. Among the technological achievements, 3D visualization is a promising result. To effectively work with 3D visualization devices, they must allow realistic representation and tangible interaction. Holographic devices will be a key technology with respect to this goal.

This global discussion opens new research directions that will be followed by the authors of this paper. A major issue will be to characterize the level of tangibility of an interaction device with respect to a specific usage context. As a result of this research, clear indicators should be formalized to choose suitable collaboration techniques.

In the future, this would allow design teams to collaborate better among its members, even for distributed design teams. In the end, this will boost the performance and quality of work of entire teams.

6 ACKNOWLEDGMENTS

The authors would like to acknowledge the work on holographic imagery that has been done at the KTH Royal Institute of Technology in Sweden and their willingness to share their research findings [2, 4].

This research is undertaken in the context of the VISIONAIR infrastructure project currently created by the European community. VISIONAIR is an acronym for VISION Advanced Infrastructure for Research. The authors are actively involved in the creation of this infrastructure.

7 REFERENCES


A Virtual Assembly Approach for Product Assemblability Analysis and Workplace Design

Z. Y. Ding¹, K. K. B. Hon¹,², F. Shao²

¹ School of Engineering, University of Liverpool, Liverpool L69 3GH, UK
² Virtual Engineering Centre, University of Liverpool, Warrington WA4 4AD, UK
hon@liv.ac.uk

Abstract

A virtual assembly approach is proposed for the analysis of product assemblability during the design stage. A visual assembly environment created in DELMIA is developed and its capability is demonstrated via a case study of a blower assembly. Based on the simulation of the entire assembly process, cycle time and human factor issues are evaluated from an ergonomics point of view. Comparison results show that the assembly cycle time and energy expenditure can be reduced when the operator posture is improved as a result of workplace redesign. Simulation results of the virtual assembly can be used for the planning and validation of the actual assembly process as well as feedback for design changes.

Keywords:
Virtual assembly environment, Manual assembly, Ergonomics, Workplace design

1 INTRODUCTION

To enhance competition in the global market, manufacturing companies are increasingly concerned with product and process validation in order to facilitate efficient and effective engineering changes [1]. Assembly process validation is an area that needs special consideration due to its direct influences on product quality, time to market and cost [2]. The minimization of the product cycle time and work induced fatigue are two significant factors during validation. Therefore, it is beneficial to incorporate a complete ergonomic analysis of the assembly process for a new product design.

In the late 90s, ergonomic analysis was mostly supported by videotaping systems, i.e. a videotape of an operator performing the assembly operations [3-5]. Such research analyzed the videotape of work methods and workplace layout. With the development of computer hardware and software, 3D simulation techniques were employed in ergonomic analysis. Generally, information on assembly sequence, operator movements, etc. from the workplace is collected and then a virtual manufacturing environment mimicking the actual physical environment is constructed. Iterative simulations can be conducted in such a virtual environment for ergonomic analysis [6-8]. However, such analyses are often based on an existing product and workplace. The construction of these physical prototypes will increase product development lead time and costs. Therefore, there are significant advantages in the concurrent study of human factors for product assembly during the design stage.

Chryssolouris et al. developed an experimental virtual environment for the verification of manual assembly processes [9]. An immersive virtual environment with a CyberGlove was used in their study of four alternative layouts for the assembly of a boat propeller. The influence of a number of process parameters and their combinations on the process cycle time were also quantified. Rajan et al. developed a Virtual Reality-based environment JGPRO for aircraft floor assembly jig design and analysis [10]. 3D CAD models of assembly product, jig and a virtual hand were imported into JIGPRO for assembly process simulation and accessibility analysis. The main purpose was to analyze accessibility during assembly and to evaluate the risk of musculoskeletal injuries. Sundin et al. described a case study of bus chassis assembly which aimed to improve efficiency and ergonomics in the early design stages [11]. ‘Jack’ was used for the construction of a computer manikin and ergonomic analysis of different work sequences including posture was conducted.

This investigation demonstrates the capability of a virtual assembly approach in product assemblability analysis and workplace design based on a blower assembly case study. A virtual assembly environment is created and a digital human model is introduced to perform all necessary assembly operations. Three metrics are used in this study: operator posture, energy expenditure and process cycle time. They are evaluated and used for the redesign of the workplace in the phase of virtual assembly.

2 METHODOLOGY

2.1 Virtual assembly environment

DELMIA V5R20 is adopted for the creation of virtual assembly environment and process simulation. Its digital manufacturing approach is built upon a Product, Process, and Resource (PPR) model providing a central hub connecting all relevant data as shown in Figure 1. Especially, its virtual ergonomics provision offers a digital human modeling capability for the creation, validation and simulation of operator/product interaction. The digital operator can perform activities such as walk to a specific location (across floors, up ladders, down stairs) based on time parameters defined by the user, move from one target posture to another, as well as pick and place parts in the work area by following the movements and paths of objects. These activities can be combined with assembly activities...
for the analysis of the relationship between operators and other entities in simulation. Generally, operator activities can be defined by three different methods. Firstly, defined by standard activities, such as walk, climb the stairs and climb the ladder. However, only a limited number of standard activities are available in DELMIA. Secondly, defined by adjusting the freedom of a segment. A digital human model in DELMIA has 68 segments. They all have their own degrees of freedom, which is usually 2 or 3. By regulating the degree of freedom of a segment, specific operator posture can be defined. Finally, the third method is via inverse kinematic. When the path of an object is defined, the operator's hand can follow this specific path in space. Meanwhile, there are 5 inverse kinematic control points in DELMIA which are line of sight, pelvis, right hand, left hand, right foot and left foot. Users can assign specific locations for these points in space and corresponding posture of the operator can be obtained through inverse kinematic.

A complete virtual environment allows the preparation and inputs of all digital models involved in assembly process. These include component 3D models, workplace models, materials storage models, tool, fasteners and an operator model. These models should be loaded and located in the virtual assembly environment.

Within the virtual assembly environment, a rapid analysis for assembly process can be conducted. The product assembleability in terms of performance metrics such as RULA scores (cf Section 2.2), assembly process cycle time and the energy expenditure of the operator can be estimated. Such analysis can be part of an approach for a single or multi objective optimization. An overview of the virtual assembly environment simulation is shown in Figure 2.

![Figure 1: PPR model in DELMIA.](image)

---

### 2.2 Analysis methodology

**RULA score**

The RULA (Rapid Upper Limb Assessment) system was developed to investigate the exposure of individual operators to risks associated with work-related upper limb disorders [12]. An RULA analysis examines the following risk factors: number of movements, static muscle work, force, working posture, and time worked without a break. All these factors are combined to provide a final score that ranges from 1 to 7:

- **1 and 2** – Indicates that the posture is acceptable if it is not maintained or repeated for long periods of time.
- **3 and 4** – Indicates that further investigation is needed and changes may be required.
- **5 and 6** – Indicates that investigation and changes are required soon.
- **7** – Indicates that investigation and changes are required immediately.

**Process cycle time**

A functional timer enables the user to record the process time of each individual step of the assembly process. Manual assembly can be divided naturally into two components, handling (acquiring, orienting and moving the parts), and insertion and fastening (mating a part to another part or group parts) [13]. Thus, the total assembly process time \( t \) is calculated as:

\[
 t = t_h + t_i
\]

Where \( t_h \) = handling time

\[
 t_i = \text{insertion and fastening time}
\]

**Energy expenditure**

GARG equations are empirical metabolic energy predictive equations. They are adopted for energy expenditure calculations in this investigation as shown in Table 1 [14]. It assumes that an assembly task can be divided into simple basic operations. Once this step has been applied, the average rate for the entire job (in kcal/min) can be estimated by summing up the energy requirements for each individual operation and the energy required to maintain the posture.
GARG equations

For stoop lift:
\[ E = 0.0109 \text{BW} + (0.0012 \text{BW} + 0.0052L + 0.0028S \cdot L)F \]
For squat lift:
\[ E = 0.0109 \text{BW} + (0.0019 \text{BW} + 0.0081L + 0.0023S \cdot L)F \]
For arm lift:
\[ E = 0.0109 \text{BW} + (0.0019 \text{BW} + 0.0081L + 0.0023S \cdot L)F \]

Where
- \( E \): energy expenditure (kcal/min)
- \( \text{BW} \): body weight (lb)
- \( S \): gender (female = 0, male = 1)
- \( F \): lifting frequency (lifts/min)
- \( L \): load weight (lb)

Table 1: Energy expenditure calculation in virtual assembly environment.

<table>
<thead>
<tr>
<th>Task</th>
<th>Handling</th>
<th>Fastening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crankshaft assembly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piston sub-assembly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piston assembly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing assembly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impeller assembly</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3 CASE STUDY

As a demonstration of the capabilities of the virtual assembly environment, a case study of an aluminum blower assembly was carried out. The blower, as shown in Figure 3, was designed for remote control model aircraft applications [15]. The heaviest part of the blower is the housing which has a mass of 5.695 kg. The assembly process consists of five tasks: 1) crankshaft assembly; 2) piston sub-assembly; 3) piston assembly; 4) housing assembly; 5) impeller assembly. Each task requires the operator to acquire parts from the storage bench, transport them, and fasten them in a fixed location on the workbench. The digital operator is based on a 50 percentile US male as provided within the DELMIA database. The operator has a height of 175.58 cm and weighs 78.49 kg. The virtual assembly environment constructed for this case study is given in Figure 4 and the individual steps of the assembly process are shown in Table 2. The objectives of this case study are:

- to predict and evaluate ergonomic issues during assembly process;
- to model the assembly cycle time and energy expenditure;
- to improve assembly cycle time and energy expenditure via workplace redesign.

![Figure 4: The virtual assembly environment.](image)

![Figure 3: Blower model for assembly [15].](image)

![Table 2: Assembly process of the blower.](image)
4 ANALYSIS

4.1 RULA analysis

A propagation of the assembly task’s RULA scores along the process time is given in Figure 5. There is only one RULA score at any instant and it is updated for each 0.1 sec interval. For this operation, the highest RULA score 7 appears in the crankshaft assembly operation. This sends a strong signal that changes must be made to this task immediately. The sequences of operator posture for this task are shown in Figure 6. It is evident that in order to insert the crankshaft into the engine block, the operator is squatting and the upper body is leaning and stooping. These postures impose stress and fatigue for the operator. With such knowledge, it is logical to hypothesize that a new and less stressful posture could be attained if the workbench is higher.

Based on this observation, the workplace layout was redesigned by increasing the height of the standard workbench from 70 cm to 90 cm. After the workplace design modification, a detailed ergonomic simulation was performed again and the new RULA scores for the assembly tasks are shown in Figure 7. For the new workplace design, the RULA score 7 in the ‘danger’ zone is totally eliminated. Further, the stressful posture of squatting is removed and other postures are also improved with the new operator posture as illustrated in Figure 8.

4.2 Process cycle time analysis

The original assembly cycle time \( t_s \) is 213 sec. After workplace redesign, the new cycle time \( t_s \) is reduced to 204 sec. The total cycle time is reduced by 4.2% simply by workbench height adjustment. An overall comparison of each assembly task for the two simulations is given in Figure 9. It is evident that 3 out 10 assembly tasks, i.e., crankshaft fastening, housing fastening and impeller fastening, show a reduction in assembly time while there are no changes for other operations. In this case study, the assembly cycle time reduction is attributed to the removal of the undesirable and stressful posture of squatting as a result of workplace redesign.

4.3 Energy expenditure analysis

As the process cycle time was reduced by 4.2% after the workplace redesign, the total energy expenditure was also reduced. Moreover, based on the GARG equations in Table 1, the energy expenditure per min of arm lift is lower than stoop lift and squat lift when other parameters, i.e. the body weight, operator gender, lifting frequency and load weight remain unchanged. Consequently, the energy expenditure will decrease correspondingly when the squat and stoop movements are reduced. As the energy expenditure related to the original workplace was 8.951 kcal and the result for the redesigned workplace layout was 8.544 kcal, the reduction in energy expenditure amounted to 4.5%. This change will contribute to an improvement of the overall operator performance.

5 CONCLUSION

A virtual assembly approach is proposed for the analysis of product assemblyability. In this investigation, a virtual assembly environment was developed and an assembly cycle time and an energy expenditure calculation model were embedded into this environment. A case study of assembling a blower for a remote control model aircraft
application is employed to demonstrate the capability of this virtual assembly environment. The entire assembly process simulation consists of four assembly tasks and one sub-assembly task. Human factors were evaluated and a new design of the workplace is proposed based on the ergonomic problems identified in the assembly process. Simulation results show that the assembly cycle time and energy expenditure can be reduced through an improvement of the operator's posture. The process cycle time and energy expenditure can be reduced by 4.2% and 4.5% respectively. Further simulations could be made to optimize individual metrics or a combination of them. Results obtained from the virtual assembly analysis can also be used for process validation and verification before actual production.

6 ACKNOWLEDGMENTS
This work was supported by the Northwest Regional Development Agency and the European Regional Development Fund.

7 REFERENCES
Abstract
In the current factory layout design domain, the factory layout is usually a geometric representation. The domain suffers from many problems such as inadequate types of geometrical information, difficulties in combining different file formats and comprehensive information management. This paper describes the problems, needs and visions in the domain. The authors have identified several existing standards that are suitable as a foundation for information related to factory layout design, for a more integrated information management and for efficiently creating and analysing factory solutions. A detailed description and discussion about diverse standards are included in the paper.

Keywords:
Factory layout design, Standards, Information representation

1 INTRODUCTION
Many manufactures today are constantly changing and developing their factories, manufacturing systems or buildings due to different needs and challenges. The reasons can be the need to manufacture a new product, expansion, demand for shorter throughput time, etc. All of these are drivers for a better factory planning and factory layout design. This study focuses on how to represent factory layout design related information using existing information standards that are not delimited to factory layout information representation. Today there are no information standards dedicated only for factory layout design, but there are some other information standards which are partially suited for factory layout design in different ways. This study focuses only on machining and assembly factories.

2 DIFFERENT TYPES OF FACTORY LAYOUTS
Factory layout is considered to be at the core of the factory design process. During the development of the factory planning pilot in the ModArt project, it was identified that the layout development is the most essential activity in factory design [1]. During factory layout, different domain models of media, machines and building etc. are merged together. It is in the factory layout where the results of the product flow simulation will appear in relation to the equipment. A good virtual layout is a layout that can be utilised for verification of realiseable combination of objects from different domains. For example, it will determine if the foundation and a machine-tool fit together in size and weight. A detailed factory layout specifies a part of the flow and has an effect on the factory during its life time, in terms of shorter through put time, better space utilisation and more. This means that while developing and realising the layout for a factory, information about machine weight, foundation load capacity, electrical port location, etc. is important. It should be possible to represent this information in the factory layout model for the efficient creation and analysis of factory solutions. This, in turn, means that a good layout is not just a graphical representation, it is the model for all required information. Fulfilment to standards and laws related to e.g. safety can also be verified in the factory layout model.

Factory layout model is a very wide concept and has different meanings. As for all kinds of models, the factory layout model has purposes, viewpoints and detailed levels for its information. It is further explained in section 4. Below are some of the views that a factory layout can have in the factory design domain. These views on layouts need to be considered in the information standard evaluation.

- Block layout:
The block layout is a layout used in an early stage of the layout design process. In 2D drawings or 3D models, machines are just represented conceptually by boxes with approximated size or just a marked sub-area, meaning that much of the information about the machines doesn’t have to be specified. The most important part in the block layout is the division of the space, e.g. the area of machines, buffers, operator space and maintenance areas [2].

- Detailed layout:
A detailed layout [2] should contain all the information needed to describe a system. A system can be a factory, a line, a cell, etc. The information in a detailed layout should be realiseable and reflect a real factory at a certain level of detail.

- Media layout:
A media layout is a generic term for the layout of electric power, process fluids, ventilation, water systems, IT/telephone, and more.
Foundation layout:
Many machines or larger pieces of equipment require their own "islands" to stand on due to weight and need to be isolated from disturbances such as vibrations from other equipment. Therefore it is necessary to build a specific foundation for each of them to meet their requirements. A foundation layout describes these specific foundations in terms of dimension, load capacity, material, and more.

Apart from the different layouts above, there are layouts such as working area layout, painting area layout, safety layout, lighting layout, building layout, and more. In the detailed factory layout, many of these layout types need to be merged together for a better verification of available space.

There are three main domains in factory layout: the manufacturing system domain, the media domain and the building domain. As a factory designer, it is important to have the ability to combine these domains.

Figure 1 is an example of an integrated geometrical factory layout of the three domains. Current work methods for integration involve converting different file formats to system acceptable formats, and re-modelling e.g., the geometry for a proper level of detailing. It means that the original layouts in Figure 1 have different file formats and different levels of detail for the geometrical information, and that they have been modified to fit into one geometrical model. Many hours of work and competence in various systems are required to integrate a factory layout with the manufacturing system, media systems and the building. But still it is not sufficient to efficiently develop and realise a factory.

Figure 1: Part of an integrated geometrical factory layout

3 GENERAL PROBLEMS, NEEDS AND VISIONS
There are several identified problems and needs for factory layout:
1. Different file formats make the file exchange problematic. It is difficult to integrate layouts from different domains. For example, in many factory layouts today the media part is usually represented in symbols on separate sheets instead of in the models. It is necessary to have media represented in factory layout models so that e.g., collision can be detected in time.
2. The file sizes of the layouts are usually large, which makes it difficult to combine several layouts for visualising the whole factory system. This problem is often related to geometrical information within each layout file or model. Currently a model contains a lot of geometrical information that is not relevant to factory designers. One typical example is geometrical models of machine-tools from machine suppliers. They usually have too detailed geometrical information about the inner parts of the machines, e.g., spindles. For factory designers the machine outer contour is the most important geometrical information together with some other information such as the openings, footprint and media interface positions. This means that the file size will decrease dramatically if we only sort out the geometrical information relevant for factory designers.

3. Besides the geometrical information, there are other types of (non-geometrical) information to be considered. For example, a machine-tool model contains the type of machine (its functions) as milling or turning, the machine-tool weight (a property), etc. Today this information usually is stored in different repositories and/or data formats such as Microsoft Word, Microsoft Excel or PDF. The information management will be easier if the object properties are stored in an organised structure related to geometry. In this way, the time spent on finding usable information will decrease and information accuracy will increase.

4 LAYOUT DRAWING VS LAYOUT MODEL
It is essential to distinguish the drawing from the model. A layout drawing is often a layout that is a 2D view. A layout model contains much more information compared to a drawing. A layout model should have multiple views: a 3D view or specific view and detailing levels for conveyors, trucks, paths, media, etc. It can support different asynchronous or synchronous collaborations between people/projects. In a layout model, information should be stored for reuse, manipulation and consolidation. In a layout model, the geometrical information should be integrated with non-geometrical information i.e., a combined model (section 3, problem 1), and layout from various disciplines can be integrated i.e., an integrated model (section 3, problem 1). Therefore the factory layout model requires a suitable information model in the future to solve these problems and support the needs and vision from section 3. Currently the integrated factory layout model normally only contains 3D geometrical information.

If a combined model for e.g., a machine-tool exists, further possibilities and advantages may be found by sharing this model between disciplines or reusing this model in other-software applications such as kinematics simulations, operation planning or manufacturing concept planning. The information that is needed to generate behaviours of the system in the factory can be stored in this model. Combining this information with a process will give operation time for the flow simulation or machine-tool power for the power consumption calculation.

To meet the needs above, a standardised information model for information related to factory layout design is essential. This will decrease the file format exchange problem and will improve information management in the factory design area. A standardised information model will also give companies a chance to own their information instead of having the information embedded in commercial system structures which they have no control of. At this stage there is no information standard developed specifically for factory layout design, but there are several which can be used for information related to the factory layout domain.

Here is a summation of reasons to use standards for information related factory layout design:
For integrated information management among different domains of factory layout design, sharing data is very important. Using the information standards for better knowledge and information management has been suggested in research project including SPECIES [3] and also in a state of art study [4].

- For improved interoperability between software programs.
- For improved availability of information.

A comprehensive assessment has been performed on information collection and evaluation. Information and knowledge about factory layout design and its process were collected, analysed and modelled in a web-based system called “production pilot” [5]. Based on this information and knowledge, the evaluation of existing standards was performed. Generally there are two parts of the evaluation: 1) study of the overall functionality and feasibility of the standard to represent general information related to factory layouts such as geometry, elements within a building, coordinate systems, etc. 2) detailed evaluation mostly based on the three typical issues described below.

5 ISSUES REGARDING FACTORY LAYOUT REPRESENTATION
Apart from the general problems and needs, there are some typical issues related to information representation in factory layout design e.g.:

1. How to describe (model) the relationship between a machine centre and its electrical cabinet with its cable? This issue includes these main descriptions: geometry of port, connector, electrical cabinet, machine centre and cable, the relationships between objects geometries, the object functions, relationship between functions, and object property e.g., voltage and the location. This issue concerns the manufacturing system domain and media domain.

2. How to describe the relationship between machine footprint and the foundation that the machine is standing on? This issue includes these main descriptions: geometry of machine-foot and foundation, definition of machine-foot-print, machine weight, load capacity of foundation, relationship between machine weight and load capacity, foundations material and the location. This issue concerns the manufacturing system domain and building domain.

3. How to describe the relationship between a robot and the machine-opening when there is no the physical contact? This issue includes these main descriptions: geometry of the machine-opening and robot, the envelope of the robot arm and opening, and the function and the relationship between the opening and the arm. This issue concerns the amount of space or distance that is needed between two objects.

These three issues above contain many specific detailed descriptions and there are many similar issues related to them. If these three issues could be represented by standards, the rest of the similar issues can be solved. A small part of the objects, properties and relationships in the issues are listed in Table 1 to exemplify the detailed evaluation.

The detailed shapes of manufacturing systems and resources are not considered in the first stage of the information model evaluations for layout. The CAD part is not the problem because it is a generic part that many domains are sharing, and it is the most developed part in the standardised data exchange world. Product management data (e.g., version, variant and status) is also an essential part of factory design, but this part is not the main focus either; because the evaluation scope will be too broad.

The evaluation of a standard mostly focuses on the questions:

1. Can it represent the objects and their properties within the factory layout design?
2. Can it represent the relationship between manufacturing resources, media resources and building resources to support development and realisation of the factory layout?

6 REFERENCE INFORMATION ABOUT STANDARDS
Three standards have been studied and evaluated in this work: ISO 10303-214, ISO 10303-225 and Industry Foundation Classes (IFC).

6.1 IFC
IFC is registered by ISO as ISO/PAS 16739 and is currently in the process of becoming an official International Standard ISO/IS 16739 [8]. IFC is developed to represent an information model structure for sharing construction and facility management data across various applications used in the building domain [7]. IFC uses the EXPRESS (ISO 10303-11) modelling language for the data schema specification. As the EXPRESS is used, an implementation can be done utilising part 22 SDAI (Standard Data Access Interface) and Part 21/28 files.

6.2 STEP
ISO 10303, industrial automation systems and integration - product data representation and exchange, named STEP (STandard for the Exchange of Product data), is developed to represent the product data model [8]. The objective of STEP is to provide a neutral mechanism capable of describing product data throughout the life cycle of a product, independent of any particular system [9]. Within STEP there are many application protocols (APs) focusing on different areas. In this paper two APs are evaluated: AP 214 and AP 225. For all APs there are two types of models: an application reference model (ARM) and an application interpreted model (AIM). The ARM is the application specific information model and the AIM is an interpretation of the common generic information model provided within STEP. This means that AIM is the conceptual link between different APs.

AP 214: Core data for automotive mechanical design processes
AP 214 is developed to exchange information between various software applications within the automotive development process [10]. AP 214 is not specifically developed to represent information related to factory layouts design, but some of the conformance classes are also suited to represent a manufacturing system, seen from the viewpoint of a factory designer. It has also been shown that the AP 214 can represent the basic information for the manufacturing system development [11]. According to M., Johansson [12], AP 214 can also represent manufacturing facilities with certain functionality such as machine shop, paint shop etc. in conceptual design.
AP 225: Building elements using explicit shape representation

AP 225 is developed “for the exchange of building element shape, property, and spatial configuration information between architecture, engineering, and construction” [13]. The information within the information model can be used in all stages of the life cycle of the building, from the designing stage to maintenance. The purpose is to assist the exchange of information between software applications in the building and construction sectors. AP 225 can e.g. integrate building structure design with service system design which is a must in the factory design process.

7 REPRESENTATION OPTIONS FOR FACTORY LAYOUT INFORMATION

Four different options (see subsections below) are identified regarding how to use existing standards to represent factory layout models based on the evaluation.

Figure 2 shows a result (summation) of the general evaluation. In the figure, “YES” means that the specific information standard has the ability to represent the modules and their information content while “NO” means that it doesn’t have the ability to represent the information. “NO*” means that the standard doesn’t have the ability to represent the information as it is, but there are some other ways to represent these information in the standards. One example is entity “IfcDistributionElement” from IFC standard originally defined for all elements that participate in a distribution system within a building such as heating system and ventilation system. But this entity can be used for the machine-tool representation. Figure 3 is a simplified instantiation of IFC with main entity “IfcDistributionElement” representing a machine-tool with its shape, location, self-defined properties (current and voltage) and relations to other objects.

Further detailed evaluation based on specific issues is required to verify if the specific information can be represented by the standards. It is not sufficient to use a general evaluation to exam the detailed feasibility of...
standards. The information models and regulations within standards have to be tested with specific information/data to make sure that the standards meet the needs within factory layout design. The detailed evaluation of the standards mostly focuses on the three issues and a part of it is presented in Table 1. In Table 1 the description of the specific objects, properties and relationships is listed in the first column, and the corresponding standard representations are shown in the remaining columns. For example, “Cable connector” is a specific object and “Item” is an entity in the standard AP 214 to represent a “Cable connector”. With AP 214, a group of other entities can be used to further classify the generic item as the specific object e.g., “Cable connector” but only the main entity e.g., “Item” is shown in Table 1. Based on the evaluation the answer to the two main questions posed in chapter 5 is: no, not all of the objects, properties and relationships can be represented by the standards AP 225 and IFC. But still, these standards can be used as a base for the factory layout representation. AP 214 is the only standard that can represent all the objects, properties and relationships in this study.

7.1 Option one
Use only AP 225 to represent the three main domains: the manufacturing system, media and building construction. The main reason to use AP 225 is that it is a domain specific AP with building and media specified.

7.2 Option two
The second option is to use only AP 214 to represent factory layout. The main reason to use AP 214 is that it can represent objects, properties and relationships in this study with its generic character.

Advantages
- AP 214's complexity is lower compared to AP 214 and IFC, which means that it is simpler for implementers and developers.
- It is a domain specific AP with building and media specified. There are fewer requirements for an external concept model or classifications since many domain specific concepts are already in AP 225 and thereby minimise the need for some other concepts in an external concept model.

Drawbacks
- AP 225 is not so widely used compared to AP 214 and IFC. (The development of AP 225 has been inactive for years.)
- AP 225 does not include the representation of processes as AP 214 and IFC do.
- The information model lacks the ability to represent some information e.g., a manufacturing system and its relationship to a building. This is because AP 225’s focus is on building construction with its geometry representation. In short, AP 225 needs to be extended in order to support factory layout design and this extension work can be comprehensive and time consuming.

Option three
The third option is to use AP 214 to represent manufacturing system information and AP 225 for building construction and media information. Then implementations of these two APs can be connected as they share the generic information model of STEP. This option is proposed mainly because of the fact that AP 225 has more domain specific concepts and relationships which make it easier in representing buildings with its service systems whereas AP 214 is more suitable for representing machine-tools. Service systems are those systems that aim to serve manufacturing systems and the buildings, and are many times called media systems. This cross AP option makes the information (data) model more complex and further research need to be carried out.

7.3 Option four
The fourth option is to use only IFC to represent factory layout. IFC is a standard with a main focus on the buildings with its service systems. The IFC is more
complete and domain specific compared to AP 225 which also has a focus on buildings and its service systems.

Advantages
- It is a domain specific information model with domain specific concepts related to building and service system which means requirement on classification is less within these two domains.
- It is a well known standard in the building sector.

Drawbacks
- The growing numbers of versions of IFC with many changes make the core information model less stable and increase the work in software development [14]. The version used in this study is IFC 2x4.
- IFC has issues regarding the representation of manufacturing resources e.g., machine-tools, conveyers and robots. A direct representation of these is not possible, see Figure 3.
- IFC has difficulty representing relationships between properties.

8 CONCLUSION AND DISCUSSION
The evaluation of the three standards is performed. The results, answers and options are given, and the details are exemplified. Even if the details described in the three domains can mostly be represented by the standards, the representation path or quality can vary depending on how they are represented in the standards. E.g., electrical port for the electrical cabinet is easier to represent in IFC than in AP 225 and AP 214. The reason is that in IFC the entity electrical port is predefined while in AP 225 and AP 214 the entity is not predefined. Another example is machine footprint. The IFC standards has specifically pointed out that the "IfcFurnishingElement" has the attribute "footprint" but this footprint represents only 2D outline of the item [15] while the factory layout domain needs footprint that can represent different types of footprint scenarios and some of them need 3D shape representation. Concluded from the evaluation, AP 214 has the best ability to represent all the details due to its generic character and that IFC is the most domain specific standard due to its domain specific concepts, see Figure 4.

However, this does not make AP 214 the most suitable standard for factory layout design domain in the near future because there are other standards which are not yet evaluated and the classification work is not yet accomplished. The classification work is time consuming and knowledge requirements are high on developers. But for a longer perspective a classification work and development of a reference data library such as in ISO 13399 cutting tool data representation and exchange is very important for the future of factory layout design. This work then can then be used for many purposes not only enriching AP 214. One application is to have it integrated with the "Production Pilot" (described earlier), for a unified...
One way to minimise the classification work in AP 214 is to use the concepts in IFC as domain specific concepts and classification to enrich the information models in AP 214. An example of this is done for the “site” concept, given in Figure 5. The IFC entity “IfcSite” is used to classify entity “Item” in AP 214. In the similar way the properties in the “Pset_Sitecommon” can be used to classify the “General_property” from AP 214.

Other standards such as ISO 10303 AP 239 Product life cycle support (PLCS) also have been considered as options but unfortunately they are not preferred for several reasons. PLCS is developed to be able to represent any complex product during its entire life cycle [16]. As an extension to this standard, there can be reference data libraries with external classes of more domain-specific concepts. By referring to these external classes, the standard can be further specialised into specific areas such as machine tools [17] or factory layout design. This means PLCS requires a reference data library for factory design which is not yet developed.

However, compared to the needs of factory layout design, PLCS demands an unnecessarily heavy system load to meet the requirements of factory layout design. Another issue with PLCS is that it does not support representation of geometry elements related to properties in a detailed level which is required by factory design.

There are also some other standards within the STEP family that can considered for future evaluations. These are AP 227: Plant spatial Configuration and AP 231: Process Engineering. STEP AP 241 for AEC (Architecture, Engineering and Construction) facilities is under development by ISO TC184 SC4. It started several years ago. The ISO 15926: "Industrial automation systems and integration—Integration of life-cycle data for process plants including oil and gas production facilities" is also a standard that can be considered in the future.

However, this study is performed on existing standards with a focus on what factory designers need to know to develop and realise a factory for the manufacturing industry. This is necessary due to the needs in the domain as described earlier. By comparing and evaluating these standards, a better overview on what these standards can represent in the domain of factory layout design is summarised.

All the options above have their advantages and drawbacks. This study shows that option two and option four are most preferred because these two requires less comprehensive work comparing to other two options which include the extension of standards. The choice can vary depending on the work situation and the focuses within the factory layout design. The factory designer is the interface between the manufacturing domain, media domain and building domain. These three domains sometimes can be very far away from each other, especially the building domain and manufacturing domain. There are two main questions related to the system usage and development that need to be answered:

- If it is important to be compatible with others in the manufacturing area. STEP is well-known compared to IFC in the manufacturing area, which means that STEP is more publicly acceptable in this area. It is always easier to develop something well-known. Unfortunately in the manufacturing community, STEP is mainly known as a file format able to exchange 3D geometry and assembly structures.
- If it is important to be compatible with the building construction domain in the future. IFC is better known in the construction domain than STEP which means it will be accepted more easily. This file format is known by many in the construction domain.

9 ACKNOWLEDGMENTS

This work is done within the Factory Design Process project, financed by the VINNOVA FFI program in collaboration with Scania and the EXPRES project.

10 REFERENCES


Product Lifecycle Management Model for Design Information Management in Mechanical Field

J. Le Duigou, A. Bernard
Ecole Centrale de Nantes, IRCCyN, 1, rue de la Noë, BP 92101, 44321 – NANTES Cedex 3 - France
julien.bernard@irccyn.ec-nantes.fr, alain.bernard@irccyn.ec-nantes.fr

Abstract
Product Lifecycle Management (PLM) is one way to improve productivity in all manufacturing companies and this is truer in mechanical Small and Medium Enterprises (SMEs). Nevertheless the implementation of a PLM system is not an easy thing for such companies. SMEs have lots of difficulties to create the right information model in order to improve design efficiency. In this paper we propose a PLM model that could help SMEs in the mechanical engineering field. Based on a literature review and on the analysis of the mechanical SMEs needs that would favour efficiency during the design process, a new PLM model for mechanical SMEs is proposed.

Keywords:
Product Lifecycle Management, Product Process Organisation Model, Unified Modelling Language

1 PLM IN MECHANICAL SMEs
Product Lifecycle Management (PLM) is “a strategic business approach that applies a consistent set of business solutions in support of the collaborative creation, management, dissemination and use of product definition information across the extended enterprise from concept to end of life - integrating people, processes, business system, and information” [1]. It provides improvements in information search, reuse, change management, etc. And those functionalities are required by the mechanical SMEs according to a 2007 survey [2].

Mechanical SMEs have difficulties to implement PLM systems. As explained by [3], the main difficulty is the modelling of the information and the formalisation of the processes. Getting the correct data model is a key for successful implementation of a PLM system. The use of the right model significantly improves the chance of success in obtaining the right data model.

The SMEs have more difficulty than the other in PLM modelling. One explanation is that they have fewer competencies in modelling than the bigger firms. The second explanation is that they have substantially different needs in terms of PLM that are not taken into account in the today’s PLM models.

The literature proposes different models that could fit the PLM modelling requirements. In section 2, after analysing different models, a new PLM model is proposed. In section 3, the analysis of this model is done through the functional requirements of mechanical SMEs. Section 4 briefly presents the integration of this model in a PLM demonstrator. The paper concludes with perspectives and discussions.

2 PLM MODELS
A PLM model defines and structures the information concerning a product during the entire lifecycle and through the extended enterprise. Since the 80’s, the concepts necessary in product information management systems evolved from Product, to Product/Process, and more recently to Product/Process/Organisation. The study is done on six Product/Process/Organisation models: NIST [4], Patterns [5-6], GRAI [7], FBS-PPRE [8], STEP AP 214 [9] and AP239 [10], IPPOP [11] and MOVES [12]. The NIST proposes a product information model including geometry, structure and assembling, and tolerances. PATTERNS propose different patterns for information reuse, like a product model or a process model. FBS-PPRE proposes a common object for Product Process Resource and External effect, with Functional, Behavioural and Structural views on this object. GRAI is more decision oriented. It distinguishes two kinds of activity: decision activity and execution activity. Those activities are managed by decision centres. IPPOP integrates in a Product Process Resource model at an organisational level. STEP AP 214 is a model from STEP dedicated to the automotive industry. STEP AP239 is more aeronautic oriented and covers the whole lifecycle of the product. MOVES is a Product/Process/Organisation model that focuses on human resources modelling. Those six models have been constructed together with large firms, mainly from the automobile and aeronautical sectors. They are analysed from Product, Process and Organisation point of views.

2.1 Product
The first word in the PLM acronym is “Product.” It was also the first model for information management. The Product models describe the structure of the products. Gamma [5], then Gzara [6] proposes a decomposition of a product in composite product and in elementary product. Krause [13] and STEP AP329 separate virtual...
product from real product. FBS PPRE and IPPOP propose functional, structural and behavioural views of the product. STEP AP 214 and AP239 propose to distinguish the product, the product version and the product view definition. They also propose different breakdown structures. This concept of version is also presented in IPPOP.

Moreover the functions are linked to the product in Patterns, FBS-PPRE, IPPOP and NIST. Patterns and FBS-PPRE distinguish technical function and service function.

So the product is composite, versioned, and could be virtual or real. But from a PLM point of view, the differentiation between virtual and real products does not make sense because this system manages only virtual products. The product is linked to the function.

2.2 Process

According to a product life cycle view, the process view is obligatory. It integrates the different stages of the life cycle.

The process and the activity are objects of the same type for Patterns, FBS-PPRE and Moves. A process is a composition of activities that can be considered as processes and decomposed in sub-activities. STEP AP 214 and AP 239 link Activity concept to Product and Resource. Patterns, IPPOP and FBS-PPRE link the product as input and output of the activity. The product changes from one state to another through the activity. The activities are linked together through a sequential link.

So the Process is a composition of activities. The activity is linked to the product and to the resource. The link of the activity with the product is input/output.

The resources of the company have to be modelled. In Patterns, the resources include material and human resources whereas STEP AP 214 and 239 do not include human resources in the resource class but in the organisation class. More recently the immaterial resources are also modelled, in IPPOP by adding informal or methodological resources, or, as in MOVES, by adding a link between resource and activity representing the capacity.

2.3 Organisation

At last, in the actual global context, the organisational view becomes more and more important. GRAI distinguishes two kinds of activity: decision activity and execution activity. Those activities are managed in decision centres. IPPOP proposes an organisational point of view adding decision centres in the Product/Process/Resource views. The decision centres allocate resources and objectives to a project. In STEP, the organisation is linked to the human resources. MOVES method defines organisational entities that are responsible for actors and have authority on processes.

So the organisation is linked to the activity and to the resource.

2.4 Product Process Organisation model based on literature review

Based on the literature, we propose a simple model including the three views: Product, Process and Organisation. The product is linked to function and is an input and output of Activity. A combination of Activities from a Process, and they are launched by triggers. Resources are used by the activities, and are specialisable in human, material and methodological resources. The organisation is responsible for an activity. The class diagram is presented in Figure 1.

Combining the literature models is a first approach, but our goal is to obtain a model that will provide an improvement in the integration of PLM in mechanical SMEs. The models from the literature are mainly constructed with large companies and especially the design departments of these firms. The purpose of this paper is to obtain a model that will help the SMEs to implement PLM systems. These systems are shared with different departments, like design departments, but also manufacturing and purchasing departments. In the next section, a needs analysis of the different departments in SMEs is done to modify and improve the model extracted from the literature.

Figure 1: Model created from the literature review.

3 FUNCTIONAL REQUIREMENTS

To get the functional requirements of mechanical SMEs in terms of PLM, the proposition is an inductive approach. The modeller is immerged in different SMEs to understand their processes, the information flows, and the communication in the company and outside the company. The functional requirements are based on immersions done in representative SMEs [14]. The needs analysis is done by interviews and document analysis of the experts, by modelling their needs and by validations with the experts. To choose SMEs representatives from the mechanical field, a typology of those companies is proposed in [15]. The typology identifies three kinds of mechanical SMEs, the assembly manufacturers, the component manufacturers and the elementary part manufacturers. According with this typology, the three companies in our study are respectively a special machines manufacturer, a sailboat rigging manufacturer, and a crankshaft manufacturer. In each company, the design department, the manufacturing department and the procurement department were studied.

3.1 Design

The design department designs the product sold by the company. The design departments that were considered by our study were the one of a special machine producer and the one of a sailboat rigging manufacturer. The special machine design department already has a PDM system but needs improvements. The sailboat rigging design department does not have any information management system. The captured requirements are listed above.
Product version
The versioning of a product allows the change of the product to be followed. The modified model includes a ProductVersion class, specialisation of the Product class. The different versions are linked by nextVersion associations. To be able to know the state of a version (in change, validated, obsolete, etc.), a state attribute is added to the other attribute of the ProductVersion class.

Product documentation
The documents associated with a product are many and varied. They can be CAD files, textual specifications, concept drawings, etc. A Document class is added and linked to the Product class.

Internal and external product
The policy, the right access, and the management of the product are different if the product is a finished product sold by the enterprise, an internal product managed by the company or a product bought from a supplier. The attributes of those products are also different. The modified model proposes to specialise Product in FinishedProduct, Component and RawMaterial. Notice that this view is a system view. The raw material could be a sub-system, if the sub-system is made by a supplier and the finished product could be an elementary part if the company sells elementary parts.

Design Bill of Material (BoM)
The BoM is an essential document for the design department, especially when designing complex systems. The solution proposed in the model is to link the Product class with itself. So the product is composed of other products (that can be components or raw materials).

Optional product
Optional components are components that are not obligatory in the construction of a product. A component could be optional in one product and obligatory in another. To distinguish the optional part from the core part of a product, the model adds a class to the product-product link that describes if the composition product-component is optional or not.

Alternative product
An alternative product is a component that could replace another component in a product. The alternative product does not change the product function, as defined in [16]. A ProductAlternative class is added to the model, linked to the Product class as alternative and context. The context link describes the assembly in which the alternative is validated whereas the alternative link describes the two interchangeable products.

Material
The material of a product is needed by the elementary parts manufacturer. To know the material of a product, a Material class is added and linked to the Product class.

From the design department point of view, all of the necessary changes of the model presented in Figure 1 are related to the Product class. Figure 2 summarizes the modifications made.

3.2 Manufacturing
The manufacturing department prepares the process planning of the products sent by the design department or by a customer. The manufacturing departments were a sailboat rigging manufacturers and a crankshaft manufacturer. Both do not have a dedicated system for information management. The captured requirements are the following:

Operation version
The manufacturing department has different standard operations that evolve with the time. A versioning of operations allows them to follow the different modifications of an operation. In the model, the operation is a sub-class of Activity. An ActivityVersion class is added to the model. The different versions are linked by nextVersion associations and a state attribute is added to the ActivityVersion class like in the ProductVersion class.

Operation documentation
Those operations are documented by different files, like CAM file. So the Document class is also linked to the Activity class.
**Supplier of a product**

The supplier of the bought products must be identified. So the *Organisation* class is linked to the *RawMaterial* class.

**Alternative supplier**

They also need to know the other suppliers capable of supplying a product or an alternative product. An *OrganisationAlternative* class is added to the model. It is linked to an organisation in the context of the supplied product.

**Supplier of an operation**

The procurement department does not manage only product procurement. It also manages outsourcing operations, like in the crankshaft company. In the model the *Organisation* is linked to the *Activity*.

**Supplier of a resource**

At last, the procurement department can also punctually rent resources, to respond to a customer order for example. The *Organisation* is also linked to the *Resource* in the model.

**Figure 4** is the class diagram of the organisation.

### 3.4 Extended enterprise context

In an extended enterprise context, the companies have to exchange data to collaborate on all of the phases of the product lifecycle (from the early design phases with co-design to the disposal phase with the creation of a recycling plant, through the manufacturing phases with the supply chain). A finished product for a company becomes a raw material for another. Even a product for a company could become a process or a resource for another. For example, in the exchange between a special machine manufacturer and its customer, the product for the manufacturer (the special machine) becomes a resource when in the factory of the customer.

**Figure 5** is the class diagram of the organisation.

### 3.5 Product/Process/Organisation model

Combining the different views presented in this section, the Product, Process, and Organisation are represented Figure 6.

### 4 IMPLEMENTATION OF THE MODEL

The proposed model is implemented in software based on a relational database. The software developed has client-server architecture. The client is developed in VB.Net and the server is based on MS SQL Server database. Figure 7 shows the interface of the software, constituted of a tool bar (1), a tree view (2), a file view (3),...
a data card (4) and a viewer (5). The following functionalities are implemented in the software:

**Documentation**
Each object is documented. A drag and drop from the desktop to the software interface adds the document to the database. A double click on the document opens it.

**Versioning**
Every object is versionnable. A right click on the object and the selection of “new version” creates a new version of the object. It is also possible to retrieve the past versions.

**Design/Manufacturing BoM**
Different views are available. The design view and the manufacturing view give two different BoMs for the same product depending of the selected view.

**Operation**
The user can define optional operations, alternative operations and external operations (appearing in green)
Then he loads the resources useful to each operation, dropping the input items, the work centres, the tools, etc.

Suppliers

The internal and external products are in the referential of the enterprise with different colors (yellow for the internal products and green for the external products). The external product can be taken directly from the supplier PLM systems. The different suppliers identified are present in a specific folder. A drag and drop from a component of the supplier to the software will import the product, all the attributes and the attached files.

The tool has been tested on three industrial scenarios to validate the proposed functionalities. The variety and representativeness of the chosen SMEs are ensured by the mechanical companies' typology. The validation scenarios also include departments not taken into account by the study to construct the model, such as the procurement department of the sailboat rigging manufacturer or the manufacturing department of the special machine manufacturer. All of those points ensure a level of generality to the model.

5 DISCUSSIONS AND PERSPECTIVES

This paper proposes a PLM model for design information management in the mechanical field. The method of construction is a two-step approach. The deductive phase is based on a literature review to create a reference model. The second one is an inductive step to modify the model depending on the functional mechanical SMEs requirements.

The main result is a more complex enterprise object that enlarges the possibility of management on the activity, resource and organisation classes. This could be explained by the source of the literature PLM models and our specific study. PLM systems first aim to manage information from the product. So the design department is well represented in the specification of the data model. But other departments also use PLM systems, and their requirements are less taken into account. The fact that some of our case studies do not have a design department (only manufacturing and procurement services) underlines that imbalance between the design department and the others in the PLM specifications.

In future work, we would like to include more steps of the lifecycle (maintenance, end of life, etc.) and more departments' requirements in the PLM model. The SMEs that manage those phases should have a different point of view on a PLM that must be included in our model.

6 REFERENCES


Method Applying System Models as Catalog Entities for Emergent Motion Design

T. Brett
Department of Machine Tools and Factory Management
Chair of Assembly Technology and Factory Management, Technische Universitat Berlin, Germany
brett@mf.tu-berlin.de

Abstract
A method for the conceptual design of motion is proposed. State of the art dynamic system models from mathematics and physics are applied as a pool for the retrieval of knowledge analog to physical effects in design catalogs. Behavior attributes of several pool models are implemented to new motion of a single physical artifact model, carried out at a single time interval. Deduction is applied to indicate that the derived concepts feature the desired attributes as well as new emergent qualities. The effectiveness of the method is demonstrated in the design of a handling robot concept with improved energy efficiency that is a factor of 10 better compared to conventional systems.

Keywords:
Emergence, Dynamic Systems Modeling, Equipment,

1 INTRODUCTION
Typical motion design in industrial robotics and service robotics follows two different strategies. In industrial robotics, a mathematical law for path planning is appreciated ensuring a successful deduction of the law for motion tasks between arbitrary initial and final coordinates in the robot’s workspace. Path planning must be calculated at low computation effort and should always be successful. In service robotics, specific motion behavior is intended such as walking [1], running [1-2] or hopping [2]. Besides availability and control of motion, it is a great challenge to identify purposeful motion behavior in the search space of system dynamics. Research in motion control led to some discoveries of new motion behavior, such as the work presented by Hussong et al. in [3], but a systematic analysis of the search space is not in the focus of the community. The contribution of Hussong et al. is basically software which, implemented for an arbitrary robot of a class of robot types, enables the system to carry out a motion behavior which is characterized by efficiency in power consumption. Software development and mechanical design remain distinct disciplines without reason to merge them.

In dynamic systems modeling, motion behavior is a subject matter. An evolution rule for the change of system states over time is often formulated with a few equations and the initial state of the system is determined with some parameters. Behavior, observed in the phase space, is described with mathematical formulations for terms such as periodic motion [4], attraction [5], synchronization [6] or stability [4]. Dynamic systems modeling is often applied to address mathematical and physical problems as being a distinct discipline. Models of mechanical systems need elaboration following conventional design methods for physical realization. A lot of simple models have been developed and together they give a good representation of many different aspects of the search space of system dynamics. Other than the laws of physical effects, motion behavior is often more fuzzy in its parameter range and in its description. A quasi periodic motion, for example, falls back into the initial state, but not at full precision. Without limiting the length of the time period, even a model of a chaotic system could shape this behavior. Thus, the simple dynamic systems are distributed throughout the search space, but in between these models, a combinatorial exposition of fuzzy behavior prevails. To identify fuzzy behavior which is purposeful for industrial robotics, designers need to evaluate a large number of solution principles, considering both aspects of mechanics and of motion behavior.

An effective design method for motion behavior could foster further discoveries in the field. To obtain a new quality in motion behavior which goes beyond stringing together known motion sequences, emergence of something new is mandatory. For a method to be effective, a systematic approach is required which supports the designer in identifying emergent behavior and matching the new qualities to the demands of an appointed group of people. For that, four criteria for

![Figure 1: The stokes wave is an emergent behavior on the global level which cannot be anticipated from individual parts.](image)

CIRP Design Conference 2011
emergence are defined which amend the general explanation for the whole being more than the sum of its parts. First, it does not reject underlying theories, but complements them with new principles of organization or new laws [7]. The formation of the Stokes wave [8] is an example for new behavior of H2O molecules when aggregated to deep water. Information on the Stokes wave is not obtained from particle physics. Second, for emergence in synthesis, an actual instance of an object of a higher order must be met before the emergent law can be discovered [9]. For the Stokes wave, it could be the reckoning of wave formations in deep water. Third, for analysis in natural science, emergence can be perceived as a metaphysical principle of protection [10-11]. The molecules of a fluid could be replaced by another chemical element without affecting the shape of the waves. Fourth, emergence is restricted by the underlying laws of particle physics [12] such as by the fact that the weight of water cannot exceed the sum of the weight of its molecules. These four criteria are defined as:

1. Formation of new laws (see section 4.2)
2. Synthesis problem of emergence (see section 3)
3. Analysis problem of emergence (see section 3)
4. Limitation of laws due to the fundamental dynamics (see section 4.1)

2 EFFECTIVE MOTION DESIGN

Many artifacts reveal their functionality and purpose from just viewing a picture of it. Geometrical properties are recognized from the picture and the observer assigns the object into a class of practical relevance such as transportation unit, measurement device or sports equipment. Besides the identification of geometrical properties as a result from the extrapolation of the two dimensional picture into a three dimensional object, the observer is often able to identify the relevant motion which is related to the artifact. The observer may identify a round geometry as a wheel before or after he has assigned it to the class of transportation units. If the artifact is a train, the motion of the wheel in relation to the steel line is identified as a simple rolling motion. Though the wheel motion varies in speed and acceleration, the attribute of rolling characterizes well the whole solution space of possible wheel motion in the regular operation mode. If the artifact is a car, it becomes more delicate to assign an appropriate attribute to the motion behavior of the wheel. In figure 2, the observer assigns the correct wheels e, h and i to the corresponding cars in motion as e/c, h/k, i/l.

The inverse perspective from the analyzing observer is the problem perspective of the designer who synthesizes the structure. He seeks to create a wheel which satisfies the requirements resulting from a special motion behavior by means of synthesis. For the snow rally car of figure 2k, a maximum traction for turns is intended. For the top fuel drag racer in figure 2f, maximum acceleration on a straight line is intended. For the energy-efficient car in figure 2c, a maximum distance is aimed for a given amount of energy. Thus for the wheel designer, the geometrical properties are correcting variables to be optimized to improve motion parameters which are the affected variables. The purpose of the artifact in creation is to meet the driver’s preferences in performing a specific motion behavior. The picture observer identifies the related motion from the analysis of the geometrical properties due to his experience in mechanics or in motorsport. Synthesis and analysis lead to the same relation of each one kind of tire to one kind of motion behavior. However this 1:1 relationship is not absolute. It results from the disposability of problem related knowledge. The engineer benefits from his experience in car dynamics. Mental projections are valuable both for the designer and for the interested observer. The mental projections of the engineer, however, may be of a more generic nature.

For some pictures of artifacts, observers cannot identify the substantial motion which is related to the artifact. Figures 2b, j and m show three clockwork mechanics. It is not obvious which mechanics corresponds to a standard clock motion of figures 2a and d and which mechanics corresponds to the specific motion of the astromonic moon calendar of the town-hall clockwork in Prague, figure 2g. The mechanics are a composition of numerous parts. A technical drawing reveals the motion after an intensive analysis to the experienced observer. Consequently, the large compositions of figures 2b, j and m are a challenge to the observer, but not a barrier.

For designers, various alternative mechanical concepts enable them to realize the standard clock motion. Systematic design as presented in the state of the art reduces the solution space to a few favorable variants. The clockworks of Bregu, London (figure 2d) and the Zylogge, Bern (figure 2a) are examples for competing solution concepts. If each clockwork solution represents a single set of graphical properties, a number of new quality of performance. A large number of different motion behavior \( m_{mp} \) is presented with the same sports equipment properties, resulting in the relationship 1: \( m_{mp} \), where \( m_{mp} \) is any large number. However, if the time period under review is increased, it is recognized that the geometrical properties of the equipment developed over time as well, indicating a relationship of \( n_{gps} : m_{mp} \).

Figure 2i shows the evolution of snowboarding from the Snuffer board on the left which originated in 1965 until the emergence of the modern board properties around the 1990s, following an \( n_{gps} : m_{mp} \) development. Since then, new motion patterns have been developed with pace in innovation. Though half pipe snowboarding became an Olympic discipline in 2002, new motion behavior has been presented at Olympia in 2010. Development of the 1: \( m_{mp} \) type took over 20 years and is still ongoing.

For many 1: \( m_{mp} \) and \( n_{gps} : m_{mp} \) relationships, the problems of combinatorial explosion, insignificance between true and wrong behavior descriptions and small parameter spectra occur. For some 1: \( n_{gps} \) problems are also persistent, but only if the description of the motion behavior \( m_{mp} = 1 \) is available in an abstract manner. Figure 3 shows on the left side the barrier problem of emergence between motion and geometry for these problems of motion design. The other aspects are discussed in the following.

3 MACRO-CAUSATION AND MICRO-COMPOSITION

Macro and micro are often applied as synonyms for the global and local level in emergence. For this work, macro-causation and micro-composition are derived from
macroeconomics and microeconomics in order to sharpen the meaning of these words. In microeconomics, the decisions of households and enterprises as well as the impact of these activities on markets are analyzed. Macroeconomics addresses economic phenomena as a whole on the aggregated level, particularly inflation, unemployment and economic growth [13]. The principles of the methodological individualists [14] are found in microeconomics as compositions of the individuals actions. In macroeconomics, the principles of the methodological holists [14] are found where cause and effect relations on the collective level are integrated into laws and principles of organization. In the following, macro-causation and micro-composition are not reduced to economics but applied for all social, economical and ecological aspects to differ between a scope view on group behavior and the view on the actions of individual persons.

Energy-efficiency in manufacturing is a good example for the two viewpoints. On the macro level, the reduction of carbon dioxide emissions and the limited non-regenerative energy sources are predominant. The political demand for the development of efficiency technologies is a result from reasoning in terms of cause and effect on the macro level. On the micro level, scientists develop conceptual ideas for efficiency improvements, but a causal reasoning on the impact of these technologies on the macro level is often still unpredictable, as the diffusion process on the market is a complex process. As probably intended, inefficient technologies could become replaced, but customers could also react in a way that the new product increases only the number of energy users. For equipment, Weinert proved that the most efficient technologies must not lead to the most efficient process [17]. A barrier disturbs the chain of causation equal to the barrier to causation between macro- and microeconomics, see figure 3.

4 APPROACH FOR EFFECTIVE MOTION DESIGN
The method is intended for application in academic research to develop sets of motion behavior for production systems. The only triggers for design are macro goals. Micro requirements for the artifacts to suit their application fields are supplemented during the design process. The approach offers three major advantages. First, emergence of an innovation is enabled without facing the synthesis problem of emergence. As illustrated in figure 3, both the goal formulation and the intended solution are developed on the abstract and global level. The designer does not have to cross the barrier to causality to gain a new inventive idea. As the second advantage, the development is initiated without market restriction to focus on facing the macro goal rather than a particular market segment. The maximum size of the solution space of motion is open to the designer to develop behavior with outstanding properties to meet the macro goal. Further, the method qualifies the designer to apply the analysis problem of emergence as an advantage to design and thus, benefitting from the barrier instead of challenging it. As is characteristic of analysis problems, different compositions of parts lead to the global behavior. While particle physicists cannot identify the true underlying particle composition, the designer facing an engineering task proposes a global motion behavior and, due to the barrier, he can chose one among multiple graphical property sets to create the part composition on the local level. Consequently, the third advantage is the analysis problem of emergence itself. The problem identification on the macro level initiates the design process, see figure 4. The designer chooses a problem which, to his belief, should be solved by facing the complexity of \( n_{gm}:m_{gm} \) relationships. In the next phase, he directly starts to develop indicators and benchmarks to quantify the effectivity of design during the whole design process.

4.1 Indicators and Benchmarks
Formulation of indicators for design should be developed in an early design phase. One reason is that effectiveness requires a project goal formulation. Another is that emergent thinking should be supported with a quantification of the limitations of emergent laws, see section 1.
In a first step, the relevant parameters for the macro goal are identified and their theoretical extremes are analyzed. The example of energy-efficiency in manufacturing is taken up for that. A fast analysis of manufacturing shows that parameters cannot be identified which are of relevance both for the macro goal and for motion behavior. The search space is reduced, leading to a focus on handling equipment with promising results. For energy-efficiency, energy itself or the quotient of energy and time which is power consumption can be applied, although they may lead to different approaches. Knowledge of the initial and final energy level is sufficient to determine the minimum required energy for some processes such as thermal processes in food processing. For material flow processes, the initial and final energy level of a product is generally equivalent if the initial and the final condition is defined as rest in a storage neglecting potential energy. Yet, there is substantial power consumption detectable for material flow. The power consumption induced to the payload as kinetic energy is chosen as the indicator of major concern for this concept phase due to the contrast between theoretical minimum and the consumption of state of the art equipment.

The physical law of power $P(t) = F(t) \cdot v(t)$ and velocity $\dot{v}(t) = ds(t)/dt$ result in

$$P(t) = \frac{\int F(t) \, dt}{at} = F \cdot v.$$  \hspace{1cm} (1)

A pick and place motion as the handling operation from an initial to a final point in the time interval $\tau$ consisting of a constant acceleration interval with $a = y = \text{const}$. $\forall t = [0 \leq t \leq \tau/2]$ and a constant deceleration interval with $a = -y \forall t = [\tau/2 < t \leq \tau]$ is defined. Disregarding the discontinuity in $\tau/2$, for each interval $F = \text{const}$. is obtained from Newton’s basic equation and a proportionality of

$$P \propto v.$$  \hspace{1cm} (2)

The payload speed of this benchmark motions results for the first interval from (4) as

$$v(t) = \int_0^\tau a \, dt = a \cdot t \quad \forall t = [0 \leq t \leq \tau/2]$$  \hspace{1cm} (3)

with a maximum in $\tau/2$. From (1) and (3), the maximum power consumption results as

$$P_{\text{max}}(\tau/2) = \frac{m \cdot a^2 \cdot \tau}{2}$$  \hspace{1cm} (4)

for indicator calculations with default pick and place time. If only the pick and place distance $x$ is known, the first pick and place interval distance is

$$X/2 = \frac{1}{2} a \cdot \frac{\tau^2}{2}$$  \hspace{1cm} (5)

resulting in

$$P_{\text{max}}(\tau/2) = m \cdot a^2 \cdot \sqrt{\frac{x}{a}}$$  \hspace{1cm} (6)

The average power is

$$P_{\text{avg}} = \frac{1}{\tau/2} \int_{\tau/2}^\tau m \cdot a^2 \cdot dt = \frac{P_{\text{max}}}{2}.$$  \hspace{1cm} (7)

The calculated values enable handling performance to be illustrated according to figure 6.

In figure 6a, the motion cycles of the state of the art are illustrated with the power consumption single-body-system benchmark. The diagram shows the typical horizontal motions of fast delta handling robots. They require in theory less than 0.5 kW effective power for the payload and the robots have an average capacity of less than 2.5 kW in the state of the art. Thus, these robots are not suitable for the motions according to figure 6b which require more than 5 kW effective power for the payload. The state of the art serial kinematics industrial robots are also not expected to handle these motions as they typically have a self-weight of more than 350 kg. An average acceleration at 10g is not possible for such industrial robots in general. Thus, figure 6b illustrates a technological gap which cannot be solved with typical industrial robots.

In contrast to the technological gap resulting from high acceleration requirements, figure 6c illustrates another technological gap resulting from low acceleration requirements. For these motions, the electrical energy sent to the robot mechanics is not of great relevance, as the power converters consume much more electrical energy for idle operation. Since industrial robots feature typically the same number of servo drives as provided degrees of freedom, a four axes industrial robot requires much more than 200W. Thus, a technological gap results in enabling handling at four degrees of freedom at approximately 100W. Regarding the mechanical system, the obvious approach to achieve that goal is to reduce the number of servo drives. Consequently, a second indicator for efficiency for a second motion behavior is the number of applied servo drives. As the number of servo drives influences the product cost, it is an indicator for energy and economic efficiency for process design.

4.2 Dynamic Systems Modeling

Many contributions to dynamic systems modeling (DSM) follow the intention to increase the pool of scientific knowledge, as is usual in the more fundamental sciences. Effects derived from this intention can be applied as lookouts points for a search in the motion space with motivation for creating artifacts with a specified purpose, as indicated in figure 5. Motion behavior, formulated with simple DSM could be aggregated in design catalogs. Fundamental behavior can also be fused to result in a new behavior of a new quality. The example of energy-efficient handling equipment is taken up again to illustrate, how purposeful behavior is identified.
Low energy consumption is the macro goal of the design. Short cycle times and short rest position for gripping products are identified as relevant for handling in the market segment. The fundamental effect of periodicity enables a minimum of energy consumption, but the system does not rest. The fundamental motion behavior of the control motion on the other hand enables to rest a single point of a robot, while the other parts of the robot maintain motion. The fundamental motion behavior of momentum transmission enables high acceleration. How would a trajectory featuring all three behaviors look and which graphical properties of the robot should be chosen for that? Applying DSM, simulations are performed for every fundamental behavior to identify graphical properties suitable for all of these behavior properties. It is illustrated in the example of periodic motion.

In figure 7, the double pendulum in the rotational field is illustrated. It is a serial kinematics of three links. The derivation of Weidemann presented in [18] applies in principle, but the dynamics of the second and third link shall not affect the constant speed of the first link.

The angle of the first link is not relevant for the simulation and the acceleration of the link is set to zero, such that

\[ q_1(t) = \delta, \quad \delta = 0. \tag{8} \]

The rotational field is determined by a kinematics and a dynamics parameter. It is the link length which is set to

\[ l_1 = 1 \text{ and } \delta. \]

The states of the dynamic system result according to figure 7 as

\[
\begin{bmatrix}
    x_{c1}(t) \\
    x_{c2}(t) \\
    x_{c3}(t) \\
    x_{c4}(t) \\
    x_{c5}(t) \\
    x_{c6}(t)
\end{bmatrix} =
\begin{bmatrix}
    -M_{d2} \cdot \dot{\theta} + M_{d1} \cdot \delta \\
    -M_{d2} \cdot \dot{\theta} + M_{d1} \cdot \delta \\
    -M_{d2} \cdot \dot{\theta} + M_{d1} \cdot \delta \\
    -M_{d2} \cdot \dot{\theta} + M_{d1} \cdot \delta \\
    -M_{d2} \cdot \dot{\theta} + M_{d1} \cdot \delta \\
    -M_{d2} \cdot \dot{\theta} + M_{d1} \cdot \delta
\end{bmatrix}
\tag{9}
\]

Friction is not integrated to the model. For the illustration of the model, the state space is formulated in Cartesian coordinates for the position of E as:

\[
\begin{align*}
x_{E,1} &= l_2 \cdot \sin(q_2) + l_3 \cdot \sin(q_2 + q_3) \\
x_{E,2} &= l_2 \cdot \cos(q_2) + l_3 \cdot \cos(q_2 + q_3).
\end{align*} \tag{10}
\]

Figure 8 shows how the periodic motion is obtained for a well selected initial state. Quasi periodic motion is also of value for the engineer. It is in the visibility range of periodic motion theory. The left example shows no purpose from the viewpoint of periodicity. It is out of the visibility range.

4.3 Optimization

The simple algorithms of DSM applied to the global attribute synthesis in the previous section need to be replaced by a more precise model for optimization. A lot of micro-requirements have already been implemented to
the concept but little information about the local parts is yet available; mainly just the kinematics information about robot links. Thus, at least it is known that the development is for a robot system. The designer can apply an arbitrary modeling approach for robot dynamics. The challenge of the optimization phase is to create suitable program code for the path planning. It must be flexible for testing various alternative patterns of motion until motion emerges comprising the behavior of periodicity, nullspace motion and momentum acceleration. As an example, an evolutionary optimization algorithm has been chosen. About 50 significant changes to the code have been carried out, until the intended behavior resulted. Graphical properties have been adapted simultaneous to the code adaptation, see [19]. Expected effects, due to changes of the mechanical properties, have often been falsified. The dynamics of the three coupled links in complicated motion are hard to predict by mechanical reasoning. For most code setups, a small parameter spectrum showed the intended results. Figure 9 shows the motion and figure 10 illustrates the results with the final code. A comparison to the single-body-benchmark asserts that the motion behavior enables a high increase of energy-efficiency which is a factor of 10 in relation to the benchmark while common industrial robots are worse than the benchmark. From the figure, further information relevant for handling on the micro level is obtained referring to payload, workspace and acceleration.

Figure 5: Applying fundamental behavior as a viewpoint for purposeful behavior.

4.4 Refinement
The refinement phase focuses on the micro requirements addressing specific market segments for the artifact. For the elaborated example, the market of solar cell handling has been identified. Additional requirements are a constant acceleration and deceleration of the product and a direct path of the handled product between the grip and the release coordinates. Experience from the previous design phase and a motion data management system are valuable for this phase. The database should comprise both trajectory data and the corresponding optimization code data. For the presented robot concept, a good adaptation to these requirements could be achieved.

4.5 Deficit compensation
For building a set of motion behavior, synergy effects should be utilized. Motions in a set should be able to compensate each other's deficits. The deficit of the designed motion is that short distances in the workspace cannot be traveled. A second artifact should be built in to compensate for this problem, addressing the indicator shown in figure 6c. Other than in the example of Aikido, the set of motion could be developed to be carried out with different artifacts working in collaboration.

5 CONCLUSION
The approach of effective motion design presented in this paper applies motion behavior from dynamic system models analog to physical effects from design catalogs. Optimization was applied to implement the three behavior attributes of periodic motion, underactuated motion and nullspace motion into a single system. The system features the emergent quality of being able for handling goods consuming less actuator power than kinetic power induced to the good. This new quality is emergent as information about this quality is not contained in the original models. It is purposeful for the market of industrial robotics as simulations reveal that the energy consumption is improved by factor 10 compared to conventional robots. The example indicates that the systematic of the proposed method could enable to attain emergent qualities which have been quantified by indicators and benchmarks at the beginning of the design. In further research, more design examples need to be elaborated to prove by means of deduction the significance of the method for emergent design.

6 REFERENCES

Figure 6: Power consumption indicator applied for a state of the art evaluation (a); for the identification of a class of high speed motions which cannot be realized with industrial robots (b) and a class of motions where the electrical losses in the electrical cabinet dominate the overall consumption (c)
Figure 8: Double pendulum in the rotational field performing chaotic motion (left), quasi periodic motion (middle) and periodic motion (right).

Figure 10: Exemplary simulation results of the optimization phase.
Design of Wireless Electric Power Transfer Technology: Shaped Magnetic Field in Resonance (SMFIR)

N. P. Suh
President, KAIST
291 Daehak-ro, Yuseong Gu, Daejeon, Republic of Korea, 305-701
npsuh@president.kaist.ac.kr

Abstract
Major advances in the development of electric vehicles (EV) – buses, cars, high speed trains, etc. – have been made using a new wireless power transfer technology named the “Shaped Magnetic Field in Resonance” (SMFIR) that can send electric power over a significant distance. SMFIR is a basic technology that can be used in many applications where electric power has to be supplied remotely without using conductive wires. The “On-Line Electric Vehicle” (OLEV) is a SMFIR-based electric vehicle, which was chosen as one of the 50 Best Inventions of 2010 by the TIME magazine. In this paper, the underlying design for SMFIR is presented, which was done applying the Independence Axiom of Axiomatic Design Theory.

Keywords:
Axiomatic Design, electric vehicles, OLEV, SMFIR

1 INTRODUCTION
Historically a series of major innovations have sustained the economic growth of the world since the Industrial Revolution. Most important steps involved in major innovations are: (1) problem identification, (2) definition of functional requirements (FRs) that a solution for the identified problem must satisfy, (3) design of the product to satisfy the FRs by through the identification of the design parameters (DPs), (4) actual implementation of DPs in the final system using chosen process variables (PVs), and (5) commercialization of the product. Creative design and implementation are the basis for successful innovations that can spur economic growth [1-2]. This paper presents one such innovation.

Since 2009, KAIST has been involved in the invention and innovation of two new green transportation technologies: Mobile Harbor (MH) and On-Line Electric Vehicle (OLEV). These basic inventions were possible because of the “Axiomatic Thinking” process, i.e., application of the Independence and the Information Axioms. Furthermore, the rapid conversion of these inventions to innovations was possible, because the detailed design and implementation of these large complex systems followed a systematic and logical process prescribed by Axiomatic Design and because of the hard work of competent researchers and engineers. Also the importance of the large funding provided by the Korean government cannot be over-emphasized. Now a group of investors has formed “OLEV Technologies, Inc.” in the United States and the “OLEV Korea Company” is being established in Korea by another group of investors. If they succeed, the last step of the conversion process – from invention to innovation – would be completed.

This paper deals with the On-Line Electric Vehicle (OLEV). The basic concept of OLEV was presented as a keynote paper in the 2010 CIRP Design Conference in Nantes, France [3]. At the time, the details of the basic technology of wireless transfer of a large amount of electric power over a significant distance were not described because of the issues related to intellectual property rights (IPR). In this paper, the design of the basic wireless electric power transfer technology – the “Shaped Magnetic Field in Resonance” (SMFIR) – is described. KAIST has filed over 180 patents on this and related technologies. The purpose of this paper is to describe the basic design process and concept of SMFIR.

The goal of OLEV is to reduce the CO2 level in atmosphere by replacing cars with internal combustion (IC) engines with electric cars (EV). The interest in EVs has intensified in recent years in order to meet the regulations on the emission of CO2 and other harmful gases from IC engines. The need to reduce CO2 has been articulated by the International Panel for Climate Control [4].

Most commercially introduced EVs and vehicles with IC engines share one common concept: both carry the energy required to propel the vehicle on board of the vehicle: EVs use large batteries and cars with IC engines use a liquid fuel tank. Although R&D has been done to reduce the size, weight, and cost of the batteries (especially, lithium polymer batteries and lithium ion batteries) they will continue to be heavy, large, and costly. Today, the battery almost doubles the cost of an EV over typical cars with IC-engines. The basic idea behind the OLEV is to minimize the size of the battery and the distances that must be travelled using the battery by supplying energy to moving vehicles from an external source (an underground power supply system).

The OLEV project was initiated in 2009 under the sponsorship of the Korean government. This was a major research project for KAIST with an annual budget of about $25 million during the first year and $15 million during the second year. A commercial scale OLEV was developed, installed, and ready to take passengers within
a year after the project was initiated. KAIST is in the process of transferring the OLEV technology to industry, in addition to continuing research on OLEV and other "green" transportation technologies that uses the SMFIR technology.

2 DESIGN OF OLEV

2.1 OLEV FRs, DPs, and Constraints at the First Level

The highest-level functional requirements (FRs) of OLEV are as follows [3]:

FR1 = Propel the vehicle with electric power
FR2 = Transfer electricity from underground electric cable to the vehicle
FR3 = Steer the vehicle
FR4 = Brake the vehicle
FR5 = Reverse the direction of motion
FR6 = Change the vehicle speed
FR7 = Provide the electric power when there is no external electric power supply
FR8 = Supply electric power to the underground cable

Constraints are as follows:

C1 = Safety regulations governing electric systems
C2 = Price of OLEV (should be competitive with cars with IC engines)
C3 = No emission of greenhouse gases
C4 = Long-term durability and reliability of the system
C5 = Vehicle regulations for space clearance between the road and the bottom of the vehicle

The design parameters (DPs) of OLEV may be chosen as follows:

DP1 = Electric motor
DP2 = Power transfer device from the underground coil to the vehicle
DP3 = Conventional steering system
DP4 = Conventional braking system
DP5 = Electric polarity
DP6 = Motor drive
DP7 = Re-chargeable battery
DP8 = Electric power supply system (Inverter, etc)

2.2 Design Matrix (DM) for the First Level FRs and DPs

The Design Matrix (DM) for the first level FRs and DPs is given in Figure 1:

```
    DP1 DP2 DP3 DP4 DP5 DP6 DP7 DP8
FR1 X   X   0   0   0   0   0   0
FR2 0   X   0   0   0   0   0   0
FR3 0   0   X   0   0   0   0   0
FR4 0   0   0   X   0   0   0   0
FR5 X   0   0   0   X   0   0   0
FR6 X   0   0   0   0   X   0   0
FR7 0   0   0   0   0   0   X   0
FR8 0   0   0   0   0   0   0   0
```

Figure 1: First level design matrix for OLEVs

The first level design is a decoupled design. The design of FR2/DP2 is the critical phase of the design task.

3 DESIGN OF SMFIR (THE SECOND LEVEL FR2X, AND DP2X OF OLEV)

The first-level FRs and DPs given in the preceding section must be decomposed until the design is completed with all of the details required for full implementation. The second-level FRs are the FRs for the highest-level DPs and at the same time, the children FRs of the first-level FRs [1,5]. The basic core technologies have been created at this second level of FRs and DPs. In this section, only FR2 and DP2 are decomposed. The design that satisfies FR2 is known as the Shaped Magnetic Field in Resonance (SMFIR). It may be defined as a disruptive technology that can be applied to many fields replacing existing products [6].

The SMFIR technology that made OLEV possible transmits a large amount of electrical energy across empty space wirelessly and captures it at the other end. To develop SMFIR, we started out with the FRs for OLEV. In this paper, only FR2 and DP2 of OLEV are decomposed and designed.

3.1 FRs and DPs for SMFIR

FR2 and DP2 are restated below:

FR2 = Transfer electric power from underground power supply system to vehicle
DP2 = Power transfer device from the underground coil to the vehicle

Since FR2 cannot be implemented without further detailed design, FR2 must be decomposed. The second level FRs are obtained by decomposition of FR2 are as follows:

FR21 = Create a magnetic field above the ground
FR22 = Control the shape of the magnetic field
FR23 = Control power level of the magnetic field
FR24 = Pick-up the energy of the magnetic field from the vehicle
FR25 = Minimize the radiation of electromagnetic field, i.e., EMF

Figure 1 shows the conceptual design that satisfies the FR2s. The lower-level DP2 are as follows:

DP21 = Electromagnet design (ferrite core inside electric field)
DP22 = Magnetic pole design (L)
DP23 = Power level (strength) of the magnetic field
DP24 = Magnetic energy pick-up unit on the vehicle that is in resonance with the magnetic field
DP25 = Shield for stray electromagnetic field, i.e., EMF

The constraints that the design must not violate are the following:

C21 = Maximum allowable EMF level of 62.5 mG
C22 = Maximum weight of the pick-up unit
C23 = Electric shock resistance of the system
C24 = Temperature rise should not exceed 20°C
C25 = High magnetic permeability of the core material
C26 = Minimize the power loss

Some of these FRs and DPs such as FR25 and DP25 have been further decomposed but are not presented in this paper.
As shown in Figure 2, the magnetic field above ground is controlled by the underground power supply. The shape of the magnetic field is controlled to change the distance between the ground and the pick-up unit attached to the vehicle, H. The DP that controls H is the distance between the two magnetic poles, L. The alternating magnetic field is picked up by the pick-up unit mounted on the vehicle at the frequency of the magnetic field. The shield imbedded in the ground and the active and passive cancellation system mounted on the vehicle minimize the radiation of the magnetic field (EMF). This is the essence of Shaped Magnetic Field in Resonance technology.

The design matrix for the FR2s/DP2s is given by Equation (1):

\[
\begin{bmatrix}
    FR21 \\
    FR22 \\
    FR23 \\
    FR24 \\
    FR25 \\
\end{bmatrix} = \begin{bmatrix}
    X0000 \\
    XX000 \\
    XXX00 \\
    000X0 \\
    XXX0X \\
\end{bmatrix} \begin{bmatrix}
    DP21 \\
    DP22 \\
    DP23 \\
    DP24 \\
    DP25 \\
\end{bmatrix}
\]

The design is a decoupled design. According to Equation (2), all DPs, except DP24, affect the shielding of radiation. DP25 was designed to cancel the radiation due to other DPs.

The frequency chosen through simulation for optimum power transmission was 20 KHz, which minimizes the loss while maximizing the power transfer. The simulation was performed by Professor J. H. Kim and his research group [7]. The current supplied to the under ground cable to generate the magnetic field was in the range of 200 amps at a voltage of around 400 volts. The actual implementation of the design was done under the direction of Professor D. H. Cho, the director of the OLEV program, Professors C. T. Rim, and In-Soo Suh.

The OLEV system installed in Seoul Grand Park in December 2009 is shown in Figure 3. It replaced the noisy and smelly diesel system. The length of the circular path around the Park is about 2.2 km. The total length of the four-segments of the underground power supply system is 372.5 m. OLEV has a small battery on board to propel the vehicle even on roads without the imbedded power system. The battery is recharged when the vehicle is on top of the underground power system. The system was designed so that the charge on the battery remains about the same after completing each round.
power supply is segmented so that only the segment right below the vehicle is activated.

An integration team of the OLEV project constructed the design matrix for the OLEV system to identify and eliminate coupling between the FRs at several levels. The final design concepts were either uncoupled or decoupled designs. When there was coupling, its effect was minimized by making the magnitude of the element of the design matrix that caused coupling much smaller than other elements through design changes.

A given FR may have several different DPs. In this case, the final DPs were selected through modelling and simulation of the design using different DPs. The final values of DPs were also determined through modelling and simulations before the hardware was actually built.

4 APPLICATION OF SMFIR

SMFIR can be applied to many fields. We are trying to apply the SMFIR technology to bus systems first, since it is more readily adopted with a minimum investment in the infrastructure. As noted in the paper given at the 20th CIRP Design Conference, the overall economy of the OLEV system is attractive. At KAIST, we are conducting research to apply this technology to buses, passenger cars, high-speed trains, subways, appliances, and even to airplanes.

5 SUMMARY

The design of the Shaped Magnetic Field in Resonance (SMFIR) was achieved by applying the Independence Axiom. This technology, which enables the wireless transmission of a large amount of energy, has many important applications, including the On-Line Electric Vehicle (OLEV). This technology is being explored for other applications where wireless power transmission is required, ranging from low to very high kilowatts. This technology has been implemented in the OLEV systems deployed in the Seoul Grand Park and at KAIST.

6 ACKNOWLEDGMENTS

This project was made possible by major grants given to KAIST by the Ministry of Education, Science and Technology, the Ministry of Knowledge Economy, and the Ministry of Strategy and Finance. The author expresses his personal gratitude to President Lee Myung Bak, whose support made this project possible. He also wishes to thank Prime Minister Han Seung Soo, National Assembly member Yoon Jin Shik, and Minister Park Jae Wan for their support. The actual development and implementation of OLEV was done under the direction of Professor Dong Ho Cho by a team of 90 researchers and administrators. I would like to thank them all. Key faculty members of the OLEV team were Professors Joung Ho Kim, Choon Taek Rim, In-Soo Suh, and Hang Kee Lee, who provided technical leadership in executing this project.

7 REFERENCES

Application of Shaped Magnetic Field in Resonance (SMFIR) Technology to Future Urban Transportation

I. S. Suh
Graduate School for Green Transportation, KAIST
Yuseong Gu, Daejeon, Republic of Korea, 305-701
insoo.suh@kaist.ac.kr

Abstract
Transportation with electrified vehicles can reduce global dependence on fossil fuels and reduce the emission of greenhouse gases. Recent developments have been focused on the development of electric vehicles, hybrid electric vehicles and fuel cell vehicles. However, the commercial deployment of electric vehicles has lagged behind due to technological issues associated with the battery including: price, weight, volume, driving distance, and limited investment in charging infrastructure. Shaped magnetic field in resonance (SMFIR) technology enables electric vehicles to overcome these limitations by transferring electricity wirelessly from the road surface while vehicle is in motion. This work describes the innovative SMFIR technology used in the KAIST online electric vehicle (OLEV) project as well as its impact on the future of urban transportation. The system integration of the power supply into the OLEV and the vehicle system architecture is also discussed.

Keywords:
Electric Vehicle, Transportation, SMFIR, OLEV, Battery

1 INTRODUCTION
Most electric vehicles (EVs) get the electric energy needed for operation from on-board storage devices (i.e. batteries). However, current battery technology provides a very limited travel range with high costs, long charging times, and lower operating efficiencies due to battery weight. These issues must be addressed in order to increase adoption of EVs in both public and personal transportation.

KAIST has been challenging the technological limitations imposed by electric vehicle batteries by developing wireless power transmission technology that allows electric vehicles to charge during operation. This technology limits the need for remote static charging stations and replaces them with charging infrastructure embedded in the road or highway system. From a technical perspective, this strategy enables the designer to transform design constraints into design variables. This not only allows for the development of EVs with substantially smaller and lighter batteries, but also gives engineers greater freedom in designing the charging infrastructure and the on-board energy storage devices. It also allows the EV power management system to be more closely integrated with the electric power train. From a business perspective, this increases vehicle performance, user satisfaction, and business competitiveness, all while protecting the environment.

In this paper, we introduce an overview of the shaped magnetic field in resonance (SMFIR) technology that was developed as part of the KAIST online electric vehicle (OLEV) project. SMFIR is a technological innovation in wireless power transmission capacity and efficiency under dynamic operation. The design parameters and process of the dynamic charging infrastructure are introduced, starting with the required electric power, required wireless power transmission and the design considerations within the vehicle of the OLEV system.

2 BACKGROUND
2.1 Internal Combustion Engine (ICE) age to green transportation
The world is facing a tough challenge in the perspective of climate change and the global energy supply, mainly caused by a heavy dependence on fossil fuels. In 2007, the United Nations Framework Convention on Climate Change (UNFCCC) took initiatives on providing authoritative, timely information on all aspects of technologies and socio-economic policies, including cost-effective measures to control greenhouse gas (GHG) emissions [1].

While there have been active debates with mixed opinions on the global petroleum production forecast, the U.S. Energy Information Administration (EIA) scenario, published in 2002, shows that peak oil production will be reached within a couple of decades, as depicted in Figure 1. The EIA applied a growth rate of 1-3% in the petroleum production profile with the assumption of R/P=10, which means that the amount of known resources (proven reserves) has 10 years of annual production at the current rate of production to create the three curves in Figure 1. The peak annual global production of petroleum will be at its peak between 2030 and 2050 [2].
introduction of electric vehicles to public and private transportation. However, the introduction of those new vehicle technologies are in a limited market penetration phase compared with the growing public’s concern on the climate changes, which has been mainly caused by the lack of consumers’ acceptance yet.

2.2 Battery technology status

While launching EVs into the market, we believe that the battery is a technological barrier against consumers’ desire for a common transportation vehicle. The cost, packaging volume and weight of these batteries in EVs have been major issues, in addition to establishing governmental policies and charging the fuel (power) distribution infrastructure in a local region or nation-wide.

The Ragone plot in Figure 3 shows the positioning of various batteries and energy storage and generation devices in view of specific power density and energy density. For the pure electric vehicle or hybrid electric vehicle application, Li-ion batteries are promising with the current technology level in market. However, the specific power and energy capacity of the batteries are falling short of the IC engines’ capacity because of the limited driving range per charging and slow charging time. The United States Advanced Battery Consortium (USABC) set the requirements needed for batteries to be used in EV application in 2003 [5]. These requirements, summarized in Table 1, cover a wide range of issues and include energy and power denoted as HEV, PHEV and EV goal in Figure 3. For example, for the EV application, the required specific power and specific energy are 400W/kg and 200Wh/kg as marked in Figure 3. The current technology status of Li-ion batteries is not enough to meet the required specification.

The projection on the future battery pack price is also an important factor. According to McKinsey & Company’s report in 2009, Battery packs now cost about $700 to $1,500 per kWh, but it could drop to as little as $420 per kWh by 2015 under an aggressive cost reduction scenario as shown in Figure 4 [6]. However, the projection is based upon a survey from automotive industries including the assumption of a technological break-through for battery materials and productivity during 2015-2020. The projected cost of the battery packs in the near future will still be a significant portion of electric vehicle cost, which is a barrier to consumers’ buying pattern toward eco-friendly transportation.
OLEV system design are covered in Ref. [7] and [8]. Detailed discussion of the design process is beyond the scope of this paper’s discussion, but we introduce the outcome of implemented design in the following sections.

3.2 Introduction to SMFIR technology
SMFIR technology (also publically known as an OLEV system) enables the electric vehicle to be charged while the vehicle is in motion. The power cable installed under the road surface can generate a 20 kHz electromagnetic field as depicted in the Figure 5, when the cable gets 20 kHz AC electricity from the power inverter which is controlled under constant current output.

The power converter gets the electricity from the grid with the typical industrial power of 3-phase 380 or 440V. For the bus application, the power capacity of the power inverter has been selected with a 100 – 200 kW range, and can be scaled-up depending on the required electric load of different applications. The pick-up coil sets attached under the vehicle’s bottom-floor are tuned to a 20 kHz resonant frequency and are designed to have maximized exposure to the generated magnetic field, which has an optimized field shape for the same purpose. In this way, the transmission efficiency can be maximized while reducing the magnetic field leakage outside of design-intended space.

The design objective is to obtain the maximum power transmission efficiency with the pre-determined level of required power capacity by optimizing the paired power supply and collection system design with the alternate current magnetic field shapes at 20 kHz of resonant magnetic power transmission.

The shaped magnetic field concept and the coverage of the magnetic field by pick-up devices are also shown in a schematic manner in Figure 6. This system is called as a dual type power supply system due to its magnetic shape.

Figure 5: Schematic diagram of SMFIR technology system [9]

3.3 Power supply system architecture design
In Figure 7, the shapes of electromagnetic field are shown as the simulation results of the mono and dual types during their numerical iteration process of optimization. The formulation and schemes for numerical simulation are well described in Ref. [9]. With different lay-ups of power cable, combined with different geometries of the ferrite core at the bottom of the cable, the magnetic field shape, paired with the pick-up devices design, and the resultant performance of wireless power transmission can be different. By placing the ferrite structures in an optimized way, the magnetic field shape can also affect the maximized exposure to the surface

<table>
<thead>
<tr>
<th>Parameter (Units) of fully burdened system</th>
<th>Minimum Goals for Long Term Commercialization</th>
<th>Long Term Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Density (W/L)</td>
<td>460</td>
<td>600</td>
</tr>
<tr>
<td>Specific Power - Discharge, 80% DOD/30 sec (W/kg)</td>
<td>300</td>
<td>400</td>
</tr>
<tr>
<td>Specific Power - Regen, 20% DOD/10 sec (W/kg)</td>
<td>150</td>
<td>200</td>
</tr>
<tr>
<td>Energy Density - C/3 Discharge Rate (Wh/L)</td>
<td>230</td>
<td>300</td>
</tr>
<tr>
<td>Specific Energy - C/3 Discharge Rate (Wh/kg)</td>
<td>150</td>
<td>200</td>
</tr>
<tr>
<td>Specific Power/Specific Energy Ratio</td>
<td>2:1</td>
<td>2:1</td>
</tr>
<tr>
<td>Total Pack Size (kWh)</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Life (Years)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Cycle Life - 80% DOD (Cycles)</td>
<td>1,000</td>
<td>1,000</td>
</tr>
<tr>
<td>Power &amp; Capacity Degradation (% of rated spec)</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Selling Price - 25,000 units @ 40 kWh/($/kWh)</td>
<td>&lt;150</td>
<td>100</td>
</tr>
<tr>
<td>Operating Environment (°C)</td>
<td>-40 to +50</td>
<td>-40 to +85</td>
</tr>
<tr>
<td>Normal Recharge Time</td>
<td>6 hours</td>
<td>3 to 6 hours</td>
</tr>
<tr>
<td>High Rate Charge</td>
<td>20-70% SOC in &lt;30 minutes @ 150W/kg (&lt;20min @ 270W/kg Desired)</td>
<td>40-80% SOC in 15 minutes</td>
</tr>
<tr>
<td>Continuous discharge in 1 hour - No Failure (% of rated energy capacity)</td>
<td>75</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 1: Requirements on advanced batteries for EV application by USABC [5]

Figure 4: Lithium-ion battery cost projection by McKinsey & Company in 2009 [6]
pick-up coils. Numerous design iterations have been performed in order to achieve the maximized power transmission to the vehicle, while reducing the magnetic field intensity to the leakage field.

![Figure 7: Magnetic field shape from simulations for Mono and Dual Types [9, 11]](image)

As a practical and exemplary application of the SMFIR technology on the road, the powered track is designed as shown in Figure 8. One powered track is composed of a set of segments with different lengths. One segment is a defined length of powered cable loop, operated with single switching mechanism controlled by the power inverter responding to the vehicle identification sensor when the vehicle is approaching to the segmented cable loop. The length of the segment can be a design variable depending on the road conditions, the vehicle speed, the operating condition of acceleration or deceleration, the presence of heavy traffic volume with possible traffic jams or highways, and the presence of BRT (Bus Rapid Transit) lanes. For example, bus stops for public transportation can have a short length of the segment that is approximately the same length of the total pick-up sets installed under the city bus.

![Figure 8: Schematic diagram of single power track architecture with six segments](image)

As an example of a practical design application, KAIST demonstrated the SMFIR technology in Seoul Grand Park with three of six-segmented powered tracks, which are composed of one of 2.5 m and five of 24 m segments as a powered track, respectively, as described in Figure 9. While three powered tracks were composed of six segments, the other had two segments of 2.5 m for the stationary charging while the vehicle was idling and waiting. In the park, the total length of powered tracks is 372.5 m, which is about 17% of the total travel distance during one-round trip of 2.2 km.

![Figure 9: Powered track installation example in the demonstration project in Seoul Grand Park [9, 11-13]](image)

This example design can be applied to other urban transportation applications by modifying design variables such as the lengths of the segments, the combination of different lengths of segments to form a powered track, and the arrangement of the powered track considering the instantaneous required power and averaged electric energy consumption during the vehicle travel. It should be noted that this demonstration project also provides a business case snapshot in applying the SMFIR technology to urban transportation.

### 3.4 Electric vehicle system architecture

An OLEV can be placed in the category of an electric vehicle because an electric vehicle is defined as a vehicle driven by an electric power train with one or more electric motors and a storage medium for electric energy, usually a battery. The unique difference between an OLEV and a more general electric vehicle is that an OLEV has a set of pick-up devices installed under the vehicle to collect the electromagnetic field energy. A set of electrical devices including rectifiers and regulators, which convert and deliver the electricity in the required form inside the vehicle, must also be installed within the OLEV.

A power control and management system is also necessary within the vehicle such as a Power Distribution Unit (PDU) to control the power flow from the electric source of the battery and the supplied power from the road. The control of the electrical power flow and communication of the necessary signals within a vehicle are also incorporated for proper operation of the vehicle. In order to show the difference in the system architecture of an OLEV vehicle, a comparison was made between the series hybrid electric vehicle and an OLEV system in Figure 10. Compared with the series hybrid power train, OLEV does not have an IC engine, thus there is no tail-out emission.
Series hybrid system (a) OLEV system

An example of the physical packaging of the pick-up devices is shown in Figure 11. Depending on the electrical load requirement for a vehicle and the power collecting capacity from one unit of the pick-up device, the proper packaging design can be derived by considering the required vehicle ground height. In the example, four sets of pick-up devices have been installed under the vehicle frame to have the power collecting capacity of 60 kW, while keeping the required ground height of 13 cm.

In buses operating in KAIST campus, the pick-up power capacity is 75 kW with five sets of pick-up devices with the variable ground height of 15-20 cm, providing enough required power to drive full size city buses with 120 kW rated and 240 kW peak power AC induction motor [8, 12].

3.5 Performance parameters and achievements

The critical performance parameters can be defined as follows: power transmission efficiency, vehicle ground height and power capacity. Thanks to the SMFIR technology, we achieved a transmission efficiency of 83% at a ground height of 20 cm and a 75 kW of power capacity. This is a ground-breaking record in the wireless power transfer field, and is a critical performance achievement for the commercial deployment to the future urban transportation.

4 APPLICATION OF SMFIR TECHNOLOGY TO URBAN TRANSPORTATION

4.1 Demonstration projects test beds

In addition to the demonstration project described above and summarized in Figure 12 (a), another test bed is under operation at the KAIST Munji campus, located in Daejeon, Korea, as shown in Figure 12 (b). This site has a series of various lengths of segments, 3m, 5m, 40m and 60m, for different application practices. Also it is fully utilized for environmental evaluation and operation monitoring purposes. KAIST is also planning to install another test bed in the KAIST main campus in 2011 as shown in Figure 12 (c).
4.2 Analytic design process for urban transportation application

As part of the engineering design and development process, a new installation has been selected to apply the power supply infrastructure. The subject site will operate a couple of OLEV buses over a 6 km one-way trip, which will take about 20 minutes for each trip. The objective is to design the most efficient and optimized power supply infrastructure. This includes identifying how long the powered track should be, where it should be installed, and what combination of the segments should be laid, in view of the overall energy balance between the energy consumption and charging during operation and the instantaneous power requirement to drive the vehicle.

From the measured (or predicted or required) vehicle velocity profile over time and distance, as shown in Figure 13, it is possible to calculate the required electric power to drive the vehicle of Figure 14, given the vehicle weight, frictional and wind resistance and other vehicle and road information, shown in the equations (1) and (2).

\[
F = W \alpha + (R_f + R_s + R_g) \frac{v^2}{2}
\]

(1)

\[
P = \frac{Fv}{m}
\]

(2)

Here, the variables are defined as follows:

- \(W\): vehicle weight [kg]
- \(\alpha\): vehicle acceleration [m/sec²]
- \(R_f\): rolling resistance coefficient
- \(R_s\): air resistance
- \(R_g\): grade resistance (= \(\tan \alpha\), \(\alpha\): grade angle)
- \(F\): force required to drive the vehicle [N]
- \(v\): velocity required for the vehicle [m/sec]
- \(P\): power required to drive the vehicle [kW].

From the calculated required power, the energy consumption during one trip can be estimated as shown in Figure 15.

From the calculated energy consumption per trip and the instantaneous power requirement for the vehicle speed and road conditions, the location and distance of the powered track can be identified with the given battery discharge capacity. In this example, the vehicle is assumed to have a battery with energy capacity of 25 kWh and the recommended discharge C-rate for the Li-polymer battery is assumed to be C/3, thus the battery in the vehicle can cover up to 75 kW of instantaneous power requirement as shown in Figure 16.

As a rule of thumb, installing the powered track is required where the required driving power exceeds the instantaneous battery discharge power capacity, so that the vehicle can have enough power to be driven. The battery capacity can be an additional design variable as well, but the market status of the battery weight, cost and packaging volume, etc. should be considered at the same time. On the other hand, the power collecting capacity of the pick-up devices is another design variable when determining the battery energy capacity and the power track length design.
Figure 17: Battery SOC changes during the vehicle operation (a) pure electric vehicle (b) OLEV

By performing the design optimization process considering the profiles of the system components, the optimized powered track length combined with the pickup power capacity and the energy storage capacity of the battery can be determined. While managing the required instantaneous power and the overall energy consumption during the vehicle travel of the closed circuit, the battery status of charge (SOC) should be monitored especially for the Li-ion family battery. One of advantages with the SMFIR technology application in the future urban transportation is being able to manage the battery SOC duty cycle with a lot less bandwidth than the one for the pure electric vehicle, as compared in Figure 17. The pure electric vehicle will have a nearly full duty cycle swinging the SOC from 20 to 90 %, however, OLEV will only swing between 40 to 60 % thanks to its dynamic charging characteristic under operation.

5 SUMMARY

In this paper, the principle of the SMFIR (Shaped Magnetic Field in Resonance) technology has been introduced so that the electric vehicle or OLEV can be dynamically charged from the road surface while the vehicle is in motion. The vehicle system architecture and power supply infra-structure design process and its examples are described and discussed.

For the practical application of SMFIR technology in future urban transportation, the demonstrative test beds are described with the design process while achieving the required performance parameters. With the technology innovation, the fixed design variables, or design constraints, in launching electric vehicles such as the battery energy storage capacity, charging station location, and operating charging distance and time, etc. have been moved to the design variable domain.

Thus, in the view of design strategy for future urban transportation systems, SMFIR technology can provide a greater deal of design flexibility in the charging facility and electric vehicle launch motivated by the CO2 reduction effort.

6 ACKNOWLEDGEMENTS

The OLEV project was supported by a research grant through the Ministry of Education, Science and Technology, Republic of Korea in 2009, and the Ministry of Knowledge and Economy in 2010. A number of colleagues in the KAIST OLEV Project Center have made important contributions in developing and implementing the SMFIR technology.

7 REFERENCES

Magnetic Field Design for Low EMF and High Efficiency Wireless Power Transfer System in On-Line Electric Vehicles

Department of Electrical Engineering and Computer Science
KAIST, Daejeon, 305-701, Korea
sahn@kaist.ac.kr

Abstract
In the mechanism of wireless power transfer in on-line electric vehicle (OLEV), the design of magnetic field is the key technology to determine its electrical performance of power transfer capacity, transfer efficiency, and electromagnetic field (EMF) level. To satisfy all the requirements, systematic approach for optimization of design parameters is required. Even though shielding for reduction of EMF can be applied independently, the shielding effectiveness of the applicable shielding method should be considered in optimization of design parameters. In this paper, we introduce the wireless power transfer mechanism and the EMF reduction techniques, and perform design parameter optimization to maximize transfer power efficiency while satisfying power transfer efficiency and EMF regulation.

Keywords:
On-line electric vehicle, Electromagnetic Field, Wireless Power Transfer, Efficiency, Optimization, Linear Programming

1 INTRODUCTION
Even though intensive research has been performed on fully electric transportation systems, we are still facing serious problems in battery-powered electric delivery systems. These issues include the large size, weight, and cost of batteries, long recharging times, and limited availability of charging service points. Moreover, diminished stocks of lithium could lead to increasingly high prices and ultimately cause electric vehicles to price themselves out of the automotive marketplace.

KAIST has introduced a novel on-line electric vehicle (OLEV), in which the automotive vehicle constantly receives and recharges its power from power lines embedded underneath the surface of the road (Figure 1). OLEV has a minimal battery capacity (about 20% compared to that of the conventional battery-powered electric vehicles) which can consequently minimize the weight and the price of the vehicle and power station.

Figure 1: Photograph of on-line electric vehicle system

One of the key design requirements of the OLEV system is the suppression of the leakage magnetic flux from power lines and the pickup module to maintain the power delivery efficiency and meet the total power needs of the OLEV. In this paper, we propose techniques for the reduction of magnetic flux from the OLEV system. Some passive and active shielding methods are applied to real vehicles based on simulations and measurements, and the application to real vehicles is shown.

2 POWER TRANSFER MECHANISM
The power transfer system for OLEV consists of an inverter, power lines, a pickup module, capacitors, a battery, and a motor, as shown in Figure 3. 60 Hz for power transfer is converted to 20 kHz at the inverter stage and a current of about 200A flows through the power lines. The magnetic flux generated from the power lines is gathered at the pickup module to generate DC power for the vehicle motor. The non-contact power transfer that occurs between the power lines and the pickup module generates a huge magnetic flux. So, the design of the power lines and the pickup module are the key technologies for effective power transfer and the solution of the electromagnetic field (EMF) problems.

Figure 3 shows the vertical magnetic flux of the power lines and pickup module. There are two power lines with opposite current directions underneath the road surface forming a current loop. Due to the current in the power lines, a magnetic flux is induced around each power line. Between the power lines, the magnetic fluxes from the two power lines are added. The pickup module catches the vertical magnetic flux through copper coils around the ferrite core. This type has the advantage of efficient power transfer because the direction of the magnetic flux from the power lines is the same as the direction of the flux to the pickup module.
3 DESIGN METHODOLOGY

3.1 Definition and Formulation of Design Criteria

In the design of the power lines and the pickup module structure for OLEV system, we consider three criteria for the electrical performance of the wireless power transfer system: power transfer capability, power transfer efficiency, and leakage from the electromagnetic field.

The power transfer capability implies the maximum power that can be transferred from the power lines under the road to the load in the vehicle, which consequently determines the maximum speed and recharging time of the vehicle. From the simplified equivalent circuit model of the wireless power transfer system with two series resonant coils as shown in Figure 4, the power at the load $R_L$ is calculated to be proportional to the frequency, mutual inductance, and magnitude of source current assuming that the system is operating at the resonance frequency as shown in (1).

$$P_L \cong \frac{\omega^2 M^2}{(R_s + R_i)^2 + \left(\frac{1}{\alpha} L_2 - \frac{1}{\alpha C_2}\right)^2}$$

$$K \cong \frac{R_i}{R_i (R_s + R_i)^2 + \alpha^2 M^2 (R_s + R_i)} \cong \frac{1}{1 + \frac{R_i R_s}{\alpha^2 M^2}} \quad (2)$$

The power transfer efficiency is also an important factor for commercialization and it should be reasonably high compared with the efficiency of other types of vehicles. To increase the efficiency, we need to minimize the loss at each stage of the power system of OLEV. With the development of power components operating at 20 kHz, which was not available tens of years ago, the efficiency of the inverter in Figure 2 is significantly increased. Also, the mutual inductance should be increased, and the parasitic resistance $R_s$ and $R_i$ which are the loss from these resistances should be decreased as derived in (2) to increase the efficiency even more.

The third criterion of leakage EMF is simply proportional to the magnitude of the current and inversely proportional to the distance between current position and measurement position without a shield as shown in Eq. (3). However, as the application of passive and active shields significantly changes the magnitude of EMF, the design of the EMF should be performed separately which will be discussed in the next section.

3.2 Previous Procedure of Wireless Power Transfer System Design

The previous design procedure for the wireless power transfer system for OLEV is shown in Figure 5. At the early stage of design, we have to determine the topology and outline of the dimensions for the physical structures.
such as the number of coils, coil size and dimension and the position of the ferrite core because the mutual inductance is roughly determined when the physical dimension is fixed and it is hard to change the value significantly in the latter stage.

Table 1 shows the result of simulated sensitivity analysis of transferred power for the change of main design parameters which is the reference for the optimization of the design. At each design stage, a sensitivity analysis on the effect of each design parameters has been performed using simulation with 3-dimensional field solver.

<table>
<thead>
<tr>
<th>Design Parameters</th>
<th>Change of Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-20%</td>
</tr>
<tr>
<td><strong>Dimension Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Air Gap</td>
<td>+46.3%</td>
</tr>
<tr>
<td>Number of Turns in Pickup Coil</td>
<td>-44.0%</td>
</tr>
<tr>
<td>Dist. between Rail Wires</td>
<td>-40.0%</td>
</tr>
<tr>
<td>Pickup Coil Width</td>
<td>-24.2%</td>
</tr>
<tr>
<td><strong>Material Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Permeability ($\mu$)</td>
<td>-1.0%</td>
</tr>
<tr>
<td>Permittivity ($\epsilon$)</td>
<td>0%</td>
</tr>
<tr>
<td>Conductance ($\sigma$)</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Electrical Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>-44.0%</td>
</tr>
<tr>
<td>Frequency</td>
<td>-44.1%</td>
</tr>
</tbody>
</table>

Table 1: Sensitivity analysis of transferred power for the change of design parameters

4 **SHIELDING FOR REDUCTION OF EMF**

4.1 **Passive Shielding**

Figure 6 shows the magnetic flux density distribution of OLEV. In the case of the vertical magnetic flux type, there is one magnetic flux path between the power lines and pickup module where the power is transferred. The return flux comes back to the power lines via the sides of the main flux path. The horizontal magnetic flux type has two magnetic flux paths. The side power lines of this type have return flux paths on the side of the main flux path. The return flux path creates the fringing magnetic flux, and this flux is measured as the EMF level of OLEV. In this work, the target EMF level of OLEV is 62.5 mG according to the regulation of Korea Communications Commission which follows the ICNIRP design guideline [4-5].

As the power supply system of OLEV generates large amounts of magnetic field to transfer 60 kW of power which is necessary for the vehicle, there are tens of thousands mG of magnetic flux between the power lines and pickup module beneath the vehicle while power is transferred. So, if even 0.1% of leakage magnetic field comes out from OLEV system, the EMF level could exceed the regulation of 62.5 mG. The distribution of magnetic field for OLEV is shown in Figure 6. The number of connections using metal brushes is significant to improve the shielding effectiveness of the passive shield, we additionally applied soft contacts between bottom plate and vertical ground plate by metal brushes as shown in Figure 8. Thus, the design of the metal brushes is critical to improving the EMF level. The photograph of implemented metal brush is shown in Figure 9.

The number of connections using metal brushes is a significant factor to improve the shielding effectiveness of the passive shield. As shown in Figure 10, the EMF level has been decreased from 144 mG to 35 mG when the number of connections using metal brushes is increased from 2 to 8.
4.2 Active Shielding

The EMF can be minimized by active shielding with or without passive shields independently, and the basic concept of active shield is shown in Figure 11. Similar to power lines, the active shield is also a metal wire which carries the same frequency with current but the phase is the opposite of the current in the pickup.

In the design of active shield, the directions of magnetic fields by the source and active shield should be carefully considered. In Figure 12, the direction of magnetic field is shown. To make the EMF level less than the regulation at all positions, the magnetic field from the active shield should be almost the same as that from pickup module at all positions. At the position above 20cm from road surface, the magnetic field vector is parallel to the metal plate because of the metallic shield at the bottom of the vehicle. So, to place the active shield close to the pickup coil is more effective. However if the active shield goes closer to the pickup coil, the current of the active shield should be larger. For this reason, the placement of the active shield is compromised considering the shielding effectiveness and current magnitude.

In Figure 13 (a) and (b), the magnetic flux density with and without active shielding is depicted. When the active shield is applied, the leakage magnetic flux is cancelled by the magnetic flux from the active shield and significantly reduced to less than the regulation of 62.5 mG. Figure 12 shows the optimization procedure of the active shield design where the position and current magnitude should be determined. At the optimal value of current, the magnetic flux density is reduced to 1/10 of the density without the active shield as depicted in Figure 14.
5 DESIGN PARAMETER OPTIMIZATION

5.1 Formulation of Design Parameters

In this section, we formulate a parameter optimization problem such that the transferred power to pickup, $P_{\text{transfer}}$, which is consumed at the load $R_L$ of Figure 2, is maximized while EMF level and power transfer efficiency, $K$, satisfy the requirements. We assume that the power transfer efficiency should be greater than or equal to 0.8, and the leakage EMF should be less than or equal to 62.5 mG.

Table 2 shows system parameters, which are divided into two categories: constant system parameters and variable system design parameters. We assume that the air-gap between power lines and pickup coils, resonance frequency, parasitic resistance of power lines, parasitic resistance of pickup coil, and load resistance are given as in Table 2. We can change three system design parameters: width of pickup coil $W_C$, current of power lines $I_S$, and number of turns in pickup coil $n$.

Accordingly, we formulate our optimization problem as follows:

variables: $W_C$, $n$, $I_S$

maximize $P_C$

such that

$$EMF \leq 62.5 \text{ mG},$$

$$K \geq 0.8,$$

$$0 \leq W_C \leq W_{C,\text{max}}, 0 \leq n \leq n_{\text{max}}, 0 \leq I_S \leq I_{S,\text{max}}.$$  (1)

Figure 15 shows the effect of $W_C$ on $V_C$ and the EMF. The difference between the simulation result and mathematical model should be minimized to improve the accuracy of the design parameter optimization procedure.

From Figure 15, we obtain the approximate expressions for $V_C$ and EMF as follows:

$$V_C \approx c_1 f n I_S \sqrt{W_C},$$  (2)

$$EMF \approx c_2 n I_S W_C^2.$$  (3)
where $C_1$ and $C_2$ are constants. Then, transfer power $P_{\text{transfer}}$ and total power $P_{\text{total}}$ at resonant frequency can be represented as

$$P_{\text{Transfer}} = \frac{C_1^2}{R} f^2 n^2 \frac{1}{2} W C,$$  \hspace{0.5cm} (4)

$$P_{\text{Total}} \approx R L C f^2 n^2 \frac{1}{2} W C .$$  \hspace{0.5cm} (5)

Therefore, the power transfer efficiency is

$$K = \frac{P_{\text{Transfer}}}{P_{\text{Total}}} = \left(\frac{1 + \frac{R L C}{C_1^2 f^2 n^2 W C}}{1} \right) .$$  \hspace{0.5cm} (6)

From (3), (4), (6), we can express the optimization problem in (1) as follows:

$$\max \alpha_1 f^2 n^2 \frac{1}{2} W C$$

subject to

$$nl_3 W C^2 \leq \alpha_2$$

$$f^2 n^2 W C \geq \alpha_3$$

$$0 \leq W C \leq W_{C, \text{max}} , 0 \leq n \leq n_{\text{max}} , 0 \leq l_3 \leq l_{S, \text{max}}$$

where

$$\alpha_1 = \frac{c_1}{R} , \quad \alpha_2 = \frac{62.5}{c_2} , \quad \alpha_3 = \frac{R L C}{c_1^2 f^2 \left(\frac{1}{0.8}\right)} .$$

Let $x = \log(n), y = \log(l_3), z = \log(W C)$ . Then, the optimization problem in (7) can be restated as:

$$\max 2x + 2y + z + \beta_1$$

subject to

$$x + y + 2x \leq \beta_2$$

$$2x + z \geq \beta_3$$

$$x \leq x_{\text{max}}, y \leq y_{\text{max}}, z \leq z_{\text{max}}$$

where

$$\beta_1 = \log(\alpha_1), \beta_2 = 1.2 \beta_3, x_{\text{max}} = \log(n_{\text{max}}),$$

$$y_{\text{max}} = \log(l_{S, \text{max}}), z_{\text{max}} = \log(W_{C, \text{max}}).$$

Note that the problem (8) is a form of typical linear programming (LP) problem.

5.2 Numerical Results of Design Parameters

In the process of finding optimal design parameters, the parameters which maximize the transfer power are determined. The width of pickup coil should be minimized because it increases EMF more significantly than current and number of turns. Similarly, the current and the number of turns should be increased unless it violates the boundary conditions. The boundary conditions on the power transfer efficiency affect the design parameters when the frequency is low or mutual inductance is small. Once the product of frequency and mutual inductance is large enough, the EMF is the only boundary condition, and then the combination of the design parameters is determined to make the EMF 62.5 mG which is the maximum value allowed in the optimization. In this EMF boundary, the current and number of turns are maximized until they reach the maximum value we set as $W_{C, \text{max}}, n_{\text{max}}, l_{S, \text{max}}$ in (7). Finally, two maximum values of $n_{\text{max}}, l_{S, \text{max}}$ determine the transferred power because the number of turns and current should reach the maximum value for maximum power.

Now, we obtain the optimal solution for problem (8) and compare it with the simulation results to investigate the validity of the approximation for LP formulation. Figure 16 shows the optimal transfer power $P_C$ and the variation of constraints such as EMF and $K$ for different values of the frequency $f$. The optimal power increases as the frequency increases because frequency simply increases the transfer power and has no effect on EMF. The efficiency and EMF should be maintained at the specific level. We can find that the simulation results are similar to the LP solution, which means that the approximation for LP formulation is reasonable. More accurate results can be obtained by applying more complex numerical models in (2) and (3) which describes the voltage and EMF more accurately.

Figure 16: Optimal transferred power, EMF, and efficiency for different values of frequency
Figure 17 plots the optimal power for different values of the frequency $f$ and the load $R_L$. When the load resistance increases, the optimal power decreases and the power transfer efficiency slightly decreases. At the center of the surface in Figure 17, there is an edge across the surface, which is generated due to the power efficiency boundary condition. The power transfer efficiency boundary is critical when the frequency and load resistance are in this range.

**Figure 17: Optimized transferred power for different values of frequency and load resistance**

6 CONCLUSIONS

In the design of the wireless power transfer system in OLEV, the design of electromagnetic field is the most important for optimal electrical performance. To maximize power transfer capacity with high transfer efficiency and without violating EMF regulation, systematic design approach is necessary. The two procedures of reducing EMF and optimizing the wireless power transfer system design parameters are performed. For a more accurate design, more complex modelling of design parameters is required. For implementation with real vehicle, the power capacity of 60 kW using 5 pickup modules, with 80% power transfer efficiency, and EMF level lower than 62.5 mG have been achieved.

7 ACKNOWLEDGMENTS

This work was supported by the OLEV project of KAIST and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0029179).

8 REFERENCES


